Evolution of structural dynamics in bilobed proteins

Giorgos Gouridis^{1,2,3,4,#,*}, Yusran A. Muthahari^{1,3,#}, Marijn de Boer^{1,#}, Konstantinos Tassis¹, Alexandra Tsirigotaki³, Niels Zijlstra², Nikolaos Eleftheriadis¹, Ruixue Xu^{3,5}, Martin Zacharias⁶, Douglas A. Griffith², Yovin Sugijo¹, Alexander Dömling⁵, Spiridoula Karamanou³, Anastasios Economou³, Thorben Cordes^{1,2,*}

¹Molecular Microscopy Research Group, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

²Physical and Synthetic Biology, Faculty of Biology, Ludwig Maximilians-Universität München, Großhadernerstr. 2-4, 82152 Planegg-Martinsried, Germany

³Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute for Medical Research, KU Leuven, 3000 Leuven, Belgium

⁴Structural Biology Division, Institute of Molecular Biology and Biotechnology (IMBB-FORTH), Nikolaou Plastira 100, Heraklion-Crete, Greece

⁵Drug Design Group, Department of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands

⁶Physics Department, Theoretical Biophysics (T38), Technical University of Munich, München, Germany

[#]these authors contributed equally to this work

*corresponding author emails: g.gouridis@imbb.forth.gr (G.G.) & cordes@bio.lmu.de (T.C.)

Abstract

Novel biophysical tools allow the structural dynamics of proteins, and the regulation of such dynamics by binding partners, to be explored in unprecedented detail. Although this has provided critical insights into protein function, the means by which structural dynamics direct protein evolution remains poorly understood. Here, we investigated how proteins with a bilobed structure, composed of two related domains from the type-II periplasmic binding protein domain family, have undergone divergent evolution leading to modification of their structural dynamics and function. We performed a structural analysis of ~600 bilobed proteins with a common primordial structural core, which we complemented with biophysical studies to explore the structural dynamics of selected examples by single-molecule Förster resonance energy transfer and Hydrogen-Deuterium exchange mass spectrometry. We show that evolutionary modifications of the structural core, largely at its termini, enables distinct structural dynamics, allowing the diversification of these proteins into transcription factors, enzymes, and extra-cytoplasmic transport-related proteins. Structural embellishments of the core created new interdomain interactions that stabilized structural states, reshaping the active site geometry, and ultimately, altered substrate specificity. Our findings reveal an as yet unrecognized mechanism for the emergence of functional promiscuity during long periods of protein evolution and are applicable to a large number of domain architectures.

protein energy landscape adapted from Kern and co-workers¹² according to the Tier description and definitions introduced by Frauenfelder and co-workers.¹⁵ A structural state is defined as the lowest point of a well on the energy surface. The populations of the Tier-0 states “closed” and “open” are defined as Boltzmann distributions and their relative probabilities (p_c , p_o) are determined in this paper by smFRET, which allow to follow large-scale domain motions. Tier-1 states describe local and fast structural fluctuations (e.g., changes in secondary structure elements like loop motions or loss of secondary structure). Tier-1 dynamics were probed in this study by HDX-MS. Rapid and localized Tier-2 dynamics (e.g., side chain rotations) were not considered here, but can be monitored via MD simulations. Changes in the chemical environment, i.e., absence or presence of a ligand, modify the energy landscape via a bias for one of the two states. Typically, thermodynamic parameters such as the free energy difference of Tier-0 wells (ΔG_{OC}) can be determined by ITC. **(B)** Structure of a representative bilobed protein, the substrate binding domain 2 (SBD2) of the ATP-binding cassette (ABC) amino acid transporter GlnPQ from *Lactococcus lactis* (PDB: 4KR5). **(C)** Summary of the structure-based phylogenetic tree with schematic representations of the different structural classes (class A-G) highlighting their termini. Complete sequence and structure-based phylogenetic trees are provided in Fig. S1 and SuppFile. 1, 2. We used the following notation: Secondary structure elements that are common between the different classes were assigned a number identifier (e.g. helix H5 in classes F and G), whereas the unique elements have a letter followed by a number identifier (e.g. HD1, first unique helix of class D proteins). An asterisk (*) marks a structural subclass (shown in detail in SuppFile. 3). Schematics are based on the crystal structures of the selected proteins. **(D)** Topology of bilobed proteins depicting the consensus of secondary structure elements: strands (s) or helices (H) belonging to the Domains (D1, D2), β -strands (β 1, β 2) and C-tails of all seven classes (A-G) are shown. Revised alignments of bilobed proteins are shown in SuppFile. 3. The secondary structure elements forming the conserved *cherry-core* structure are depicted on the top row.

Tier-0 dynamics were observed and characterized in various settings, e.g., in motor proteins, where they are used in propelling movement along filaments¹⁶, in the transport of molecules or biopolymers across biological membranes¹⁷⁻²¹, or in the activity of proteins that perform mechanical work.²² Tier-1 dynamics drive the actions of various signaling proteins for transmission of signals.²³⁻²⁵ Structural and biochemical data indicate that also enzymes show varying degrees of structural dynamics²⁶, although it is not well understood what precise role this plays for catalytic activity. The current belief is that extensive structural dynamics in enzymes are not necessarily required for catalysis²⁷, but rather enact in regulation. For instance, many protein kinases exploit Tier-0 dynamics to generate active or inactive structural states.²⁸ Tier-2 dynamics have been shown to be important for the evolution of enzymatic function.²⁹ In addition to domain motions occurring within a structure, protein oligomerization and the possible quaternary dynamics can also be relevant for function.³⁰ A well-characterized example is the allostery of haemoglobin that occurs on the transition between two distinct quaternary states (relaxed and tense).^{31,32} All this highlights that multi-Tier dynamics of proteins occur on various time- and length-scales (Fig. 1A) and are often the basis for function. It is, however, not well understood how structural dynamics are designed and optimized during evolution to tailor protein function.

Analysis of protein sequences and structures has provided important insights into the evolution of protein function.³³ Proteins can be decomposed into common constituent domains, a domain being a contiguous polypeptide chain within a protein that can fold and function independently. A powerful approach is to assign the domain components of proteins to families and superfamilies on the basis of sequence alone (Pfam database)³⁴ or in combination with structural information (CATH and SCOP databases).^{35,36} The CATH and SCOP databases combined have identified ~3000 domain superfamilies comprising >50 million domains that account for ~70% of domains in completed genomes.³⁷ More recently, the database ECOD groups domains by considering their evolutionary relationships.³⁸ In ECOD ~760000 domains have been assigned to 3700 homologous groups.³⁹ The most highly populated domain superfamilies/homologous groups are universal to all kingdoms of life.^{40,41} The prevalence of proteins with two or more domains, i.e., proteins with multidomain architectures, and the recurrent appearance of the same domain in non-homologous proteins suggests that functional domains are reused when creating new proteins and functions.⁴² Analyses of selected domain groups^{43,44}, and more recent large-scale investigations^{45,46} have shown that domains in such groups generally share a common structural core (40-50% of the domain) that is highly conserved, even for relatives separated by billions of years. To achieve functional promiscuity of such a structural core, it has been proposed that larger structural embellishments (secondary structure elements or even entire domains) need to be added to the core for altered biochemical function.⁴⁷ Such fusions occur frequently at the N or C termini⁴⁰ which has led to the belief that structural elements act as Lego bricks, which are recombined in various ways for new functions to emerge during evolution.^{42,47}

Much less is known about the evolution of structural dynamics or the role played by structural dynamics in the evolution of protein function. A recent model suggests that the native state of an evolved protein is the most abundant state of all possible structural states, which was selected for a specific function.^{48,49} The “avant-garde” evolvability theory proposes the existence of a highly promiscuous primordial protein structure. It is assumed that for the emergence of the evolved native state, several synergistic rounds of neutral mutations and plastic modifications were required, during which, the ability to evolve (evolvability) was traded for ligand-functional specificity.^{50,51} These observations were experimentally verified⁵² and agree with the proposal that changes in structural dynamics serve as a mechanism for the evolution of specialist enzymes from promiscuous generalists.⁵³ Though, discordant examples are also observed in evolution.⁵⁴ Neutral mutations do not compromise the native state and the original function of a protein, whereas plastic modifications are referred to as adaptive structural changes in response to a few mutations leading to functional specialization. Other recent studies indicate that this evolvability theory can explain the short-period evolution, e.g., variation of enzyme local flexibility (variation of Tier-1/2 dynamics) to acquire new functions, particularly well.^{48-50,55} Remarkably, this short-period evolution can be faithfully reproduced *in vitro* using directed evolutionary approaches based on consecutive rounds of single point mutations.^{48,55} Ancestral protein reconstruction has also been useful in elucidating the role of structural dynamics in the emergence of specialized amino acid binding proteins from a promiscuous ancestor.⁵⁶ However, it still remains unclear how during longer periods of evolution, a primordial core structure evolves to modulate or diverge Tier-0 or quaternary dynamics and with that “generates” completely new functionalities.

