Light allows for photosynthesis and other essential light-dependent chemical reactions. Light also triggers photo-oxidative stress via generation of reactive oxygen species, which rapidly damage the cell. Organisms in all domains of life produce proteins capable of sensing light. These biological photoreceptors participate in vision, harness light energy, regulate circadian clocks, and mediate gene expression and major developmental processes. Different classes of photoreceptor proteins with a variety of chromophore cofactors exist to sense light over the visible and ultraviolet spectrum. This group of photoreceptors was recently expanded by a new class that is widespread in bacteria and uses the vitamin B12 derivative adenosylcobalamin (AdoCbl) as the chromophore to orchestrate light perception and response. The prototypes of this class, the CarH-type photoreceptors, have been characterized in *Myxococcus xanthus* and *T. thermophilus* and typically regulate light-induced expression of carotenoid biosynthetic genes, which result in carotenoid-mediated protection against photo-oxidative damage. This class of photoreceptors thereby allows bacteria to mitigate the detrimental effects of sunlight, a critical determinant of survival in light-exposed conditions, while avoiding unnecessary production of carotenoids in the absence of light.

The CarH-type photoreceptors consist of an amino (N)-terminal DNA-binding domain and a carboxy (C)-terminal AdoCbl-binding and oligomerization domain to directly sense light and regulate gene expression. In the dark, AdoCbl-bound CarH, a tetramer, binds to the promoter region of target genes to repress transcription. Exposure to blue, green, or ultraviolet light disrupts the photosensitive Co–C bond in AdoCbl, leading to tetramer disassembly, loss of operator-binding, and activation of gene expression. Thus, AdoCbl, which is typically used as a cofactor for radical-based enzyme reactions, is now being used for a new biological function as a light sensor. Here, we sought to determine the structural basis for the functional repurposing of one of nature’s most complex metallocofactors and for this new mode of light-dependent gene regulation.

**CarH has a modular architecture**

To visualize the CarH domain structure and the architecture of the repressor ‘dark’ state, we first determined two independent structures of AdoCbl-bound CarH from *T. thermophilus* to 2.15 Å and 2.80 Å resolution (Extended Data Table 1). To prevent cleavage of the light-sensitive AdoCbl Co–C bond, we performed all crystallization experiments under red light and all diffraction data collection at $T = 100$ K. Both single-crystal ultraviolet–visible (UV–vis) spectra and the electron density confirmed that AdoCbl remained intact with a Co–C bond length of 2.0 Å (Extended Data Fig. 1), indicating that the structures are in the ‘dark’ state. Each monomer of CarH has a modular three-domain architecture with an N-terminal winged-helix DNA-binding domain followed by the light-sensing domain, composed of a four-helix bundle and a C-terminal Rossmann-fold cobalamin (Cbl)-binding domain (Fig. 2a).

The DNA-binding domain is structurally similar to those of MerR family transcription factors and to that of the AdoCbl-independent CarH parologue CarA, featuring a canonical recognition helix and a β-hairpin wing for DNA binding (Fig. 2a). In our DNA-free structures, the DNA-binding domain samples different orientations for the different CarH protomers in the asymmetric unit, enabled by a flexible linker region and stabilized by crystal lattice contacts (Extended Data Fig. 2).

In contrast to the flexible DNA-binding domains, the four-helix bundle and the Rossmann domain are structurally rigid, together forming a module that binds the AdoCbl light sensor. AdoCbl is sandwiched between the four-helix bundle, which interacts with the upper axial $\delta$-deoxyadenosyl (5′-dAdo) ligand, and the Rossmann-fold domain, which binds the Cbl lower face in the base-off/His-on mode with the side chain of His177 displacing the Cbl dimethylbenzimidazole base (Fig. 2a, d and Extended Data Fig. 1b). Instead of closely resembling an AdoCbl-dependent enzyme, the CarH light-sensing domain is structurally homologous to the methylcobalamin...
Figure 1 | Schematic of CarH-mediated light-dependent gene regulation.
Structures of all three relevant states are reported herein. Red circles depict
AdoCbl, filled red semicircles photocystolysed Cbl, and open red semicircles
4′,5′- anhydroadenosine, the recently identified photolysis product of CarH-bound
AdoCbl33. See main text for details.

(MeCbl)-binding module of methionine synthase MetH (Fig. 2b)17, even
though the AdoCbl 5′-dAdo group is much bulkier than the MeCbl methyl group in MetH. Modelling AdoCbl into MetH leads
to several steric clashes (Fig. 2c), but in CarH, a small but important
shift of 2.5 Å of the four-helix bundle enlarges the cavity on the Cbl
upper face (Fig. 2d, e). Additionally, four hydrophobic residues at the
Cbl upper face in MetH are replaced in CarH, providing the 5′-dAdo
group a larger binding pocket (Leu715, Cbl upper face in MetH are replaced in CarH, providing the 5′-dAdo group a larger binding pocket (Leu715→Val1138), a hydrogen bonding
interaction (Val718→Glu141), and more polar environment in general (Val719→His142, Phe708→Trp131) (Fig. 2d, f). Although
Trp is larger than Phe, the orientation of the Trp side chain on the
side of the upper ligand rather than directly above perfectly accom-
modates the 5′-dAdo group (Fig. 2e, f). Notably, Trp131, Glu141, and
His142 are highly conserved in CarH homologues, suggesting that
this mode of AdoCbl binding is conserved as well (Extended Data
Fig. 3).

CarH is a dimer-of-dimers type tetramer
AdoCbl-bound CarH is a tetramer in the crystal structure (Fig. 3),
consistent with results from size-exclusion chromatography (SEC)
and analytical ultracentrifugation38,39. Four light-sensing domains
comprise the core of the tetramer (Fig. 3a–d) with the DNA-binding
domains extending outward (Fig. 3e, f). The core has a dimer-of-
dimers architecture, with each constituent dimer composed of two
CarH light-sensing domains in a head-to-tail orientation (Fig. 3a).
The extensive head-to-tail dimer interface is formed by the four-helix
bundles and the Cbl-binding domains and features a solvent-buried
area of 1,430 Å² on each protomer as well as numerous hydrogen
bonds and ionic interactions involving various side chains and the
5′-dAdo group of AdoCbl (Fig. 3a, b). Two such head-to-tail dimers
assemble to a tetramer in a staggered fashion (Fig. 3c–f). This dimer-
dimer interface is again formed by the light-sensing domains, creating
a buried surface area of 1,590 Å² on each of the two head-to-tail
dimers, whereas the four DNA-binding domains are positioned on
the surface of the tetramer and only make minor contributions to the
interface (Fig. 3e, f). It was previously demonstrated by SEC that CarH
lacking the DNA-binding domain still forms tetramers40, and here we find that CarH adopts the same tetramer architecture when the DNA-
binding domains are proteolytically removed during crystallization
(Fig. 3c, e). Thus, the light-sensing domains appear to mediate tetra-
merization, positioning the DNA-binding domains on the surface to
engage DNA.

To confirm the CarH tetramer architecture and the mode of AdoCbl
binding, we mutated residues in the AdoCbl binding site, at the head-to-
tail dimer interface, and at the dimer–dimer interface and analysed these
mutants using SEC and electrophoretic mobility shift assays (EMSAs)
(Extended Data Figs 3 and 4). Non-conservative mutations in the bind-
ing site for the 5′-dAdo group (W131A, E141A; Fig. 2d) impaired
tetramer formation and mutations near the head-to-tail dimer interface
(H142A, D201R; Fig. 3b) completely abolished it. Remarkably, the most
dramatically adverse H142A and D201R mutations also appeared to impair
AdoCbl binding (Extended Data Fig. 4a, c; absorbance traces at 522 nm).
Moreover, DNA binding affinity weakened with decreased ability to form
tetramers in the W131A, E141A, H142A, and D201R mutants
(Extended Data Figs 3a and 4g). For comparison, we also introduced the
conservative W131F mutation, which behaved like wild-type (WT)
CarH in its oligomerization and DNA binding properties (Extended Data
Fig. 4a, g). The inability of the D201R mutant protein to oligomerize
is consistent with the observed Asp201/Arg176 interaction playing an
important role in stabilizing the head-to-tail dimer interface (Fig. 3a).
We expected a second compensatory mutation, R176D or R176E, to restore
the interaction and, indeed, the R176D/D201R and R176E/D201R double
mutants could form tetramers (Extended Data Fig. 4e) and bind to
DNA, albeit less efficiently than WT CarH (Extended Data Fig. 4g).
Finally, replacing Gly160 and Gly192 at the CarH dimer–dimer interface
(Fig. 3d) by the bulkier Gln resulted in dimers in the presence of AdoCbl,
a form previously never observed for WT CarH (Extended Data Fig. 4e).
Although both these mutants bound to the DNA probe in the dark, they
failed to engage DNA.