Here, we test this aspect of the evolvability theory for proteins separated by long evolutionary periods during which Tier-0 or quaternary dynamics were introduced to an existing protein core structure. For this, we identified proteins with a bilobed domain structure with a high degree of plasticity and neutrality^{29,49}, having (i) a conserved core-structure, (ii) large sequence diversity and related functional divergence and (iii) occurrence in all kingdoms of life. The selected structural core is composed of two Rossman folds, believed to be amongst the most ancient architectures⁵⁷, connected by a single β -sheet (Fig. 1B). We analysed ~600 proteins with this structural core that contain two related domains from the type-II periplasmic binding protein (PBP) domain homologues (ECOD: X-, H- and T-groups Periplasmic Binding Protein-like II). We show here that different members of these homologues domains have undergone divergent evolution by the acquisition of domains or secondary structure elements, predominantly at their termini (Fig. 1C, D). Using a combination of structural analysis and biophysical investigations, we demonstrate that such structural embellishments confer multi-Tier conformational dynamics that diversify the function of the core-structure to yield transcription factors, enzymes or extra-cytoplasmic transport related proteins. To understand both the mechanistic distinction of these proteins and their evolutionary process, we used single-molecule Förster resonance energy transfer (smFRET)⁵⁸ and Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS)⁵⁹. smFRET allows to monitor Tier-0 dynamics with a temporal resolution down to microseconds and sub-nanometer spatial resolution at the single-molecule level, even for highly-heterogeneous conformational ensembles.^{12,60-62} HDX-MS complements smFRET, as it can probe Tier-1 dynamics throughout the structure that occur within the structural states identified by smFRET (Fig. 1A).⁶³ Using this combination of techniques, we show how the C-terminal extensions confer distinct Tier-0 dynamics to the structural core in a manner specific to their three-dimensional orientation. This arrangement also dictates specific geometrical criteria, crucial for establishing specific ligand interactions, all of which we detail in this work. On the other hand, we reveal the means by which N-terminal domain additions enable oligomerization to provide distinct quaternary dynamics in LysR-type transcriptional regulators. The remarkable modularity of these proteins permits us to confirm and expand the evolvability theory for the primordial core-structure during long-period evolution, which is largely facilitated, seemingly, by genetic recombination events.

RESULTS

Structure, classification and evolution of the selected bilobed proteins. In this study, our focus was on proteins composed of two globular lobes (bilobed), each of 3 layers ($\alpha/\beta/\alpha$), articulated around a

central β -sheet hinge (Fig. 1B). This focus meant that not all multidomain architectures harbouring the type-II PBP domains were included. To investigate the evolution of structural dynamics in the selected proteins, we constructed phylogenetic trees based on both sequence and structural information (Fig. 1C). This analysis indicated that the proteins evolved from a common ancestor that diversified into seven distinct structural classes A to G (Fig. 1C/D), members of which are found throughout all kingdoms of life with some even present in viruses (Fig. S1A, SuppFile. 1, 2). They have a consensus structure (Fig. 1B, top row in Fig. 1D and SuppFile. 3), which we have dubbed the *cherry-core* (hereafter CC, and proteins harbouring this core, *cherry-core proteins*, CCPs) because of the bilobed structure's resemblance to a cherry (Fig. S2A), that contains the type-II PBP domains.⁶⁴ According to ECOD database, the two domains of CCPs belong to the same X-, H- and T-groups (Periplasmic Binding Protein-like II; SuppFile. 3), supporting the common ancestry.³⁸ In the consensus structure (Fig. 1C/D), domains D1 and D2 adopt a face-to-face mirror-fashioned geometry with the active site, which is typically a ligand-binding site, located at their interface (Fig. S2B).

The majority of the selected proteins have distinct segments, N-terminally linked to the CC (Fig. 1C/D, Fig. S1). Class A proteins have the addition of a LysR winged helix-turn-helix (HTH)-type DNA binding domain (ECOD: X-, H-group: HTH, T-group: Winged; Fig. 1C). This element has been shown to be responsible for oligomerization and binding to promoter DNA.⁶⁵⁻⁶⁷ Most proteins of classes B-D, and F, G contain N-terminal localization signals for export via the general secretion (Sec) pathway or the twin-arginine translocation (Tat) pathway.⁶⁸⁻⁷¹ The presence of the various address tags was obtained from UniProtKB⁷² and verified manually by inspecting all 600 protein sequences with PRED-TAT.⁷³ Class E proteins are predominantly cytosolic and lack such an N-terminal signal peptide.

In addition, classes B-G have distinct C-terminal structural embellishments, hereafter termed C-tails (Fig. 1D). For example, Helical-tail 1 (H1) is common to all classes, except A and C, and importantly, it has a similar placement in the 3D architecture of the proteins. All the other C-tails are unique to a specific class, such as the Helical-tail G (HG), present only in class G proteins.

Function, ligand specificity and conformational states of the selected bilobed proteins. The function or ligand specificity of the CCPs, as documented in UniProtKB⁷², was found to correlate with the assigned structural class (Fig. S1A). Class A proteins are bacterial transcription factors of the LysR-type transcriptional regulator (LTTRs) family. Class E proteins are predominantly eukaryotic single-turnover enzymes. The majority of the remainder (class B-D, F, G) are found in prokaryotes, and associate with the translocator domains of ABC transporters, or with the membrane-embedded domains of chemoreceptors, where they mediate unidirectional solute transport and signal transduction respectively.

To investigate what role the distinct C-tails of these proteins might have played in the evolution of structural dynamics and function, we examined CCPs for which high-resolution structures of unliganded (*apo*) and liganded (*holo*) states were available (Fig. 2, Table S1). Interestingly, class A and E proteins, in the majority of cases, display nearly identical *apo* and *holo* structures. They also show the widest variety of substrates with little chemical structure similarity (Fig. S3). For most members of the other classes, D1 and D2 of the CC undergo a rigid body rotation of varying degrees (Fig. 2, Table S1). For the solute binding proteins this mode of substrate binding has been called the "Venus-Fly Trap mechanism".⁷⁴ These proteins also recognize ligands with a specific pharmacophore⁷⁵: amino acids, ethanolamines, phosphonates, iron-phosphate complexes and carbohydrates are recognized by classes B, C, D, F and G, respectively (Fig. S3). Another striking difference is that proteins in classes B-G are monomeric, whereas those in class A are oligomeric (see sections below).

Figure 2. Structures of CCPs highlighting their C-tails. Structures (top) and schematics (bottom) of the identified structural classes (A-G). The *apo* and *holo* structures of the indicated CCPs are superimposed to highlight the role of the C-tail. The domain to which the two structures were superimposed is represented in grey (D1 in A-E and D2 in F, G), whereas the one that is displaced relative to the C-tail in the open state is represented in green. Arrows and schematics indicate domain displacement (see also Table S1). Interaction contact maps are presented in Table S2. The most prominent contacts between the indicated secondary structure elements of the C-tail and D1 (F, G) or D2 (B-E) that stabilize the open state are indicated by a yellow dashed line. The PDB codes and protein names are indicated.

smFRET to monitor large domain motions (Tier-0) in bilobed proteins. To investigate structural changes and dynamics of the CCPs, we used smFRET^{12,60-62} to probe Tier-0 dynamics at near-physiological conditions in aqueous buffer at room temperature. The following representative proteins (details in Fig. S1) were investigated: the effector binding domain (EBD) of CynR, representing the CC of full-length CynR⁶⁶ (class A), SBD2⁷⁶ (class B), OpuAC⁷⁷ (class C), CmpA⁷⁸ (class E) and MalE⁷⁹ (class G). For these experiments, D1 and D2 were stochastically labelled using cysteines that were substituted for non-conserved, surface-exposed residues, one in each domain (Fig. 3A). Fluorophore labelling was performed by maleimide-thiol conjugation.^{22,80} Labelling positions were selected based on the crystal structures, and the residues chosen display large changes in separation between the *apo* and *holo* states. smFRET was performed by confocal microscopy with alternating laser excitation (ALEX).⁶⁰

As predicted by our structural analysis (Fig. 2, Table. S1), and in line with our previous observations⁸¹, the FRET efficiency histograms and fitted distributions (Fig. 3B-D) shifted towards higher FRET efficiency (*E*) values upon addition of saturating concentrations of ligand, for SBD2, OpuAC and MalE. This indicates that in the *apo* state, the donor and acceptor dyes are further apart (Fig. 3B-D upper panels, low FRET) compared to the *holo* state (Fig. 3B-D, lower panels, high FRET). Thus, our data suggest that ligand binding drives Tier-0 dynamics in these CCPs.