CarH binds the promoter – 35 element
To reveal the mode of transcriptional repression, we next sought to
visualize CarH bound to its cognate DNA operator. The CarH operator
lies within a 110-base-pair (bp) segment of the intergenic region
between carH and the carotenogenic crtB40. Using systematically
truncated DNA probes in EMSAs (Extended Data Fig. 5a, b), this operator
...
The structure revealed a unique mode of DNA binding involving three of the four DNA-binding domains of tetrameric CarH (Fig. 4a). The fourth DNA-binding domain is disordered and not visible in the electron density. The overall architecture of the tetramer is the same before and after DNA binding except for a reorientation of the DNA-binding domains (Extended Data Fig. 6e). All three DNA-binding domains face the same way on the DNA segment and bind to a set of three 11-bp repeats with a consensus sequence (A/G)(A/G)(A/C)T(A/G/T) (Fig. 4a). This parallel orientation is stabilized by specific interactions between adjacent DNA-binding domains (Extended Data Fig. 6f). The central DNA-binding domain comes from one head-to-tail dimer, whereas the two flanking DNA-binding domains come from the second head-to-tail dimer (Extended Data Fig. 6g, h). These structures suggest that the two individual head-to-tail dimers would bind to DNA less tightly, consistent with the reduced affinity and cooperativity for the dimeric G160Q and G192Q CarH mutants (Extended Data Figs 3a and 4g).

To obtain support for this unusual DNA binding mode, hydroxyl radical and DNase I footprinting were used to examine the regions of DNA protected by CarH. The DNase I footprint, which was obtained after DNA binding, protected the same element. Nucleotides protected from hydroxyl radical cleavage are indicated by grey. AdoCbl shown as in Fig. 2a. DNA-binding domains are shown in cyan.

Figure 3 | CarH oligomerization. a, CarH protomers arranged in a head-to-tail dimer, coloured by domain (helix bundle: yellow; Cbl-binding domain: green) with left protomer shown in lighter colours. AdoCbl shown as in Fig. 2a. DNA-binding domains hidden for clarity. b, Close-up of selected residues at the dimer interface. c, Core of CarH tetramer, assembled from two head-to-tail dimers. Top dimer is coloured as in a; bottom dimer is coloured in black and grey. Gly160 and Gly192 at the dimer–dimer interface are shown as red spheres. Structure is from a sample of CarH that degraded during crystallization and lacks the DNA-binding domains (crystal form 2, see Methods). d, Close-up of Gly160 and Gly192 (red spheres) from two protomers at the dimer–dimer interface. e, Tetramer of full-length CarH including DNA-binding domains (crystal form 3), coloured as in a. DNA-binding domains of coloured dimer are shown in dark cyan, those of grey dimer are shown in light cyan. f, Additional view of CarH tetramer, revealing how DNA-binding domains are positioned on the protein surface.

Figure 4 | CarH DNA binding. a, CarH tetramer bound to a 26-bp DNA segment (yellow). CarH is shown in ribbons with one head-to-tail dimer in green (Cbl-binding domains) and yellow (helix bundles) and the other dimer in grey. AdoCbl shown as in Fig. 2a. DNA-binding domains are shown in cyan. Sequence of DNA segment used for crystallization (larger font) as well as flanking sequences in the operator (smaller font) are shown below. Cyan bars indicate base pairs covered by each DNA-binding domain. Base pairs covered by the recognition helix are boxed; red box highlights the promoter −35 element. Nucleotides protected from hydroxyl radical cleavage are indicated by bullets. The orientation of the DNA in the structure was confirmed by heavy atom labelling (Extended Data Fig. 6b–d). b, Schematic of CarH–DNA interactions for the central DNA-binding domain, denoted as follows: black arrows, hydrogen bonds/electrostatic interactions; green arrows, van der Waals interactions; solid lines, contacts from protein side chains; dashed lines, contacts from main chain; (M), contacts in the DNA major groove; (m), contacts in the DNA minor groove. c, Close-up of interactions between CarH (cyan) and DNA (yellow). Hydrogen bonds and ionic interactions to the phosphate backbone are shown as black dashed lines. Contacts to DNA bases are shown in purple. Side-chain orientation is not unambiguous owing to the modest resolution of this structure, but many of the contacts shown are supported by mutagenesis (Extended Data Fig. 7).
using a longer DNA segment (130 bp) than the one used in the crystal structure (26-bp), still matches the crystal structure footprint (Extended Data Fig. 5a, c). Additionally, three evenly spaced 4-bp hydroxyl radical footprints are observed on both the sense and the antisense strand (Extended Data Fig. 5c) that correlate with where the ‘wings’ of the three DNA-binding domains contact the minor groove (Fig. 4a). Taken together, the size of the DNAse I footprint, the presence of three hydroxyl radical footprints, and the observation that the DNA sequence contains three repeats (see above), suggest that CarH binds to DNA using three of its four DNA-binding domains.

To determine whether all three repeats are important for high-affinity binding of CarH to DNA, we tested the effect of mutating DNA bases in the CarH operator (Extended Data Fig. 5d, e). Mutating dinucleotides in any single repeat only led to a small decrease in affinity as evidenced by the intense retarded bands for the CarH–DNA complex and the small amounts of free DNA in EMSAs (mutants 1–3 or 8–10 in Extended Data Fig. 5e). In contrast, simultaneously mutating dinucleotides in any two of the three repeats or in all three repeats almost completely abolished DNA binding (mutants 4–7 and 11–14 in Extended Data Fig. 5e). As a control, we also mutated DNA bases in the operator that CarH does not contact directly, and as expected, WT CarH bound to these mutants with similar affinity (mutants 15–18 in Extended Data Fig. 5e). Given that the results of mutations were similar for each of the three repeats, it appears that all three repeats are important in determining CarH–DNA affinity.

Each DNA-binding domain forms hydrogen bonds and electrostatic interactions to the phosphate backbone, contributed from both the peptide backbone and the side chains of Thr26, Tyr30, Arg28, Arg43, and Lys67 (Fig. 4b, c), and each domain also inserts His42 of its surface to the Cbl upper face, where it binds to the cobalt to form bis-His ligation site (Fig. 5d), prompting movement of the helix bundle to occupy this void and cover the Cbl (Fig. 5e).

Strikingly, the helix bundle motion brings His132 from the protein surface to the Cbl upper face, where it binds to the cobalt to form bis-His ligation Cbl (Fig. 5d, e and Extended Data Fig. 8b). Such bis-His ligation, common for haems, has not been reported for Cbl, although bis-His Cbl ligation was recently proposed on the basis of mass spectrometry for the Cbl-dependent transcription factor AerR.[19] We therefore validated formation of bis-His-ligated Cbl using UV–vis spectroscopy. Spectra of light-exposed WT CarH and a H132A mutant, which is unable to form the bis-His ligation, resemble those of free Cbl with two or one nitrogen-based ligands,[20], respectively (Extended Data Fig. 8c, d). In contrast, the spectra of the AdoCbl-bound proteins are identical (Extended Data Fig. 8e). These results provide unambiguous evidence for a bis-His-ligated Cbl in light-exposed CarH and suggest that this mode of coordination might be used more frequently in non-haem proteins.

Notably, both WT and H132A CarH undergo light-dependent tetramer disassembly, indicating that bis-His ligation is not required with a role of these residues in DNA binding. Notably, the Q25A mutant retained DNA-binding capacity, indicating that Glu25 could be more important for mediating DNA specificity than affinity (Extended Data Fig. 7c). Finally, the H42A and Y30A mutants only showed mildly reduced affinity, suggesting that their interactions are not essential for DNA binding (Extended Data Figs 3a and 7c).

**Light triggers helix bundle movement**

Finally, to examine how light exposure causes tetramer disassembly, we determined the crystal structure of light-exposed CarH to 2.65 Å resolution (Fig. 5, Extended Data Fig. 8a, b and Extended Data Table 1). The structure contains monomeric CarH with bound Cbl but without the 5′-dAdo group, which dissociated as a consequence of light exposure (Fig. 5a). The four-helix bundle and the Cbl-binding domain individually do not exhibit major conformational changes compared with the dark AdoCbl-bound structure. However, the orientation of the helical bundle relative to the Cbl-binding domain has changed drastically with a >8 Å displacement (Fig. 5a, b). This helix bundle movement would disrupt the head-to-tail dimer interface (Fig. 5c), leading to tetramer disassembly, dissociation from DNA, and transcriptional activation.