In contrast, the distributions of class A and E proteins, CynR⁸² and CmpA⁷⁸ respectively (Fig. 3E, F), were virtually identical in the absence or presence of saturating ligand concentrations. Ligand binding was confirmed via isothermal titration calorimetry (ITC) showing binding affinities of both proteins in the micromolar range (Fig. 3G, H). Thus, we conclude that the CC of class A and E proteins lack Tier-

0 dynamics on the probed reaction coordinates for the selected FRET-distance pairs in contrast to the other structural classes.

Interestingly, this observation is in line with the known biological function. Class E proteins are predominantly single turnover enzymes (Fig. S1A), for which the rigidity of their active site is a prerequisite for catalysis⁸³ (Fig. S4A). As previously suggested by us and others, Tier-0 dynamics in periplasmic-binding proteins (class B, C, D, F and G) are utilized in the regulation of transport in ABC importers (see discussion). A remaining question is, however, how ligand binding in class A proteins triggers transcriptional processes without major structural changes related to ligand binding.

Figure 3. Monitoring structural changes and ligand binding in CCPs using smFRET and ITC. (A) Schematic of experimental strategy to monitor structural states by FRET efficiency via stochastic labelling of D1 and D2 with donor and acceptor fluorophores. (B-F), Solution-based apparent FRET efficiency histograms in the absence (top) or presence (bottom) of saturating ligand concentrations for each protein indicated. Grey bars are experimental data and solid line is the fit. Centre position of Gaussian fits are given in Table S5. (G, H) Binding isotherms of the calorimetric titration of azide (G) and calcium-carbonate (H) to CynR (G) or CmpA (H), respectively with the indicated thermodynamic parameters. According to high-resolution structural data⁸⁴, both Ca^{2+} and CO_3^{2-} are present in the CmpA binding cleft. Indeed, we observed heat release upon titration of Ca^{2+} to a CO_3^{2-} bound CmpA. Data points represent the heat of reaction per injection and the line is the fit.

HDX-MS to study local flexibility (Tier-1) of a class A protein, CynR. The DNA binding domain of class A proteins typically comprises a ~58 aa HTH motif. This is followed by a ~20 aa long helix that provides a dimerization interface and a connecting loop that attaches the DNA binding domain to the CC (Fig. 4A). The CC acts as the tetramerization interface within the full-length CCP or the dimerization interface within the effector binding domain (Fig. 4B, Fig. S4B-E). Indeed, by using SEC-MALS we observed that (full length) CynR is tetrameric, whereas the CC of CynR is dimeric (Fig. 4C).

To determine whether ligand binding to CynR resulted in conformational changes that were not detectable along our selected reaction coordinate, or that were too small or fast for smFRET (Fig. 3E), we performed HDX-MS. In contrast to smFRET, which reports on a single distance along a single reaction coordinate, HDX-MS probes structural dynamics at near-residue level resolution providing global insights into Tier-1 dynamics. HDX-MS detects the exchange of hydrogens with deuterium at solvent-accessible and non-hydrogen-bonded backbone amides.^{85,86} Hydrogens involved in stabilizing the secondary, tertiary or quaternary structure of a protein via hydrogen bonds are exchanged more slowly through structural transitions that disrupt these bonds. Deuterium incorporation into the protein can then be determined, following proteolysis, by mass spectrometry. The mass difference between hydrogen (^1H) and deuterium (^2H) results in a mass shift between non-deuterated and deuterated peptides that is a measure of the number of exchanged hydrogens.^{59,87,88}

For such investigations, the CC of CynR or full-length CynR were isotopically labelled (pD, 7.4; 25 °C) for different time periods (10 – 10⁵ sec) either in its free or in its DNA bound state, and the peptic fragments were identified by MS (Fig. 4D, SuppFile. 4). For each peptide, the fraction of deuterium uptake relative to the maximum determined deuterium incorporation was calculated (SuppFile. 4). The data reveal a rigid character of the CC of CynR, whereas the dimerization helix and the rest of the DNA binding domain turned out to be more flexible (Fig. 4D). To identify the regions in which pronounced structural changes were induced by azide binding, we performed comparative HDX-MS. For this, we determined the difference of deuterium uptake (ΔD) for each peptide between different conditions, e.g., CynR *apo* versus *holo* states. Observed differences would indicate a decrease (protection) or an increase (deprotection) of deuterium uptake upon azide binding (SuppFile. 4). No statistically significant change in the deuterium uptake was observed upon azide binding (Fig. 4E). From this, we can conclude that no detectable change in structural dynamics is induced by the ligand in the CC of CynR in its free or DNA bound form (Fig. 4E, SuppFile. 4). This includes Tier-0 dynamics (detected by smFRET and HDX-MS), Tier-1 or quaternary dynamics (both detected by HDX-MS).

Figure 4. CynR tertiary and quaternary assemblies probed by MALS and HDX-MS. (A) Crystal structure of the CynR CC (PDB: 2HXR) with colored secondary structural elements discussed in the text that are critical for its quaternary dynamics (top panel). A homology model of full-length CynR obtained from the SWISS-MODEL server with CbnR (PDB: 1IZ1) as a template (bottom panel). The HTH domain is colored green and the loop connecting it to the CC splitpea. W64 at the tip of the dimerization helix and L88 of the connecting-loop are indicated with an arrow. (B) top panel: Schematic representation of the CC of CynR using the same color-coding as in A. middle panel: Schematic representation of the CynR homology model, with one of the protomers in the compact and the other in the extended configuration. The CC of CynR protomers self-associate to form dimers, or for the full-length protein, tetramers. For clarity, two of the protomers have been omitted. Bottom panel: The tetrameric assembly formed by the self-association of the dimerization helices and the CC. Of the two interacting protomers, one is present in the compact configuration, whereas the other is extended. The two protomers with the compact configuration are shown at the top of the plane, whereas those that are extended are at the bottom. (C) SEC-MALS analysis of full-length and CC of CynR (3 μ M). UV traces of the chromatograms were superimposed on the measured mass (black cycles). (D) Structural dynamics of full-length CynR in the absence of ligand and DNA by HDX-MS. Deuterium uptake values are reported for the incubation times in deuterated buffer and expressed relative to the fully deuterated control. These values are mapped onto the CynR homology model (as in A), using the indicated color gradient. Proline residues as well as the first residue of each peptide were excluded from mapping, as they do not contribute to the observed D-uptake. (E) Scatter plot visualization of the statistical analysis of D-uptake differences between the apo and holo states of full-length CynR (upper panel) and the CC of CynR (bottom panel). Three statistical criteria were used (Materials and Methods), as described previously.⁸⁹ Statistically significant differences would appear as black spheres (indicating that ΔD -uptake $> 2 \times SD$ for a specific peptide), lying outside the 99% confidence threshold ($1-p \geq 0.99$; indicated on y-axis) and outside the $\pm 4 \times$ pooled average SD cut-off (indicated on x-axis; value given on the right).

An asymmetric C-tail drives large-scale tertiary (Tier-0) structural changes. From our results on the selected CCPs, only those with a C-tail, display Tier-0 dynamics (Fig. 2, 3). To investigate how the C-tails introduced Tier-0 dynamics to the CC, we compared crystal structures of the *apo* and *holo* states of the CCPs to identify interactions between their C-tails and the CC. Fig. 2 summarizes the results of these comparisons. Interestingly, we found that the *holo* structures are similar for all CCPs, but the *apo* structures are class-specific. Contact mapping of the interactions between the CC and C-tail using the protein interaction calculator web-server⁹⁰, showed that the number and characteristics of the interactions depends on the structural class (Fig. 2, Table. S2). Strikingly, these interactions predominantly stabilize the open conformation of the CC, the only exception being for the class A and E. In the former cases, stabilization is associated with an asymmetrical placement of the C-tail with respect to D1 and D2. In contrast, the C-tail of class E (predominantly H3 and HE1; Fig. 2) is placed symmetrically around D1 & D2, and thus cannot provide the required structural asymmetry needed to create a stable open state. The stabilizing interactions of the C-tail to form the open state involve the consensus CC-helices of D1 and D2 (Fig. 2, Table. S2). In classes B, C and D such interactions involve D2, whereas D1 is contacted in classes F, and G. These asymmetrical interactions create active sites with distinct geometries (Fig. 2, Fig. S2).