Tetramer disassembly is triggered by loss of the AdoCbl 5′-dAdo group: its presence in the AdoCbl-bound structure blocks movement of the helix bundle, owing in large part to the positioning of W131 against the upper Cbl ligand (Fig. 2d), keeping the CarH protomers in the extended ‘upright’ conformation required for tetramerization. Loss of the 5′-dAdo group leaves a large cavity on the Cbl upper face (Fig. 5d), prompting movement of the helix bundle to occupy this void and cover the Cbl (Fig. 5e).

Strikingly, the helix bundle motion brings His132 from the protein surface to the Cbl upper face, where it binds to the cobalt to form bis-His ligation Cbl (Fig. 5d, e and Extended Data Fig. 8b). Such bis-His ligation, common for haems, has not been reported for Cbl, although bis-His Cbl ligation was recently proposed on the basis of mass spectrometry for the Cbl-dependent transcription factor AerR.[19] We therefore validated formation of bis-His-ligated Cbl using UV–vis spectroscopy. Spectra of light-exposed WT CarH and a H132A mutant, which is unable to form the bis-His ligation, resemble those of free Cbl with two or one nitrogen-based ligands,[20], respectively (Extended Data Fig. 8c, d). In contrast, the spectra of the AdoCbl-bound proteins are identical (Extended Data Fig. 8e). These results provide unambiguous evidence for a bis-His-ligated Cbl in light-exposed CarH and suggest that this mode of coordination might be used more frequently in non-haem proteins.
for disassembly (Extended Data Fig. 8f, g). However, Cbl dissociation after light exposure is faster for H132A CarH than for WT CarH, which forms a very tight and stable complex with the photolyzed Cbl, as indicated by the relative abilities of the protein to be reconstituted with fresh AdoCbl (Extended Data Fig. 9a, b) and by the observation of a CarH:Cbl adduct in mass spectrometry for WT CarH but not for the H132A mutant (Extended Data Fig. 9c, d). Thus, the bis-His ligation could be important to retain the Cbl cofactor after photolysis.

Discussion

Impressively, CarH-type photoreceptors are found in hundreds of bacterial genomes, including bacteria that uptake rather than biosynthesize AdoCbl. CarH is distinguished from most known classes of photoreceptors (with the exception of some light-oxygen-voltage (LOV)-type photoreceptors21–23) in that it can bind to DNA directly, instead of requiring additional proteins for gene regulation. Beyond gene regulation, the CarH light-sensing domains can be found fused to effector domains such as histidine kinases and in stand-alone modules that could undergo light-dependent protein–protein interactions. This versatility probably explains the broad distribution of CarH-like proteins in bacteria.

The use of AdoCbl as a light-sensing chromophore by CarH is biologically unprecedented. AdoCbl is structurally and photochemically distinct from known photoreceptor chromophores such as bilin24, flavins25–27, retinal4, or Trp side chains28: light exposure leads to breakage of a covalent Co–C bond, whereas other chromophores undergo less drastic changes such as light-induced electron transfer or cis–trans isomerizations. In all cases, however, light energy is ultimately harnessed to drive a large-scale conformational change, highlighting the convergence of different light-sensing mechanisms. AdoCbl was previously best-known as a cofactor for radical-based enzyme reactions, in which reversible homolytic cleavage of the Co–C bond provides access to the 5'-dAdo radical, but rather 4',5'-anhydroadenosine34, which differs by one proton and one electron, and cannot cause radical damage. Thus, AdoCbl now joins the list of enzyme cofactors that have been repurposed as sensors; a list that already includes flavins (as light sensors in LOV, blue light sensor using flavin (BLUF), and cryptochrome photoreceptors25–27) and haem (as sensors of oxygen and other small molecules25).

Our CarH structures additionally reveal the functional repurposing of two different protein modules. The CarH light-sensing domain mirrors the Cbl-binding module of methionine synthase MetH, an enzyme that uses a MeCbl intermediate in the transfer of a methyl group from methyltetrahydrofolate to homocysteine, generating tetrahydrofolate and methionine34. But whereas MetH uses its helix bundle to position Phe708 over the MeCbl methyl group and protect it from photolysis34, CarH, enabled by specific substitutions at the Cbl upper face, uses this fold as an AdoCbl-binding light-sensing domain, in which the corresponding Trp131 senses the presence of the 5’-dAdo group and transmits the signal of AdoCbl photolysis by leading a conformational change of the helix bundle. Furthermore, whereas the Cbl-binding module of MetH is embedded in a 136 kDa multi-domain protein and juxtaposed to different substrate-binding domains via transient domain–domain interactions during a catalytic cycle35, the light-sensing domain of CarH is used to assemble a tetramer that is stable enough to occlude the −35 element from RNA polymerase. Thus, this module has been repurposed from a methyl group carrier in primary metabolism to a light-sensing modulator of oligomerization state.

Similarly, the CarH DNA-binding domain resembles those of MerR-type transcription factors such as MerR, BmrR, and SoxR, whose role as transcriptional activators in the presence of heavy metals or other stresses has been established and whose DNA-bound structures have been reported36,37,38,39. Whereas Mer proteins bind as dimers to a (pseudo-)palindromic DNA sequence and distort the DNA, which brings promoter elements into alignment for transcriptional activation (Extended Data Fig. 6i), the CarH tetramer uses its DNA-binding domains to bind to three contiguous repeat sequences in a parallel mode, which occludes the −35 element and represses transcription. This unique DNA binding mode rationalizes the tetramer architecture of two-head-to-tail dimers of CarH, which is unusual for transcription factors but here enables the DNA-binding domains to arrange in a parallel fashion and cooperatively engage the repeat sequences.

AdoCbl is a biologically expensive molecule, requiring arduous pathways for biosynthesis or specialized machinery for uptake. For CarH-using organisms, it would not be surprising if there were a recovery mechanism for Cbl following tetramer disassembly, and it is tempting to suggest that formation of bis-His ligated Cbl in some CarHs might be the first step of such a recovery pathway. In this regard, it is interesting to note that His132 is strictly conserved in thermophilic bacteria (Extended Data Fig. 3b), where perhaps the bis-His ligated Cbl is an adaption to elevated temperatures. Although use of AdoCbl as a light sensor comes at a price, it appears that the physiological benefits make this repurposing worthwhile.

Altogether, our results provide fundamental insight into a new mode of light-dependent gene regulation and reveal an exquisite example of cofactor and protein domain repurposing. The structures furthermore provide a basis for deployment of the modular CarH photoreceptors, in which the light-sensing and DNA-binding activities rest on different domains, for engineering light-modulated transcriptional control or protein–protein interactions.

Note added in proof: A paper describing a detailed photochemical mechanism for CarH based on time-resolved spectroscopic data has just been published40.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.
11. Elias-Arnanz, M., Padmanabhan, S. & Murillo, F. J. Light-dependent gene regulation in nonphototrophic bacteria. *Curr. Opin. Microbiol.* **14**, 128–135 (2011).

12. Takano, H. et al. LdrF, a cAMP receptor protein/FNR family transcriptional regulator, serves as a positive regulator for the light-inducible gene cluster in the megaplasmid of *Thermus thermophilus*. *Microbiology* **160**, 2650–2660 (2014).

13. Frey, P. A. in *Comprehensive Natural Products II Chemistry and Biology Vol. 7* (eds Mander L. & Liu, H.-W.) 501–546 (Elsevier, 2010).

14. Banerjee, R. Radical carbon skeleton rearrangements: catalysis by coenzyme B_{12}-dependent mutases. *Chem. Rev.* **103**, 2083–2094 (2003).

15. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27**, 145–163 (2003).

16. Navarro-Avilés, G. et al. Structural basis for operator and antitarget recognition by *Mycobacterium xanthus* CarR repressor. *Mol. Microbiol.* **63**, 980–994 (2007).

17. Drennan, C. L., Huang, S., Drummond, J. T., Matthews, R. G. & Ludwig, M. L. How a protein binds B_{12}: a 3.0 Å X-ray structure of B_{12}-binding domains of methionine synthase. *Science* **266**, 1669–1674 (1994).

18. Diez, A. I. et al. Analytical ultracentrifugation studies of oligomerization and DNA-binding of TtCarH, a *Thermus thermophilus* coenzyme B_{12}-based photosensory regulator. *Eur. Biophys. J.* **42**, 463–476 (2013).

19. Cheng, Z., Li, K. R., Hammad, L. A., Karty, J. A. & Bauer, C. E. Vitamin B_{12} regulates photosystem gene expression via the Crf1 antitargetressor AerR in *Rhodobacter capsulatus*. *Mol. Microbiol.* **91**, 649–664 (2014).