In alternative phylogenetic trees, based on the protein sequence using either D1/D2-domains or the C-tail, the clustering remains similar (Fig. S1B-D, SuppFile. 1, 2); indicating that D1/D2 (CC domains) and the C-tail of a specific class co-evolved to be part of the same polypeptide. This is in line with the role of the C-tail to interact with one specific domain, D1 or D2, to stabilize the open structural state.

Experimental verification of the role of C-tail interactions in stabilizing the open conformation.

To confirm the role of the C-tail interactions with the rigid domains D1 or D2 for stabilization of the open state, we manipulated the relevant ones (Fig. 2; Table. S2). The impact was tested via assessing structural states and ligand binding affinities, monitored in smFRET experiments. Test cases were selected from class B and G, as these CCP classes are only remotely related by the first major clade in the evolutionary trees (Fig. 1C) and their open state is stabilized by distinct helices within the C-tails (Fig. 2), contacting either D2 (class B) or D1 (class G).

In SBD2, a hydrophobic interaction between L480 in the C-tail and P419 in D2 was weakened by substitution of L480 with alanine (L480A; Fig. 5A, Table. S2). The mutation resulted in the appearance of a subpopulation of molecules (~15%) that were in the closed state in the absence of glutamine (Fig. 5C vs B). To rule out the possibility that this might have been due to an artefact introduced by the choice of fluorophores, a second pair of fluorophores was tested, which showed a comparable result (Fig. S5). We also examined whether residual endogenous ligand might account for this subpopulation by performing smFRET measurements on diluted samples, but these experiments showed subpopulations of a similar size (Fig. S5). A closed-unliganded conformation was also observed for the wild-type SBD2 previously, but with a much lower abundance (~1%). Detection of this small subpopulation required the use of scanning confocal microscopy⁸¹, because populations <5% cannot be detected reliably with ALEX microscopy. We also observed small differences in the mean *E* values for *apo* and *holo* states of SBD2 (L480A) as compared to SBD2 suggesting that the structural landscape had been altered (Fig. S5). Destabilizing the open state is supposed to decrease the K_d ⁹¹, consistent with a 2-fold increase in its glutamine binding affinity of SBD2 (L480A) as compared to the wild-type (K_d of 840 ± 270 nM and 1990 ± 130 nM, respectively; Fig. 5D).

Figure 5. Experimental verification of the role of C-tail interactions in stabilizing the open conformation of SBD2 and MalE by smFRET and HDX-MS. (A, E) Dotted rectangles (left panels) on the SBD2 (in A PDB: 4KR5) and MalE (in E PDB: 1OMP) structures highlights the critical contact region between the C-tail and the CC that stabilize the open state. Zoom in of rectangle regions (middle and right panels) depicting interactions between the C-tail helix H1 and D2 (A) or between the C-tail helix H4 and D1 (E) in the indicated *apo* (open) or *holo* (closed) states. Distances (Å) between L480 and P419 (A) or between M321 and Y90/F92 (E) are shown as black dotted lines. (B, C) Solution-based apparent FRET efficiency histograms of SBD2 (B) and SBD2 (L480A) (C) at different conditions as indicated. (D) Fraction of the closed state (high FRET state; Fig. S5) of SBD2 and the indicated derivative as a function of glutamine concentration. (F, G) Solution-based apparent FRET efficiency histograms of MalE (F) and MalE (M321K) (G) at different conditions as indicated. (H) Fraction of the closed state (high FRET state; Fig. S6) of MalE and its derivatives as a function of maltose concentration. (I) Maltose release from MalE (M321K) over time determined by solution-based smFRET (see detailed values in Figure S6). Data points (panel D, H) and grey bars (panel B, C, F, G and I) are the experimental data, and the solid line is the fit. (J) Map of regions in MalE structure that show statistically significant increase in deuterium uptake caused by M321K (numerical values and complete statistical analysis presented in SuppFile. 5). $n=3$. (K, L) Deuterium uptake for the indicated MalE C-tail helices. Standard Deviations (SD) are shown in the deuterium uptake plots.

For MalE, we constructed the derivatives MalE (M321A)⁹² and MalE (M321K), with weakened interactions between the C-tail H4 and D1, by disrupting the hydrophobic interaction of M321 with Y90 and F92, and the aromatic sulphur interactions with Y90 (Fig. 5E, Table. S2). For MalE (M321A), the stabilizing interactions of M321 are partially abolished. This resulted in an 8-fold increase of maltose affinity [K_d from 2400 ± 400 nM in MalE to 300 ± 50 nM in MalE (M321A) (Fig. 5H, Fig. S6)]. An even stronger effect was observed for MalE (M321K) with an affinity enhanced by 3000-fold (K_d of 0.81 ± 0.15 nM; Fig. 5G vs Fig. 5F, Fig. 5H). Shilton and co-workers have also proposed the hydrophobic interactions of M321 as being important structural determinants of the open state⁹², affecting the affinity of MalE for maltose, in agreement with our results.

We next compared the lifetime of the closed maltose-bound conformation of MalE (M321K) with that of wild-type. Addition of 10 nM maltose allowed MalE (M321K) to occupy the closed state exclusively (Fig. S6I, top panel). 20 μ M unlabelled MalE (M321K) protein was subsequently added to scavenge maltose, which is stochastically released from the labelled protein. In a time-course experiment the decrease in the population of closed maltose-bound MalE (M321K) was then followed as a function of time⁹³ (Fig. 5I, Fig. S6I). From these experiments, we found that the lifetime of the closed maltose-bound conformation of MalE (M321K) was 122 ± 12 s, and that this value was 2500-fold higher than the value determined for wild-type MalE (0.048 ± 0.010 s; Fig. S6). This result is consistent with the observed

increase affinity of the derivatives for maltose. Our observations clearly demonstrate that the C-tail is crucial for the conformational landscape of the CC and the modulation of the states by ligands.

The C-tail dynamics control the structural transitions of the *cherry-core*. In the above-mentioned derivatives of SBD2 and MalE, the structural basis for the destabilization of the open state could not be addressed mechanistically. Our smFRET experiments report predominantly on Tier-0 dynamics that affect the protein's tertiary structure (Fig. 1A). Considering that Tier-0 dynamics, as a consequence of C-tail-D1/D2 destabilization, could only be observed for SBD2; we performed comparative HDX-MS (SuppFile. 5) to monitor changes in the secondary structure of MalE in comparison to MalE (M321K), (Fig. 5J). For this, we determined the difference of deuterium-uptake (ΔD) for each peptide between MalE and MalE (M321K) in their *apo* states. Remarkably, comparative HDX-MS indicates that the differences between wild-type and MalE (M321K) are localized almost exclusively at the C-tail and specifically in regions interacting with D1 (SuppFile. 5, Fig. 5J). The D-uptake of the C-tail helices H2/H4, that are critical for stabilizing the open state (Fig. 5E, Table S2) denotes that their rigidity was significantly reduced in the MalE derivative (Fig. 5 K, L). As might be expected, reduced rigidity occurred also in a region within the CC containing Y90 and F92. Notably, the same regions become allosterically destabilized in wild-type MalE upon maltose binding (SuppFile. 5). These results support the idea that the mutation leads to a weaker interaction between the C-tail and the CC resulting in a destabilization of the open conformation, because these C-tail elements and the region containing contact residues were found to be more flexible and solvent-exposed.

Taken together, these results provide compelling evidence that the interactions between the CC and C-tails are involved in stabilizing the open conformation of the CC and thus evolved to stabilize a novel conformational state.