20. Marques, H. M., Marsh, J. H., Mellor, J. R. & Munro, O. Q. The coordination of imidazole and its derivatives by aquocobalamin. *Inorg. Chim. Acta* **170**, 259–269 (1990).

21. Linden, H. & Macino, G. White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in Neurospora crassa. *EMBO J.* **16**, 98–109 (1997).

22. Froehlich, A. C., Liu, Y., Loros, J. J. & Dunlap, J. C. White collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science* **297**, 815–819 (2002).

23. Nash, A. I. et al. Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl Acad. Sci. USA* **108**, 9449–9454 (2011).

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

25. Losi, A. & Gärtner, W. The evolution of flavin-binding photoreceptors: an ancient photosensory LOV protein. *Nature* **49–72 (2012).

26. Conrad, K. S., Manahan, C. C. & Crane, B. R. Photochemistry of flavoprotein light sensors. *Nature Chem. Biol.* **10**, 801–809 (2014).

27. Herrou, J. & Crosson, S. Function, structure and mechanism of bacterial photosensory LOV proteins. *Nature Rev. Microbiol.* **9**, 713–723 (2011).

28. Jenkins, G. I. The UV-B photoreceptor UVR8: from structure to physiology. *Annu. Rev. Plant Biol.* **649–664 (2011).

29. Burgie, E. S. & Vierstra, R. D. Phytochromes: an atomic perspective on photoactivation and signaling. *Plant Physiol.* **26**, 4568–4583 (2014).

30. Losi, A. & Gärtner, W. The evolution of flavin-binding photoreceptors: an ancient photosensory LOV protein. *Annu. Rev. Plant Biol.* **49–72 (2012).

31. Nash, A. I. et al. Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl Acad. Sci. USA* **108**, 9449–9454 (2011).

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

33. Jost, M., Simpson, J. H. & Drennan, C. L. The transcription factor CarH safeguards use of adenosylcobalamin as a light sensor by altering the photosynthesis products. *Biochemistry* **54**, 3231–3234 (2015).

34. Girvan, H. M. & Munro, A. W. Heme sensor proteins. *J. Biol. Chem. 288*, 13194–13203 (2013).

35. Matthews, R. G. Cobalamin-dependent methionine synthases. *Acc. Chem. Res.* **34**, 681–689 (2001).

36. Jarrett, J. T. et al. Protein radical cage slows photolysis of methylenylcobalamin in methionine synthase from *Escherichia coli*. *Bioorg. Med. Chem.* **4**, 1237–1246 (1996).

37. Bandarian, V., Ludwig, M. L. & Matthews, R. G. Factors modulating conformational equilibria in large modular proteins: a case study with cobalamin-dependent methionine synthase. *Proc. Natl Acad. Sci. USA* **100**, 8156–8163 (2003).

38. Heldwein, E. E. Z. & Brennan, R. G. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409**, 378–382 (2001).

39. Watanabe, S., Kita, A., Kobayashi, K. & Miki, K. Crystal structure of the [2Fe-2S] oxidative-stress sensor SoxR bound to DNA. *Proc. Natl Acad. Sci. USA* **105**, 4121–4126 (2008).

40. Kutta, R. J. et al. The photochemical mechanism of a B_{12}-dependent photosensory protein. *Nat. Commun.* **6**, 7907 (2015).

Acknowledgements We thank F. Murillo for his contributions and support, J. A. Madrid for technical assistance (Universidad de Murcia), and A. Cohen (SSRL) and N. Ando (MIT) for help with single crystal UV–vis spectroscopy. This work was supported in part by the National Institutes of Health (NIH, grant GM069857 to C.L.D.), the Ministerio de Economía y Competitividad, Spain (grants BFU2012-40184-C02-01 co-financed by FEDER (Fondo Europeo de Desarrollo Regional) funds to M.E.-A.; BFU2012-40184-C02-02 to S.P.; PhD fellowship to J.F.-Z.), a CSIC-JAE-Predoc (Spain) fellowship (to J.M.O.-G.), and an MIT Poitras pre-doctoral fellowship (to M.I.). C.L.D. is a Howard Hughes Medical Institute Investigator. This work used the Northeastern Collaborative Access Team beamlines of the Advanced Photon Source, which are supported by National Institute of General Medical Sciences (NIGMS) grant P41GM103403 and the NIH. The Advanced Photon Source is a US Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-06CH11357. Use of the Stanford Synchrotron Radiation Lightsource is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE-AC02-76SF00515, the Department of Energy Office of Biological and Environmental Research, and by the NIH and NIGMS (including P41GM103393).

Author Contributions M.J. performed crystallographic experiments, J.F.-Z., M.C.P., and J.M.O.-G. performed in vitro mutant analyses, and G.K. and P.Y.-T.C. assisted with crystal structure refinement. M.J., S.P., M.E.-A., and C.L.D. designed experiments, analysed the data, and wrote the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5CB8 (AdoCbl-bound CarH crystal form 2), 5CD0 (AdoCbl-bound CarH crystal form 3), 5C8E (AdoCbl- and DNA-bound CarH), and 5CBF (light-exposed CarH). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.P. (padhu@iqfr.csic.es), M.E.-A. (melias@um.es) or C.L.D. (cdrennan@mit.edu).
METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Protein constructs. Cloning of the PET15b-CarH construct encoding for Thermus thermophilus CarH with an N-terminal His<sub>6</sub>-tag was described previously<sup>a</sup>. The H132A mutation was introduced into PET15b-CarH using QuikChange PCR mutagenesis (Stratagene) with Pfu Turbo DNA polymerase. All other mutants were obtained by gene synthesis (Genscript) with appropriate 5’ and 3’ restriction sites for cloning into the PET15b expression vector.

Protein purification. WT CarH and mutants were purified as described previously<sup>b</sup>. A slightly modified protocol was used for His<sub>6</sub>-tagged CarH for crystallization. After expression and affinity chromatography, performed as described previously<sup>c</sup>, a threefold molar excess of AdoCbl (Sigma) was added and the mixture was incubated at ice for 1 h. All subsequent handling was performed in a dark room under red light. The protein solution was applied to a HiLoad 26/60 Superdex 200 size exclusion column (GE Healthcare) pre-equilibrated with CarH buffer (0.1 M NaCl, 0.05 M Tris-HCl, pH 8). Under these conditions, tetrameric AdoCbl-bound CarH eluted as a single peak, separate from residual amounts of monomeric CarH. Fractions containing AdoCbl-bound CarH were combined and concentrated to about 8 mg ml<sup>−1</sup>. The protein underwent proteolysis at the linker region between the two domains. AdoCbl-bound CarH eluted as a single peak, separate from residual amounts of CarH but no DNA. Crystals were transferred in three steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 20% (v/v) glycerol, soaked in that solution for 10 s, and then flash-frozen in liquid nitrogen.

A third crystal form of AdoCbl-bound CarH containing full-length protein was obtained by the sitting drop vapour diffusion technique at 25 °C. An aliquot (0.15 μl) of a protein solution (6 mg ml<sup>−1</sup> AdoCbl-bound CarH in CarH buffer, supplemented with 70 μM of a 28-bp DNA segment) was mixed with 0.15 μl of a precipitant solution (20% (w/v) PEG 3350, 0.2 M KCl) using a Phoenix liquid handling robot (Art Robbins Instruments). The drop was equilibrated against 70 μl of the precipitant solution. Rectangular crystals appeared within 6 months. Again, the protein underwent proteolysis and the crystals only consisted of the C-terminal light-sensing domains. Crystals were transferred in two steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 20% (v/v) glycerol, soaked in that solution for 5 s, and then flash-frozen in liquid nitrogen.

Light-exposed CarH was crystallized by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 μl) of a protein solution (4.5 mg ml<sup>−1</sup> light-exposed CarH in CarH buffer) was mixed with 1 μl of a precipitant solution (3.4 M NaCl, 0.1 M Bis-Tris pH 6) on a glass cover slip. The cover slip was sealed with glue over a reservoir containing 500 μl of the precipitant solution. Octahedral crystals appeared within 8 months. Crystals were transferred in three steps of increasing glycerol concentration into a cryogenic solution containing the precipitant supplemented with 18% (v/v) glycerol, incubated in that solution for 10 s, and then flash-frozen in liquid nitrogen.