DISCUSSION

Common (structural) origin represents the hallmark of Darwinian evolution. Homology, or descent from a common ancestor is often deduced from similarities in protein sequences or better from structures, as the latter are more conserved during evolution.⁹⁴ However, similar structures can originate from divergent, convergent or parallel evolution.⁹⁵ The most common “tricks” nature uses to vary a protein domain are: β -strand invasion/withdrawal, insertions/deletions/substitutions of secondary structure elements, domain flip/swaps and circular permutations.^{96,97} The currently established evolvability theory relies on investigations involving a fixed length polypeptide chain by observing the effects of sequence variations, accomplished primarily by directed evolutionary approaches or by investigating closely related functional homologues. The functional promiscuity originating from structural variability is altered by the few amino acid modifications that can yield alterations of local structural fluctuations. For this reason, this theory can well explain protein evolution during short time periods.

In this study, we focused on the analysis of structures that have diverged over longer evolutionary periods. We analyzed a group of ~600 proteins that share a core structure with the same topology of secondary structure elements giving rise to identical three-dimensional structures. The structure was predominantly varied by terminal embellishments and exhibits detectable sequence identity, used for constructing the sequence-based phylogenetic trees (Fig. S1). The identified proteins show divergent long-term evolution from a common ancestor, which spreads throughout the tree of life. This common ancestor is represented by the consensus core structure (dubbed *cherry-core*, CC, Fig. S2), and encountered within the type-II class of periplasmic binding proteins.^{64,98} As proposed previously⁶⁴, the CC derived possibly from a gene duplication of a type-II periplasmic binding protein domain [(ECOD); SuppFile.3] that has been connected by two β -strands (Fig. 6A).

CCPs with asymmetric C-tails and N-terminal signal peptides. When the CC is fused N-terminally to a signal peptide and C-terminally to an asymmetric C-tail, the CC acts as an extra-cytoplasmic monomeric protein, that associates with the translocator domains of ABC transporters or with chemoreceptors (Fig. S1, Fig. 6B). The two structural states are *apo*-open vs. *holo*-closed and originate from Tier-0 dynamics, which are critical determinants of their biological function: the membrane-embedded partners can discriminate between open vs closed states to activate or inactivate a biological process^{80,99-102}, such as solute transport. These structural transitions rely on the “generation” of an open state of the CC, accomplished by the C-tail, through its asymmetric interactions with either of the domains of the CC (D1 or D2, Fig. 2). Evidently, the open state is stabilized by such enthalpic contributions and destabilized by mutations or ligand binding that increase the flexibility of the interacting regions (Fig. 5J) [the C-tail elements with either D2 (classes B, C, D) or D1 (classes F, G), (Fig. 2)]. Seemingly, entropic contributions (protein conformational entropy) bias the structural equilibrium towards the closed state. Given the fact that the *holo* is the closed state driven by ligand binding⁸¹, we anticipate that the interactions of the ligand with the CC cleft allosterically induce order- to disorder transitions to alter the structural equilibrium. Depending on the placement of the asymmetric C-tail in the three-dimensional space, we identified five different “flavors” of open states, establishing distinct geometries of active sites; all resembling a triangle distinctly oriented in space (Fig. 2, Fig. 6E). We verified that each active site geometry can recognize a specific chemical structure (Fig. S3), in full agreement with ligand binding triggering closing.

The evolvability theory of Tawfik and co-workers^{29,51} explains the functional promiscuity encountered within the different structural classes. We selected two closely related proteins in our phylogenetic tree in class B (Fig. 1C) to illustrate this (SuppFile. 2, red asterisk). SBD1 (PDB: 4LA9) mediates the unidirectional transport of two substrates (glutamine & asparagine), whereas SBD2 (PDB: 4KR5) transports one of them (glutamine), though captures it with higher affinity.⁸⁰ SBD1 can be transformed to bind glutamine with higher affinity alike SBD2 by mutating three amino acids.⁷⁶ Clearly, functional promiscuity is traded seemingly with ligand specificity.⁵⁶ However, only the modularity notion introduced in this study can explain a greater ligand-functional promiscuity. Class B proteins recognizing and mediating amino acid transport (Fig. S1), would gain tertiary structural variability by acquiring four more helical elements at their C-termini (H2, H4, H5, HG, Fig. 1C, D) (Fig. 6). This would potentially

allow to switch them to class G, to bind and mediate transport of carbohydrates (Fig. S1A). According to another evolutionary trajectory, class B proteins, by acquiring helices H3, HE1, HE2; could divert to class E, and switch from transport related proteins to enzymes (Fig. 6) by restricting their tertiary structural variability to a unique closed state. An extreme case of functional divergence has been experimentally verified in a recent study that restored the evolutionary history of the enzyme cyclohexadienyl dehydratase¹⁰³, a class B member according to our classification. The reconstructed ancestor of this enzyme was a highly promiscuous transport-related protein possessing the open-unliganded and the closed-liganded structural state and was able to bind four different cationic amino acids. On the other side, cyclohexadienyl dehydratase binds a single substrate and forms a unique closed state for its catalytic function, since it is believed that structural sampling represent a constrain for catalytic activities.^{104,105} Our study can now explain the means by which the addition of modular elements to a conserved core-structure diversify its function, i.e., by modulating its structural landscape.

CCPs with symmetric C-tails. When the CC is combined C-terminally to a symmetric C-tail during evolution (Fig. 6C), the CC either operates as an enzyme or is associated with ABC transporters (Fig. S1). Those proteins are present in a unique (*apo* and *holo* state) closed structure (Fig. 2), as asymmetrical interactions are an essential prerequisite for open state formation. Interestingly, these ABC transporter-associated proteins are both interacting with the actual translocator from the extra-cytoplasmic side (alike the classes B, C, D, F, G; described in the previous paragraph and in such a case possess also an N-terminal signal peptide), but are also tethered covalently to the translocator ATPase motor domains.⁸⁴ As the switching (open to closed) behaviour – to activate the transport cycle – is missing^{80,81,99}, the non-canonical arrangement of these ABC transporters could only trigger transport activation using a yet uncharacterized mechanism. On the other hand, the lack of Tier-0 dynamics observed in the single-turnover enzymes is conforming to their enzymatic mechanism demanding an extremely rigid active site (Fig. S4).⁸³ Evidently, the rigidity required for the chemistry is so dramatic that the flexibility of a non-covalently (to the polypeptide chain) bound histidine to the cleft would render the reaction unproductive.

CCPs with N-terminal domains. Lastly, the CC bearing an N-terminal domain harboring the HTH type DNA binding domain, but no C-tail (Fig. 6D), yield transcription factors of the LTTR family. Apparently, in such a case, Tier-0 dynamics of the CC are not required for function (Fig. 3E). The rigidity of the CC (Fig. 4D) is evidently required for the stability of the quaternary assemblies, as those have large cavities and holes (Fig. S7).¹⁰⁶⁻¹¹¹

Structural basis of the transcriptional activation by the LysR-type regulators. Our data and analysis gave insights into the steps taken during evolution to shape the CC into functional ligand-receptor associated proteins (classes B, C, D, F, G) but also one-turn enzymes (class E). Yet, how the ligand-driven signal propagates within the LTTR family, such as CynR, remains largely elusive.

The CC of CynR constitutes only one part of its structure, i.e., the sensory effector-binding domain (EBD), of a full-length transcription factor of the LTTR family with an additional winged helix-turn-helix (HTH)-type DNA binding domain (Fig. 4A/B). Although only little direct experimental evidence is available, it has been proposed that changes in the tetrameric assembly of LysR-type transcription factors can be induced by ligand binding.^{66,112} These quaternary dynamics were suggested to activate transcription by a transition from a bent DNA (transcription OFF) to an unbent state (transcription ON)^{66,112}. Our smFRET and HDX-results (Fig. 3E, Fig. 4E), however, did not reveal any detectable changes of the CC upon azide binding. This suggests that ligand-driven structural changes within the CC protomers are minor and might thus be very hard to detect.

To gain insights into the quaternary dynamics that govern function, we analysed the available high-resolution structures of the oligomeric assemblies belonging to the LysR transcription factors¹⁰⁶⁻¹¹¹; to identify the evolutionary relevant ones by EPIC.^{113,114} Subsequently, we modelled the CynR sequence after these structures in the presence or absence of DNA (Fig. S7). The structural basis of the distinct

tetrameric assemblies is the positioning of the dimerization helix with respect to the CC, dictated by the connecting loop (Fig. 4A/B). In the extreme case that an additional helix (RD1-CH5, see Supp. File 3) is linked to- and displaces the connecting loop, an octameric assembly is obtained (Fig. S7H). To shed light on this, we monitored intrinsic Trp fluorescence during thermal melting (Table. S3). From the two tryptophanes of CynR; one present in the dimerization helix (aa 64) and the other on D1 (aa 274) of the CC (Fig. 4A, B); only the former one contributes to a fluorescent signal (data not shown). Under DNA-free *apo* conditions, CynR displays a $T_{m(app)} \sim 55^\circ\text{C}$ that is significantly destabilized ($\sim 6^\circ\text{C}$) after binding to DNA. Addition of azide at free or DNA bound CynR causes a ligand-characteristic signature throughout the Trp-temperature spectrums, giving rise to a secondary $T_{m(app)} \sim 27^\circ\text{C}$ in both cases (Table. S3).

To understand the signal propagation originating from ligand binding, we inspected the available structural information (Fig. S7). In agreement with the experimental findings (Fig. 3E, Fig. 4E), our structural analysis indicated that D1-D2 motions are not required (Table. S4) for the oligomeric assemblies of the same transcription factor (OxyR) to differ dramatically in order to trigger structural changes on DNA (Fig. S7E-G). What varies between those assemblies is the order-disorder of specific secondary structure elements within D2 of the CC (S4 till $\alpha 5$, see SuppFile. 3). We anticipate that such changes in the flexibility of secondary structure elements are induced by ligand binding, alike in the case of MalE (Fig. 5J-L).

We conclude that signal propagation in CCPs comprising large D1/D2 rearrangements (Fig. 2; classes B, C, D, F, G) driven by the C-tail/D1-D2 interactions involves bending/unbending of the “spring-hinge”.⁹⁸ On the other side, propagation in class A is initiated by small/localized rearrangements of secondary structure element within the CC somehow transmitted to the N-terminal HTH domain leading to global quaternary structural changes (Fig. S7, S8). Additional analysis regarding the molecular mechanisms of LTTR-transcriptional regulators will be the subject of future studies.

The energy landscape of the CCPs over time. Figure 6 summarizes our proposed evolutionary path taken by the CC-proteins over long time periods. The terminal modules that regulate the multi-Tier structural dynamics can be described by the folding funnel model.⁸ The CCPs with asymmetric C-tails displaying Tier-0 dynamics (i.e., open-closed states) have the two characteristic wells in the funnel separated by a large energetic barrier (Fig. 6B & Fig. S8C). For that reason, such proteins are found predominantly in the lowest energetic level open state (*apo* energetic funnel; Fig. S8C), with infrequent transitions to the closed one (Fig. 5B, C). Only in the SBD1/2 protein (class B), $\sim 1\%$ occurrence of a closed state has been experimentally observed.⁸¹ By destabilizing the open state of SBD2 further and shifting it to higher energetic levels, we obtained a $\sim 15\%$ occurrence of a closed state (Fig. 5C). As in the *holo* state energetic funnel, the lowest energy state is the closed one (Fig. S8C); addition of the ligand shifts the equilibrium towards the closed state. Since the dissociation constant (K_D) derives from the difference between the lowest energetic levels of the *apo* and *holo* funnels (Fig. S8), destabilization of the C-tail (Fig. 5D, H) that directs the open state of the *apo* funnel at a higher energetic level, leads to an increased affinity.

Alternatively, CC-proteins with a symmetric C-tail (Fig. 6C) have a single main energetic well, defining the unique closed structural state both in the *apo* or *holo* funnels (Fig. S8D).

The CCPs with N-terminal domain additions (Fig. 6D) have multiple energetic funnels corresponding to oligomerization states (Fig. S8E). The energetic funnel of such CCPs is having multiple minima corresponding to the many different arrangements of the flexible N-terminal domain with respect to the CC (Fig. S8D). Self-association (oligomer formation) and/or binding to their partners/ligands deepens specific wells that are required for function.

We anticipate the energetic funnel of the primordial consensus CC to be extremely rugged, with small energetic barriers between the wells, with a single well somewhat deeper (corresponding to the closed conformation), thus allowing sampling of multiple structural states (Fig. S8A). Such a structural variability would lead to increased substrate promiscuity, however exploiting extremely weak interactions (i.e. low binding affinities; Fig. S8B). The function of the primordial CC having no modules could be solely that of binding. In line with the evolvability theory, a specific energetic well of the funnel

can be shaped and deepened with multiple rounds of neutral mutations/plastic modifications, trading promiscuity with specificity. The evolutionary pressure for this, is to select the already deeper closed-state well, as its ruggedness (frustration)¹¹⁵ can account for high-affinity recognition of very diverse substrates. Addition of a module (internal to the core, or at its termini) can also simultaneously direct the evolutionary pressure to another well specifically, and by that create new functionalities (Fig. 6, Fig. S8): (i) In the case of proteins modulating transport or chemotaxis, an open state well is needed (being much deeper than the closed state well), and different open state wells can be created. The module also has to specifically alter not only the *apo* but also the *holo* funnel, as in the latter, the well of the closed state is expected to be the deeper one (Fig. S8C) (ii) In the case of proteins exhibiting enzymatic functions, the module has to deepen and smoothen the unique closed state well of the *apo* and *holo* funnels, since extreme rigidity is required for catalytic function (Fig. S8D). (iii) The N-terminal module of proteins destined to become transcription factors, by bending-unbending transitions of DNA, is an entire domain that induces oligomerization and tethers the CC on DNA, as there is a need to probe a larger conformational space (Fig. S8E). The CC protomers have multiple iso-energetic wells corresponding to the different positionings of the N-terminal module with respect to the CC (Fig. S8E). Oligomerization and binding to DNA and to their ligands deepens specific wells.

Clearly, the structural promiscuity achieved by the modularity introduced in this study, expands the protein evolvability theory and establish the notion that in order to comprehend protein evolution, it is essential to decode the energetic funnel of structural homologues. Structural elements or even domains added alike Lego bricks to a structural-core, being the Lego Board, trigger distinct evolutionary trajectories.

Evolution of ligand binding mechanisms. Although extensive structural and biochemical characterization of periplasmic solute binding proteins has improved our understanding of the ligand binding mechanism (lock-and-key, induced-fit and conformational selection) by which various CCPs bind their substrates, the evolutionary origin itself is still unknown. Based on our previous work, an induced-fit binding mechanism (or Venus-fly trap⁹⁸) seems the most likely solution for CCPs from classes B,C,D,F,G⁸¹, while class A and E use an apparent lock-and-key mechanism as shown here. Additional insight from this work is that structural embellishments of the CC represent likely the evolutionary mechanism by which the large amplitude structural dynamics that underpin how an induced-fit (Venus-Fly trap) emerged. This is most clearly illustrated by the class B proteins, SBD2 in particular (Fig. 2), in which there is a single C-tail helix mediating the only intra-protein interactions that stabilize the open conformation. The new inter-domain interaction would have expanded the conformational space (since the CC-protein domain of CynR in class A use a lock-and-key binding mechanism), increasing the conformational diversity of the substrate-free molecular ensemble, and creating conformers able to accommodate non-native substrates. In line with the Tawfik and Tokuriki model, these likely low-abundance structural sub-states may have been selected and enriched leading to the evolution of extant substrate-specific class B proteins. For class C proteins, a similar evolutionary trajectory seems likely. Class G proteins such as MalE, the largest of the proteins we examined, have the additional twist that structural embellishments within the type II PBP structural core, and not only at the C-terminus, may be involved in creating structural dynamics. This raises the possibility that in some class G proteins, the dynamics may also derive from the internal embellishments, freeing the C-tails to take on additional functional roles unrelated to dynamics. A more detailed study of other class G proteins is needed to explore this further. And finally, our results and observation on class E proteins suggest that C-terminal structural embellishments do not always function to enable Tier-0 dynamics. To our knowledge, this is the first proposal of a molecular mechanism for the evolution of the structural dynamics underlying a ligand-binding mechanism.

Figure 6. Model for the evolution of the *cherry-core* and hypothetical energetic funnels: A gene duplication of a type-II periplasmic binding protein domain connected by a single β -sheet gave rise to the CC, being composed of a unique closed structural state, represented by a single-well in the energetic funnel. The function of the CC was uniquely binding. Extant proteins acquired different extensions that altered their localization and dynamics and by that their function and specificity. An N-terminal signal peptide and an asymmetrical C-terminal tail generated extra-cytoplasmic proteins evolving an additional open state. The different flavors of open states, conferred distinct substrate specificities. The 2 states of CCPs, predominantly, signal substrate transport by their association to the membrane-embedded translocator domains of (ABC) transporters. A symmetrical C-tail rigidified the closed state and yielded primarily single-turnover enzymes. The N-terminal domain addition of a flexible dimerization helix and a DNA binding motif (HTH-type) conferred distinct oligomeric assemblies with different quaternary dynamics, yielding transcription factors. For details see text and Fig. S8.