CarH bound to both AdoCbl and to a 26-bp DNA segment was crystallized by the hanging drop vapour diffusion technique at 25 °C. An aliquot (1 μl) of a protein solution (6 mg ml<sup>−1</sup> AdoCbl-bound CarH in CarH buffer, supplemented with 67.5 μM of a 26-bp DNA segment, 1.5-fold molar excess) was mixed with 1 μl of a precipitant solution (16% PEG 3350, 0.2 M l-proline, 0.1 M HEPES pH 7.5) on a glass cover slip. The cover slip was sealed with glue over a reservoir containing 500 μl of the precipitant solution. Tetragonal bipyrimal crystals appeared within 3 weeks. Crystals were transferred in three steps of increasing PEG 400 concentration into a cryogenic solution containing the precipitant supplemented with 15% (w/v) PEG 400, incubated in that solution for 20 s, and then flash-frozen in liquid nitrogen.

Data collection and processing. All data were collected at the Advanced Photon Source (Argonne, Illinois, USA) at beamline 24-ID-C using a Pilatus 6M pixel detector at a temperature of 100 K. Crystals of AdoCbl-bound CarH crystal form 1 belonged to space group P<sub>4</sub>1<sub>2</sub>1<sub>2</sub>1. An initial AdoCbl-bound CarH crystal was used for a fluorescence scan to determine the C<sub>e</sub> peak wavelength for anomalous data collection. Another crystal was then used for collection of both native data and anomalous peak data. Native data were collected in a single wedge of 75° at a wavelength of 0.9792 Å (12,662 eV). The crystal was displaced continuously along its major monoclinic axis during data collection. Anomalous peak data were collected in a single wedge of 345° at a wavelength of 1.6039 Å (7,730 eV). The crystal was aligned using a mini-x goniometer such that Bijvoet mates were recorded on the same frame.

All other data except for iodine anomalous data and native data of light-exposed CarH (see below) were collected at a wavelength of 0.9795 Å (12,658 eV). Crystals of AdoCbl-bound CarH crystal form 2 belong to space group P<sub>2</sub>1<sub>2</sub>1<sub>2</sub>. Data were collected in a single wedge of 100°. Crystals of AdoCbl-bound CarH crystal form 3 belong to space group P<sub>1</sub>. Data were collected in a single wedge of 270° and the crystal was displaced continuously along its major monoclinic axis during data collection. Crystals of light-exposed CarH belong to space group I<sub>4</sub>2<sub>d</sub> and data were collected in a single wedge of 180°.
were collected at a wavelength of 0.9791 Å (12.663 eV) in a single wedge of 180°. Crystals of DNA-bound CarH both with and without the iodide label belong to space group P2₁2₁2₁. Data for crystals with unlabelled DNA were collected in a single wedge of 180°. Data for crystals of CarH in complex with iodine-labelled DNA were collected at a wavelength of 1.7365 Å (7.140 eV) in a single wedge of 200° and the crystal was displaced continuously along its major macroscopic axis during data collection.

Data for the AdoCbl-bound CarH (crystal form 1) Co peak data set were integrated in HKL2000 and scaled in Scalepack. Data for all other data sets were integrated in XDS and scaled in scxscale. Data collection statistics are summarized in Extended Data Table 1.

**Structure building and refinement.** The structure of AdoCbl-bound CarH in crystal form 1 (space group P4₁2₁2₁) was determined to 2.80 Å resolution using single-wavelength anomalous diffraction. Positions of two cobalt sites, corresponding to two CarH protomers in the asymmetric unit, were located using ShelXD in the HKL2MAP shell and refined using SHARP/autoSHARP. The initial overall figure of merit (acentric) was calculated by SHAIR to be 0.43 to 5.1 Å resolution. Experimental maps from the SHARP output, solvent flattening using SOLomitON52 and extended to 3.3 Å resolution, were of sufficient quality to place two copies of the Cbl-binding domain of MethH (PDBe accession number 1BMT, residues 745–868), eight additional helices, and AdoCbl in the electron density. This initial model was used to better define solvent boundaries in another round of solvent flattening of SOLomitON. Using the resulting electron density, loop regions were modified and side chains with visible electron density were added. A near-complete model of AdoCbl-bound CarH (containing 374 amino-acid residues and bound AdoCbl) was then used for rigid body refinement in Phenix against the native AdoCbl-bound CarH data set (crystal form 1) using data from 100 to 2.80 Å resolution. The resulting R-factors were 42.0% and 44.1% for the working and the free R-factor, respectively. The model was refined by manual adjustment in Coot until rigid body refinement in Phenix yielded R-factors of 30.8% and 34.7% for the working and the free R-factor, respectively. Subsequent cycles of refinement included positional refinement with non-crystallographic symmetry restraints and individual B-factor refinement in Phenix until the R-factors were 20.9% and 24.2% for the working and the free R-factor, respectively. This model was not refined to completion. The near-complete model was used to determine the structures of AdoCbl-bound CarH in crystal form 2 (space group P2₁2₁2₁) and crystal form 3 (space group P1), which are of higher resolution (crystal form 2) or contain the full-length protein (crystal form 3).

The structure of AdoCbl-bound CarH in crystal form 2 was determined to 2.15 Å resolution by molecular replacement in Phaser. The structure in crystal form 2 contains four CarH protomers in the asymmetric unit, corresponding to a tetramer. After molecular replacement, ten cycles of simulated annealing refinement were performed in Phenix to remove model bias. The model was then refined by iterative cycles of manual adjustment in Coot and refinement in Phenix. For EMSAs, a 20 μl reaction volume containing the DNA probe (1.2 nM, approximated by 13000 counts per minute) and protein with a fivefold excess of AdoCbl in 0.1 M KCl, 0.025 M Tris-HCl, pH 8, 1 mM DTT, 10% (v/v) glycerol, 200 μg/ml BSA, and 1 μg of sheared salmon sperm DNA as non-specific competitor was incubated for 30 min at 65°C (177-ppm probe) or 30°C (shorter probes). They were then loaded onto 6% native polyacrylamide gels (37:51:acrylamide: bisacrylamide) pre-run for 98.1%, 1.9%, and 0.0% of residues are in the favoured, allowed, and disallowed regions, respectively; for AdoCbl-bound CarH (crystal form 2), 97.8%, 2.2%, and 0.0% of residues are in the favoured, allowed, and disallowed regions, respectively; and for AdoCbl- and DNA-bound CarH, 97.1%, 2.7%, and 0.2% of residues are in the favoured, allowed, and disallowed regions, respectively. The larger number of residues in the disallowed region of the Ramachandran plot of DNA-bound CarH is due to the modest resolution of the structure. Figures were generated using PyMOL. Interfaces between subunits were analysed using the ‘Protein interfaces, surfaces and assemblies’ service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Crystallography software packages were compiled by SBGrid. All DNA binding assays were repeated three to five times for each experimental condition. EMSAs were performed in the dark as described previously. A 177-ppm DNA probe PCR-amplified using primers with one 5′-end 32P-labelled with T4 polynucleotide kinase (T4PNK, Takara) before the PCR or shorter HPCL-purified synthetic probes (Biologo) were used in the EMSAs. With the latter, one strand was 32P-labelled at the 5′-end with T4PNK and then mixed with a twofold excess of the unlabelled complementary strand to ensure that all of the labelled strand was present as double-stranded probe. The strand mixture was incubated at 100°C for 2 min and then slowly left to cool down for hybridization. For EMSAs, a 20 μl reaction volume containing the DNA probe (1.2 nM, approximated by 13000 counts per minute) and protein with a fivefold excess of AdoCbl in 0.1 M KCl, 0.025 M Tris-HCl, pH 8, 1 mM DTT, 10% (v/v) glycerol, 200 μg/ml BSA, and 1 μg of sheared salmon sperm DNA as non-specific competitor was incubated for 30 min at 65°C (177-ppm probe) or 30°C (shorter probes). They were then loaded onto 6% native polyacrylamide gels (37:51:acrylamide: bisacrylamide) pre-run for 200 μm.
30 min in 0.5× TBE buffer (0.045 M Tris base, 0.045 M boric acid, 1 mM EDTA) and subjected to electrophoresis for 1.5 h at 200 V, 10 °C. Gels were vacuum-dried and analysed by autoradiography. Autoradiograms were scanned using an Image Scanner II imager with LabScan 5.0 software (GE Healthcare). Band intensities were quantified using ImageJ (NIH) with those of free DNA used to estimate the fraction bound, which was fitted to the three-parameter Hill equation using Sigmaplot (Systat Software) to estimate $K_{d}$, the apparent equilibrium dissociation constant equivalent to the protein concentration for half-maximal binding, and $n$, the Hill coefficient. The latter, for example expected to be 2 for dimer or 4 for tetramer DNA-binding models, can vary owing to cooperativity effects, contributions from monomer-tetramer equilibria, and/or deviations from true equilibrium.

**DNase I and hydroxyl radical footprinting.** DNase I and hydroxyl radical footprinting analyses were performed under solution conditions similar to EMSA using previously described protocols. A 130-bp CarH operator-promoter DNA probe (Extended Data Fig. 5a) was 32P-radiolabelled at the 5′ end of its sense or anti-sense strand by PCR using appropriately labelled primers, as described above. For DNase I footprinting, 20 μl of 5′-radiolabelled DNA probe (~20,000 counts per minute) with 800 nM CarH and fivefold excess of AdoCbl in EMSA buffer lacking glycerol and with 0.01 M MgCl₂ were incubated for 30 min at 37 °C, then treated with 0.07 units of DNase I for 2 min and finally quenched with 0.025 M EDTA. For hydroxyl radical footprinting, samples (as for DNase I footprints but without MgCl₂) were treated with 2 μl each of freshly prepared Fe(II)-EDTA solution (1 mM ammonium iron (II) sulfate, 2 mM EDTA), 0.01 M sodium ascorbate, and 0.6% hydroxyperoxide for 4 min at 25 °C. The reaction was stopped with 2 μl each of 0.1 M thiourea and 0.5 M EDTA (pH 8). Footprinting reactions were done under dim light and, after quenching, under normal light. DNA from each sample was ethanol precipitated, washed twice with 70% ethanol, dried, and resuspended in formamide loading buffer. The 5 μl samples were heated at 95 °C for 3 min and loaded onto a 6% polyacrylamide-8 M urea sequencing gel together with G + A chemical sequencing ladders. Gels were vacuum-dried and analysed by autoradiography, and the bands quantitated using GelAnalyzer 2010a (http://www.gelanalyzer.com). Each experiment was repeated at least three times.

**Analytical SEC.** Analytical SEC for all CarH mutants except for H132A CarH was performed using an ÄKTA basic unit and a Superose 200 analytical SEC column (GE Healthcare). The calibration curve was $log(M_r) = 7.885 - 0.221V_e$, where $M_r$ is the apparent molecular mass and $V_e$ is the elution volume. Pure protein (100 μl, 50–100 μM) was injected onto the column and measured at 280 nm, 522 nm, and 405 nm with deuterium and halogen lamps, UV solarization-resistant optical fibres, reflective Newport Schwarchild objectives, and an Ocean Optics QE65000 Spectrum Analyzer. Spectra were acquired as 50 averages with an integration time of 0.03 s and a boxcar width of 3. A crystal of AdoCbl-bound CarH was cryoprotected, transferred to a nylon fibre loop, and frozen in liquid nitrogen as described above. A background spectrum was acquired on a region of the fibre loop containing just cryoprotectant. A sample spectrum was then acquired on the crystal.

41. Barker, H. A. et al. Isolation and Proteins of crystalline cobamide coenzymes containing benzimidazole or 5,6-dimethylbenzimidazole. J. Biol. Chem. 235, 480–488 (1960).

42. Hill, J. A., Williams, R. J. P. & Pratt, J. M. Chemistry of vitamin B₁₂ derivatives. Biochemistry 6, 2178–2189 (1967).

43. Hill, J. A., Williams, R. J. P. & Pratt, J. M. Chemistry of vitamin B₁₂. Part I. The valency and spectrum of the coenzyme. J. Chem. Soc. 5149–5153 (1964).

44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).

45. Kabsch, W. Xds. Acta Crystallogr. D 66, 125–130 (2010).

46. Sheldrick, G. M. & Schneider, T. R. Substructure solution with SHELXD. Acta Crystallogr. D 58, 1772–1779 (2002).

47. Schneider, T. R. & Pape, T. HKL2MAP: a graphical user interface for macromolecular phasing with SHELXL programs. J. Appl. Cryst. 37, 843–844 (2004).

48. Vontz, C., Blanc, E., Roversi, P. & Brignone, G. Automated structure solution with autoSHARP. Methods Mol. Biol. 364, 215–230 (2007).

49. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).

50. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Cryst. 40, 658–674 (2007).

51. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr. D 62, 439–450 (2006).

52. van Dijl, M. & Bonvin, A. M. J. J. 3D-DART: a DNA structure modelling server. Nucleic Acids Res. 37, W235–W239 (2009).

53. Ten Eyck, L. F. Fast Fourier transform calculation of electron density maps. Methods Enzymol. 115, 324–337 (1985).

54. Winn, M. D. et al. Overview of the CCP4 suite and current developments. Acta Crystallogr. D 67, 235–242 (2011).

55. Brnger, A. T. Version 1.2 of the Crystallography and NMR system. Nature Protocols 2, 2728–2733 (2007).

56. Chen, V. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D 66, 12–21 (2010).

57. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291 (1993).

58. Schrödinger, L. C. L. The PyMOL Molecular Graphics System v.1.3r1 (Schrödinger, LLC, 2010).

59. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Appl. Cryst. 37, 772–779 (2004).

60. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815 (1993).

61. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Appl. Cryst. 36, 763–768 (2003).

62. Schrödinger, L. C. L. The PyMOL Molecular Graphics System v.1.3r1 (Schrödinger, LLC, 2010).

63. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 772–779 (2007).

64. Morin, A. et al. Collaboration gets the most out of software. eLife 2, e01456 (2013).

65. Jain, S. S. & Tullius, T. D. Footprinting protein-DNA complexes using the hydroxyl radical. Nature Protocols 3, 1092–1100 (2008).

66. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, W320–W324 (2014).
Extended Data Figure 1 | CarH crystals contain intact AdoCbl.  

**a**, UV–vis spectra obtained from AdoCbl-bound CarH crystals at $T = 100$ K (red trace) or AdoCbl-bound CarH in solution at $T = 298$ K (black trace) exhibit good qualitative agreement and similar features, including a peak centred around 540 nm with a shoulder around 560 nm. Because many band intensities are orientation-dependent and the crystal spectrum changes with orientation but molecules are rotationally averaged in solution, quantitative comparison of the spectra is difficult. Note also that individual bands appear sharper in the crystal spectrum because the molecules have fewer rotational degrees of freedom and because fewer vibrational states are populated at $T = 100$ K.

**b**, Simulated annealing composite omit electron density (2.15 Å resolution) contoured around AdoCbl at 1.0σ (grey). The electron density covers the entire AdoCbl molecule including the Co–C bond, indicating that the Co–C bond remained intact during crystallization and data collection. AdoCbl is shown in stick representation with Cbl carbons in pink and 5′-dAdo group carbons in cyan. Co is shown as a purple sphere. The Co-coordinating His177 is shown in sticks with carbons in green. CarH is shown in ribbons with the helix bundle in yellow and Cbl-binding domain in green.
Extended Data Figure 2 | The CarH DNA-binding domain is flexible in the absence of DNA.  

**a**, Overlay of five CarH protomers, including the protomer shown in Fig. 2a, highlighting flexibility of DNA-binding domains. Structures are aligned by the Cbl-binding domains (green) and helix bundles (yellow) and shown in the same orientation as Fig. 2a. DNA-binding domains are coloured in dark cyan, light cyan, dark blue, black, and grey. AdoCbl is shown with Cbl carbons in pink, 5’-dAdo group carbons in cyan, and cobalt in purple. **b–e**, Individual CarH protomers shown side by side. Orientation and colouring as in **a**.
| Mutation  | Description | Oligomerization | DNA binding |
|-----------|-------------|----------------|-------------|
| WT        |             | ++++           | ++++        |
| Q25A      |             | ++++           | +           |
| R29A      |             | ++++           | -           |
| Y30A      | DNA contact | ++++           | +*          |
| H42A      |             | ++++           | +*          |
| R43A      |             | ++++           | -           |
| W131A     |             | +             | +           |
| W131F     | 5-dAdo      | ++            | +++         |
| E141A     |             | ++            | +           |
| H142A     |             | -             | -           |
| D201R     | Dimer       | ++            | +           |
| R176D/D201R | Dimer      | +++           | +           |
| R176E/D201R | Dimer       | +++           | +           |
| G160Q     |             | ++*           | +*          |
| G192Q     |             | ++*           | +*          |