ACKNOWLEDGMENTS

This work was financed by an NWO Veni grant (722.012.012 to G.G.), an ERC Starting Grant (No. 638536 – SM-IMPORT to T.C.), Deutsche Forschungsgemeinschaft within GRK2062 (project C03 to T.C.), SFB863 (project A13 to T.C.), FWO CARBS (#G0C6814N to A.E. and G.G.). G.G. also acknowledges an EMBO long-term fellowship (ALF 47-2012 to G.G.) and financial support by the Zernike Institute for Advanced Materials and by the Rega foundation. Y.A.M. was supported by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan Republik Indonesia-LPDP RI PhD scholarship). N.Z. acknowledges an Alexander von Humboldt postdoctoral fellowship. N.E. acknowledges an MSCA SoE FWO fellowship (195872). R.X. was supported by the Chinese Scholarship Council grant. T.C. was supported by the German Academic Exchange Service (DAAD), Center of Nanoscience Munich (CeNS), LMUexcellent and the Center for integrated Protein Science Munich (CiPSM). We thank Eitan Lerner for useful comments and critical reading of the manuscript, Monique Wiertsema for help with smFRET experiments and Florence Husada for support with ITC experiments.

AUTHOR CONTRIBUTIONS

GG., M.d.B., Y.A.M., and T.C. conceived and designed research. G.G. and T.C. supervised the project. G.G., K.T., Y.S., and R.X., designed and performed molecular biology. G.G., Y.A.M., R.X., D.A.G., Y.S. and N.E. prepared protein samples. G.G., Y.A.M., M.d.B., N.Z., Y.S. and N.E. conducted single-molecule experiments. A.T. designed and conducted HDX-MS experiments. S.K. designed and conducted MALS and intrinsic tryptophan fluorescence experiments. Y.A.M., M.d.B., R.X. conducted ITC experiments. M.d.B., G.G. and Y.A.M. analysed single-molecule data. G.G. and Y.A.M. conducted structural and evolutionary analysis of proteins and established the structural classification. N.E. and Y.A.M. classified ligands. G.G., Y.A.M., M.d.B., D.A.G., and T.C. interpreted data and wrote the manuscript. G.G. and Y.A.M. prepared figures and artwork. All authors discussed the results and contributed to the final version of the manuscript, which was approved for submission by all authors.

AUTHOR INFORMATION

The authors declare no competing financial interest. Correspondence and requests for material should be addressed to G.G. (g.gouridis@imbb.forth.gr) and T.C. (cordes@bio.lmu.de)

REFERENCES

1. Cooper, G.M. & Hausman, R.E. *The Cell: A Molecular Approach*, (Sinauer Associates: Sunderland, MA, Washington, D.C, 2013).
2. Campbell, I.D. Timeline: the march of structural biology. *Nat Rev Mol Cell Biol* **3**, 377-81 (2002).
3. Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I.C. Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355-73 (1975).
4. Frauenfelder, H., McMahon, B.H. & Fenimore, P.W. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc Natl Acad Sci U S A* **100**, 8615-7 (2003).
5. Frauenfelder, H., Sligar, S.G. & Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **254**, 1598-603 (1991).
6. Nienhaus, K., Deng, P., Kriegl, J.M. & Nienhaus, G.U. Structural dynamics of myoglobin: effect of internal cavities on ligand migration and binding. *Biochemistry* **42**, 9647-58 (2003).
7. Okazaki, K., Koga, N., Takada, S., Onuchic, J.N. & Wolynes, P.G. Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proc Natl Acad Sci U S A* **103**, 11844-9 (2006).
8. Leopold, P.E., Montal, M. & Onuchic, J.N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* **89**, 8721-5 (1992).
9. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-95 (1995).
10. Liu, S.Q. et al. *Protein Folding, Binding and Energy Landscape: A Synthesis*, (InTech, Rijeka, 2012).
11. Wales, D.J. *Energy landscapes*, (Cambridge University Press, Cambridge, UK, 2003).
12. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-72 (2007).
13. Schuler, B. & Hofmann, H. Single-molecule spectroscopy of protein folding dynamics--expanding scope and timescales. *Curr Opin Struct Biol* **23**, 36-47 (2013).
14. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annu Rev Biophys* **45**, 207-31 (2016).
15. Ansari, A. et al. Protein states and proteinquakes. *Proc Natl Acad Sci U S A* **82**, 5000-4 (1985).
16. Vale, R.D. The molecular motor toolbox for intracellular transport. *Cell* **112**, 467-80 (2003).
17. Cabezon, E., Lanza, V.F. & Arechaga, I. Membrane-associated nanomotors for macromolecular transport. *Curr Opin Biotechnol* **23**, 537-44 (2012).
18. Husada, F. et al. Conformational dynamics of the ABC transporter McjD seen by single-molecule FRET. *EMBO J* **37**(2018).
19. Locher, K.P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487-93 (2016).
20. Noinaj, N. & Buchanan, S.K. Structural insights into the transport of small molecules across membranes. *Curr Opin Struct Biol* **27**, 8-15 (2014).
21. Yang, M. et al. Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD. *Nat Chem Biol* **14**, 715-722 (2018).
22. Gouridis, G. et al. ABCE1 Controls Ribosome Recycling by an Asymmetric Dynamic Conformational Equilibrium. *Cell Rep* **28**, 723-734 e6 (2019).
23. Cox, A.D. & Der, C.J. Ras history: The saga continues. *Small GTPases* **1**, 2-27 (2010).
24. Smock, R.G. & Gierasch, L.M. Sending signals dynamically. *Science* **324**, 198-203 (2009).
25. Tzeng, S.R. & Kalodimos, C.G. Dynamic activation of an allosteric regulatory protein. *Nature* **462**, 368-72 (2009).

26. Kamerlin, S.C. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78**, 1339-75 (2010).
27. Loveridge, E.J., Behiry, E.M., Guo, J. & Allemann, R.K. Evidence that a 'dynamic knockout' in *Escherichia coli* dihydrofolate reductase does not affect the chemical step of catalysis. *Nat Chem* **4**, 292-7 (2012).
28. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275-82 (2002).
29. Petrovic, D., Risso, V.A., Kamerlin, S.C.L. & Sanchez-Ruiz, J.M. Conformational dynamics and enzyme evolution. *J R Soc Interface* **15**(2018).
30. Marsh, J.A. & Teichmann, S.A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem* **84**, 551-75 (2015).
31. Monod, J., Wyman, J. & Changeux, J.P. On the Nature of Allosteric Transitions: A Plausible Model. *J Mol Biol* **12**, 88-118 (1965).
32. Perutz, M.F. Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726-39 (1970).
33. Reid, A.J., Yeats, C. & Orengo, C.A. Methods of remote homology detection can be combined to increase coverage by 10% in the midnight zone. *Bioinformatics* **23**, 2353-60 (2007).
34. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-30 (2014).
35. Andreeva, A. et al. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* **36**, D419-25 (2008).
36. Sillitoe, I. et al. CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res* **43**, D376-81 (2015).
37. Lees, J.G., Dawson, N.L., Sillitoe, I. & Orengo, C.A. Functional innovation from changes in protein domains and their combinations. *Curr Opin Struct Biol* **38**, 44-52 (2016).
38. Cheng, H. et al. ECOD: an evolutionary classification of protein domains. *PLoS Comput Biol* **10**, e1003926 (2014).
39. Schaeffer, R.D., Liao, Y., Cheng, H. & Grishin, N.V. ECOD: new developments in the evolutionary classification of domains. *Nucleic Acids Res* **45**, D296-D302 (2017).
40. Buljan, M. & Bateman, A. The evolution of protein domain families. *Biochem Soc Trans* **37**, 751-5 (2009).
41. Chothia, C., Gough, J., Vogel, C. & Teichmann, S.A. Evolution of the protein repertoire. *Science* **300**, 1701-3 (2003).
42. Das, S. & Smith, T.F. Identifying nature's protein Lego set. *Adv Protein Chem* **54**, 159-83 (2000).
43. Dessailly, B.H., Redfern, O.C., Cuff, A.L. & Orengo, C.A. Detailed analysis of function divergence in a large and diverse domain superfamily: toward a refined protocol of function classification. *Structure* **18**, 1522-35 (2010).
44. Galperin, M.Y. & Koonin, E.V. Divergence and convergence in enzyme evolution. *J Biol Chem* **287**, 21-8 (2012).
45. Cuff, A. et al. The CATH hierarchy revisited-structural divergence in domain superfamilies and the continuity of fold space. *Structure* **17**, 1051-62 (2009).
46. Reeves, G.A., Dallman, T.J., Redfern, O.C., Akpor, A. & Orengo, C.A. Structural diversity of domain superfamilies in the CATH database. *J Mol Biol* **360**, 725-41 (2006).
47. Das, S., Dawson, N.L. & Orengo, C.A. Diversity in protein domain superfamilies. *Curr Opin Genet Dev* **35**, 40-9 (2015).
48. Otten, R. et al. Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat Commun* **9**, 1314 (2018).
49. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203-7 (2009).
50. Aharoni, A. et al. The 'evolvability' of promiscuous protein functions. *Nat Genet* **37**, 73-6 (2005).