**Thermus thermophilus**

**Thermus aquaticus**

**Thermus fилiformis**

**Neisseria meningitidis**

**Deinococcus marinopiscis**

**Roseiflexus castenholzi**

**Bacillus megaterium**

**Sorangium cellulorum**

**Myxococcus xanthus**

**Thermus thermophilus**

**Thermus aquaticus**

**Thermus fилiformis**

**Neisseria meningitidis**

**Deinococcus marinopiscis**

**Roseiflexus castenholzi**

**Bacillus megaterium**

**Sorangium cellulorum**

**Myxococcus xanthus**
Extended Data Figure 3 | CarH mutant analysis and multiple sequence alignment. a, Results summary for in vitro CarH mutant analysis. Table footnotes are as follows: * oligomerization was probed by SEC and DNA binding by gel shift analysis. † Y30A, H42A: weakened binding at 100 nM protein. ‡ G160Q, G192Q: dimer, no tetramer. § G160Q, G192Q: binds with reduced affinity and cooperativity and as a higher mobility (smaller size) complex. b, Alignment of CarH sequences from different bacterial species. Sequence identity is shown in white font with red background, sequence similarity in red font. Coloured triangles highlight functionally important positions, with filled triangles indicating residues analysed by mutagenesis in this study and empty triangles indicating residues not analysed by mutagenesis. Mutating the highly conserved His177 of the Cbl-binding motif, the lower axial ligand of bound AdoCbl, has previously been shown to impair AdoCbl binding and tetramerization. Colouring is as follows: hydrogen bonds/ionic interactions to DNA, orange; contact to 5′-dAdo, green; histidines coordinating Cbl (His132 only coordinates after light exposure), red; hydrogen bonds/ionic interactions at dimer interface, black; hydrogen bonds/ionic interactions as well as Gly160 and Gly192 at the dimer–dimer interface, cyan. Residues involved in more than one type of interaction are coloured half/half. Residues at protein interfaces are less well conserved than other functionally important residues, probably because compensatory mutations and local structural deformations are possible. Note, however, that the *T. thermophilus* Arg176–Asp201 pair observed in our structure is inverted in *Myxococcus xanthus*, suggesting that the interaction is conserved. Alignment generated using ESPript64.
Extended Data Figure 4 | Characterization of CarH mutants affecting oligomerization state. a–f, SEC traces (Superdex 200 analytical SEC column) of CarH carrying mutations (a, b) near the 5′-dAdo group; c, d, at the head-to-tail dimer interface; and e, f, at the dimer–dimer interface. Shown are traces of mutants incubated with AdoCbl in the dark (top panels) and after light exposure (bottom panels). In all panels, both absorbance $A_{280}$ nm (tracking protein) and $A_{522}$ nm (tracking Cbl) traces are shown. Molecular masses are calculated from the observed elution volumes as described in Methods, and are consistent with a tetrameric species (137 kDa), a dimeric species (89 kDa), and a monomeric species (39 kDa). Notably, mutant CarH proteins that do not tetramerize in the presence of AdoCbl (D201R, H142A) also do not appear to bind AdoCbl (see 522 nm traces of dark samples). This finding is consistent with previous studies that show cooperativity of AdoCbl binding and tetramerization, a feature that does not hold for other forms of Cbl (methylcobalamin, CNCbl and Cbl$^\beta$). Both of these mutant proteins can still bind Cbl (see 522 nm traces of light-exposed samples), which further suggests that these mutants are properly folded and that the lack of AdoCbl binding stems from inability to oligomerize. Although the degree of tetramerization of CarH mutant proteins in the dark varied, all of these mutant proteins form Cbl-bound monomers after light exposure. g, DNA-binding capacity of WT and mutant proteins (800 nM) as determined by EMSAs after incubation with AdoCbl (4 $\mu$M) in the dark. h, EMSA data for WT CarH and the G160Q and G192Q mutants fit to the Hill equation, as described in Methods. $K_d$ (in nM) and Hill coefficients from the fits are, respectively, (67 ± 2) and (5.1 ± 0.7) for WT CarH, (111 ± 18) and (3.0 ± 0.2) for G160Q CarH, and (253 ± 17) and (2.5 ± 0.3) for G192Q CarH. The data shown are the mean values and standard errors of three to five repeat experiments.
Extended Data Figure 5 | Identification and validation of CarH operator sequence by EMSAs and footprinting. a. Location of CarH operator in the intergenic region between carH and the carotenogenic crtB of the T. thermophilus genome. Structural and biochemical data are mapped onto the sequence. Three 11-bp CarH binding sites are shown in cyan font and the promoter −35 element is highlighted with a red box. Nucleotides protected from hydroxyl radical cleavage are indicated with bullets. The −42-nucleotide DNase I footprint on the sense strand is shown above the sequence and that of the antisense strand has been omitted for clarity. Nucleotide numbering on the sense strand is relative to the carH transcription start site (underlined, +1). To identify suitable DNA constructs for crystallization, operator sequences were systematically trimmed around a 40-bp segment, as indicated by the black bars, and binding was assessed by EMSAs (shown in b). The sequences of two 26-bp DNA segments used for co-crystallization are also shown. The blunt-ended 26-bp segment was used for determination of the CarH–DNA structure. The second 26-bp segment contained one-nucleotide 3′-overhang and 5-ido-deoxycytidine in position −25 (red) and was used to validate the mode of DNA binding. b. Binding of CarH (800 nM) to DNA segments of different lengths after incubation with AdoCbl (4 μM) in the dark. Substantial DNA binding was observed for a probe as small as 30-bp. c. DNase I and hydroxyl radical footprints of CarH on a 130-bp operator DNA segment. Disappearance of bands in the presence of CarH indicates protection from cleavage. Protected regions are marked on the side and were mapped onto the operator sequence using G + A chemical sequencing experiments performed in parallel. d, e. CarH binding to 40-bp operators carrying mutations. d. Sequences of tested operator variants. WT operator sequence shown at the top and bottom, with repeat sequences that CarH recognizes shown in cyan; 6-bp stretch contacted by CarH recognition helix is boxed. Mutations are as follows: Mut1–7: single (1–3), pairwise (4–6), and triple (7) mutations of AC to GT (positions 8/9); Mut8–14: single (8–10), double (11–13), and triple (14) mutations of (A/C)T to GC (positions 4/5); Mut15–18: pairwise (15–17) and triple (18) mutations of (A/G)A) to TT (positions 1/2). e. EMSAs with WT CarH (800 nM) and each of the 40-bp operator variants after incubation with AdoCbl (4 μM) in the dark. Note that two additional lower mobility complexes are observed, most apparent with the WT operator and its variants with comparable binding. The origin of these complexes is unknown, but they probably arise from oligomeric equilibria and residual amounts of light-exposed protein in the sample.
Extended Data Figure 6 | CarH DNA binding, conformational changes upon binding, and comparison with BmrR. a, $2F_o - F_c$ omit electron density (3.89 Å resolution) for DNA-bound CarH, calculated after performing full refinement of the model with DNA omitted and contoured at 1.0σ. DNA is shown with carbons in yellow and recognition helix of a CarH DNA-binding domain with carbons in cyan. b, c, Validation of DNA-binding mode using heavy-atom-derivatized DNA segments. CarH was crystallized with a DNA segment containing 5-iodo-deoxycytidine in position −25 of the sense strand. Shown is the resulting anomalous difference density (purple mesh), contoured at 6σ, for both CarH–DNA complexes in the asymmetric unit, with peaks directly adjacent to the C5 atom of deoxycytidine in position −25. d, Chemical structure of 5-iodo-deoxycytidine. e, Comparison of CarH before and after DNA binding, revealing rearrangement of DNA-binding domains. CarH before DNA binding is shown with helix bundles and Cbl-binding domains in green and DNA-binding domains in pink. CarH bound to DNA is shown with helix bundles and Cbl-binding domains in green and DNA-binding domains in cyan. The fourth DNA-binding domain of DNA-bound CarH is disordered and not modelled. DNA is shown in yellow. AdoCbl is shown with Cbl carbons in pink and 5'-dAdo group carbons in cyan. f, Contacts between residues in neighbouring DNA-binding domains, coloured by domain. Each interface between two DNA-binding domains buries 280 Å² of surface from solvent on each DNA-binding domain. Interactions of Arg72 to Tyr7 and Glu11 are indicated by black dashed lines. Colouring as in e. g, h, Models of individual CarH head-to-tail dimers bound to DNA. g, Head-to-tail dimer contributing the middle of the three DNA-binding domains, coloured by domain with DNA-binding domain in cyan, helix bundles in yellow, and Cbl-binding domains in green. The DNA-binding domain of the second protomer (right) is disordered and not modelled. DNA and AdoCbl are shown as in e. h, Head-to-tail dimer contributing the flanking DNA-binding domains. Helix bundles and Cbl-binding domains are shown in grey, remaining colouring as in e. i, BmrR bound to DNA (PDB accession number 1EXJ(8)). A BmrR dimer is shown in ribbon representation in orange and red. DNA is shown in yellow. BmrR binds as a dimer to a palindromic sequence and distorts the DNA double strand from its ideal conformation.
Extended Data Figure 7 | In vitro characterization of CarH DNA binding mutants. a, b, SEC traces (Superdex 200 analytical SEC column) of CarH carrying mutations in the DNA-binding domain. Shown are traces of mutants (a) incubated with AdoCbl in the dark and (b) after light exposure. In all panels, both $A_{280}$ nm (tracking protein) and $A_{522}$ nm (tracking Cbl) traces are shown. Molecular masses are calculated from the observed elution volumes as described in Methods, and are consistent with a tetrameric species (137 kDa) and a monomeric species (39 kDa). c, DNA-binding capacity of mutants (800 nM) as determined by EMSAs after incubation with AdoCbl (4 μM) in the dark.
Extended Data Figure 8 | Light-exposed CarH has bis-His ligated Cbl.
a, Structure of light-exposed CarH including the DNA-binding domain (cyan) and other domains coloured as in Fig. 5a. b, Close-up view of the Cbl in light-exposed CarH, with both coordinating His side chains shown in sticks. Simulated annealing composite omit electron density (2.65 Å resolution) is shown in blue, contoured at 1.0σ. c, UV–vis spectra of light-exposed WT CarH (black) and H132A CarH (red) exhibit pronounced differences, indicating that the bis-His ligation is also formed in solution. d, UV–vis spectra of free OHCbl (50 μM) with increasing imidazole concentration. The spectrum of bis-imidazole ligated Cbl (black, Cbl with 400 mM imidazole contains 60% bis-imidazole ligated Cbl and 40% Cbl with dimethylbenzimidazole and imidazole as ligands20) resembles that of light-exposed WT CarH, whereas the spectrum of free OHCbl (pink) resembles that of light-exposed H132A CarH. Note that the latter two are expected to be slightly different because free OHCbl contains a dimethylbenzimidazole group as the lower axial ligand, whereas light-exposed H132A CarH contains a histidine imidazole as the lower axial ligand. Experimental conditions chosen were similar to those reported elsewhere20. e, UV–vis spectra of AdoCbl-bound WT CarH (black) and H132A CarH (red) are virtually identical, suggesting that the mode of AdoCbl binding is unchanged, as is expected from the structure. f, g, Size-exclusion chromatograms (Superose 6 10/300 GL column) of AdoCbl-bound and light-exposed (f) WT CarH and (g) H132A CarH, demonstrating that H132A CarH, like WT CarH, forms a tetramer in the dark and undergoes light-dependent tetramer disassembly.
Extended Data Figure 9 | Disruption of bis-His ligation by H132A mutation facilitates Cbl dissociation after photolysis. a, b. WT and H132A CarH were exposed to light, rendering them monomeric, and then incubated with free AdoCbl at the indicated temperatures and periods. For AdoCbl to bind to CarH and induce tetramerization, the photolysed Cbl has to dissociate from the protein first. Thus, the extent of tetramer formation, as assessed by SEC (Superose 6 10/300 GL column), is indicative of the affinity of the protein for photolysed Cbl. That is, lack of tetramer formation in the presence of fresh AdoCbl indicates that the photolysed Cbl is still bound to the protein. The observed differential in tetramer formation between H132A and WT CarH is substantial: WT CarH retains its photolysed Cbl, showing only a small amount of tetramer formation, whereas H132A CarH loses its photolysed Cbl, reforming tetramers upon AdoCbl addition. c. ESI–TOF mass spectra of WT and H132A CarH after light exposure also reveal differential affinity for photolysed Cbl. Light-exposed WT CarH is 1,329 Da larger in molecular mass than light-exposed H132A CarH, a number corresponding to the molecular mass of Cbl. This difference in molecular mass suggests that, even under the harsh conditions of this experiment, photolysed Cbl remains bound to the WT CarH monomer but not to H132A CarH, indicating that Cbl dissociates more readily without the bis-His ligation. The minor peak next to WT CarH arises from protein bound to a potassium ion (mass shift 39 Da). Species marked with an asterisk correspond to an unidentified impurity in the H132A CarH sample. d. Control experiment showing ESI–TOF mass spectra of WT and H132A CarH in the AdoCbl-bound dark state. The very similar molecular masses obtained (differing only because of the H132A mutation) indicate that the mutation has no effect on AdoCbl binding, consistent with the fact that His132 is not coordinated to Cbl when the upper 5′-dAdo ligand is present. Both WT and H132A CarH lose their AdoCbl cofactor as the tetramer disassembles into monomeric units.
## Extended Data Table 1 | Crystallographic data collection and refinement statistics

| PDB code | 5C8A | 5C8D | 5C8E | 5C8F | light-exposed CarH |
|----------|------|------|------|------|--------------------|
| **Data collection** |      |      |      |      |                    |
| Space group | $P4_12_2$ | $P4_12_2$ | $P2_12_1$ | $P1$ | $P2_12_2$ | $P2_12_2$ | $I4_122$ |
| Cell dimensions \(a, b, c (\text{Å})\) | 94.5, 94.5, 180.5 | 94.5, 94.5, 180.5 | 51.4, 99.7, 144.0 | 78.7, 79.7, 118.4 | 177.9, 141.8, 162.7 | 176.7, 141.7, 162.6 | 126.9, 126.9, 149.5 |
| \(\alpha, \beta, \gamma (^{\circ})\) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Wavelength (Å) | 0.9792 | 1.6039 | 0.9795 | 0.9795 | 0.9795 | 1.7365 | 0.9791 |
| Resolution (Å) | 200 – 2.80 | 200 – 3.30 | 100 – 2.15 | 100 – 2.80 | 100 – 3.89 | 100 – 5.00 | 100 – 2.65 |
| \(R_{	ext{free}}\) (%) | 6.5 (69.0) | 12.4 (38.6) | 5.3 (60.8) | 10.5 (83.2) | 9.8 (126.4) | 7.2 (107.1) | 6.9 (167.8) |
| \(R_{	ext{free}}\) (%) | 7.2 (76.5) | 6.2 (71.5) | 11.7 (91.9) | 10.2 (131.9) | 8.6 (127.2) | 7.2 (174.4) | 6.9 (167.8) |
| \(CC_{12}\) | 99.9 (77.1) | 99.9 (87.7) | 99.7 (84.3) | 99.9 (87.0) | 99.9 (54.3) | 100.0 (68.5) | 100.0 (68.5) |
| \(<1/(\sigma(I))>\) | 16.4 (2.3) | 15.7 (7.0) | 15.2 (2.1) | 12.0 (2.0) | 15.7 (2.0) | 7.4 (1.1) | 26.8 (1.8) |
| Completeness (%) | 99.5 (99.9) | 99.5 (100.0) | 98.5 (99.4) | 94.9 (95.5) | 99.9 (100.0) | 97.7 (98.2) | 100.0 (99.9) |
| Redundancy | 5.3 (5.4) | 11.9 (11.5) | 3.6 (3.6) | 5.6 (5.6) | 12.2 (12.4) | 3.5 (3.3) | 12.8 (13.4) |
| **Refinement** |      |      |      |      |                    |
| Resolution (Å) | 100 – 2.15 | 100 – 2.80 | 100 – 3.89 | 100 – 2.65 |
| No. reflections | 40511 (2970) | 59284 (4419) | 38480 (2819) | 18067 (1319) |
| \(R_{	ext{free}}/R_{	ext{cc}}\) | 0.183/0.227 | 0.183/0.230 | 0.250/0.257 | 0.172/0.203 |
| No. atoms | 5766 | 14668 | 14500 | 2090 |
| protein | 364 | 728 | 728 | 91 |
| Cbl | 364 | 728 | 728 | 91 |
| 5'-deoxyadenosine | 72 | 144 | 144 | – |
| water | 259 | – | – | 19 |
| DNA | – | – | 2120 | – |
| glycerol | 6 | – | – | 6 |
| chloride | – | – | – | 2 |
| B-factors |      |      |      |      |                    |
| protein | 48.5 | 76.4 | 164.4 | 85.9 |
| Cbl | 45.1 | 65.2 | 153.9 | 100.8 |
| 5'-deoxyadenosine | 46.1 | 75.8 | 158.5 | – |
| water | 47.3 | – | – | 71.2 |
| DNA | – | – | 231.3 | – |
| glycerol | 54.3 | – | – | 95.3 |
| chloride | – | – | – | 94.7 |
| R.m.s deviations |      |      |      |      |                    |
| Bond lengths (Å) | 0.004 | 0.005 | 0.005 | 0.004 |
| Bond angles (°) | 0.82 | 0.92 | 0.79 | 0.74 |
| Rotamer outliers (%) | 5 (0.9%) | 8 (0.6%) | 10 (0.8%) | 0 (0.0%) |

* Structure was not refined to completion.
† Bijvoet pairs were not merged during data processing.
‡ Values in parentheses indicate highest-resolution bin.
§ Values were not reported in the version of Scalepack used for scaling.