51. James, L.C. & Tawfik, D.S. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. *Trends Biochem Sci* **28**, 361-8 (2003).
52. Wouters, M.A., Liu, K., Riek, P. & Husain, A. A despecialization step underlying evolution of a family of serine proteases. *Mol Cell* **12**, 343-54 (2003).
53. Zou, T., Risso, V.A., Gavira, J.A., Sanchez-Ruiz, J.M. & Ozkan, S.B. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol Biol Evol* **32**, 132-43 (2015).
54. Wheeler, L.C., Lim, S.A., Marqusee, S. & Harms, M.J. The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol* **38**, 37-43 (2016).
55. Harel, M. et al. Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. *Nat Struct Mol Biol* **11**, 412-9 (2004).
56. Clifton, B.E. & Jackson, C.J. Ancestral Protein Reconstruction Yields Insights into Adaptive Evolution of Binding Specificity in Solute-Binding Proteins. *Cell Chem Biol* **23**, 236-245 (2016).
57. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem Mol Biol Educ* **43**, 206-9 (2015).
58. Ha, T. et al. Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci U S A* **93**, 6264-8 (1996).
59. Konermann, L., Pan, J. & Liu, Y.H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* **40**, 1224-34 (2011).
60. Hohlbein, J., Craggs, T.D. & Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem Soc Rev* **43**, 1156-71 (2014).
61. Kapanidis, A.N. et al. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* **101**, 8936-41 (2004).
62. Lerner, E. et al. Toward dynamic structural biology: Two decades of single-molecule Forster resonance energy transfer. *Science* **359**(2018).
63. Zheng, J., Strutzenberg, T., Pascal, B.D. & Griffin, P.R. Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry. *Curr Opin Struct Biol* **58**, 305-313 (2019).
64. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol* **286**, 279-90 (1999).
65. Henikoff, S., Haughn, G.W., Calvo, J.M. & Wallace, J.C. A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A* **85**, 6602-6 (1988).
66. Maddocks, S.E. & Oyston, P.C.F. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology (Reading)* **154**, 3609-3623 (2008).
67. Schell, M.A. Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* **47**, 597-626 (1993).
68. Albinia, A.M., Baglieri, J. & Robinson, C. Targeting of luminal proteins across the thylakoid membrane. *J Exp Bot* **63**, 1689-98 (2012).
69. Frain, K.M., Gangl, D., Jones, A., Zedler, J.A. & Robinson, C. Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochim Biophys Acta* **1857**, 266-73 (2016).
70. Natale, P., Bruser, T. & Driessen, A.J. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. *Biochim Biophys Acta* **1778**, 1735-56 (2008).
71. Orfanoudaki, G. & Economou, A. Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). *Mol Cell Proteomics* **13**, 3674-87 (2014).
72. UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).

73. Bagos, P.G., Nikolaou, E.P., Liakopoulos, T.D. & Tsirigos, K.D. Combined prediction of Tat and Sec signal peptides with hidden Markov models. *Bioinformatics* **26**, 2811-7 (2010).
74. Mao, B., Pear, M.R., McCammon, J.A. & Quijcho, F.A. Hinge-bending in L-arabinose-binding protein. The "Venus's-flytrap" model. *J Biol Chem* **257**, 1131-3 (1982).
75. Wermuth, C.G., Ganellin, C.R., Lindberg, P. & Mitscher, L.A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). in *Pure and Applied Chemistry* Vol. 70 1129 (1998).
76. Fulyani, F. et al. Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria. *Structure* **21**, 1879-88 (2013).
77. Wolters, J.C. et al. Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS One* **5**, e10361 (2010).
78. Maeda, S., Price, G.D., Badger, M.R., Enomoto, C. & Omata, T. Bicarbonate binding activity of the CmpA protein of the cyanobacterium *Synechococcus* sp. strain PCC 7942 involved in active transport of bicarbonate. *J Biol Chem* **275**, 20551-5 (2000).
79. Shilton, B.H. The dynamics of the MBP-MalFGK(2) interaction: a prototype for binding protein dependent ABC-transporter systems. *Biochim Biophys Acta* **1778**, 1772-80 (2008).
80. Gouridis, G. et al. Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat Struct Mol Biol* **22**, 57-64 (2015).
81. de Boer, M. et al. Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife* **8**(2019).
82. Lamblin, A.F. & Fuchs, J.A. Functional analysis of the *Escherichia coli* K-12 *cyn* operon transcriptional regulation. *J Bacteriol* **176**, 6613-22 (1994).
83. Lai, R.Y. et al. Thiamin pyrimidine biosynthesis in *Candida albicans* : a remarkable reaction between histidine and pyridoxal phosphate. *J Am Chem Soc* **134**, 9157-9 (2012).
84. Koropatkin, N.M., Koppelaar, D.W., Pakrasi, H.B. & Smith, T.J. The structure of a cyanobacterial bicarbonate transport protein, CmpA. *J Biol Chem* **282**, 2606-14 (2007).
85. Ramisetty, S.R. & Washburn, M.P. Unraveling the dynamics of protein interactions with quantitative mass spectrometry. *Crit Rev Biochem Mol Biol* **46**, 216-28 (2011).
86. Wales, T.E. & Engen, J.R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* **25**, 158-70 (2006).
87. Engen, J.R. & Smith, D.L. Investigating the higher order structure of proteins. Hydrogen exchange, proteolytic fragmentation, and mass spectrometry. *Methods Mol Biol* **146**, 95-112 (2000).
88. Englander, S.W. & Kallenbach, N.R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **16**, 521-655 (1983).
89. Tsirigotaki, A., Papanastasiou, M., Trelle, M.B., Jorgensen, T.J. & Economou, A. Analysis of Translocation-Competent Secretory Proteins by HDX-MS. *Methods Enzymol* **586**, 57-83 (2017).
90. Tina, K.G., Bhadra, R. & Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res* **35**, W473-6 (2007).
91. de Boer, M., Gouridis, G., Muthahari, Y.A. & Cordes, T. Single-Molecule Observation of Ligand Binding and Conformational Changes in FeuA. *Biophys J* **117**, 1642-1654 (2019).
92. Telmer, P.G. & Shilton, B.H. Insights into the conformational equilibria of maltose-binding protein by analysis of high affinity mutants. *J Biol Chem* **278**, 34555-67 (2003).
93. de Boer, M. Maximum likelihood analysis of non-equilibrium solution-based single-molecule FRET data. *physics.bio-ph* (2020).
94. Illergard, K., Ardell, D.H. & Elofsson, A. Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins* **77**, 499-508 (2009).
95. Elias, M. & Tawfik, D.S. Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem* **287**, 11-20 (2012).
96. Grishin, N.V. Fold change in evolution of protein structures. *J Struct Biol* **134**, 167-85 (2001).

97. Peisajovich, S.G., Rockah, L. & Tawfik, D.S. Evolution of new protein topologies through multistep gene rearrangements. *Nat Genet* **38**, 168-74 (2006).
98. Scheepers, G.H., Lycklama, A.N.J.A. & Poolman, B. An updated structural classification of substrate-binding proteins. *FEBS Lett* **590**, 4393-4401 (2016).
99. Cui, J. & Davidson, A.L. ABC solute importers in bacteria. *Essays Biochem* **50**, 85-99 (2011).
100. Falke, J.J. & Hazelbauer, G.L. Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem Sci* **26**, 257-65 (2001).
101. Higgins, C.F. ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* **8**, 67-113 (1992).
102. Zhang, Y. et al. Model of maltose-binding protein/chemoreceptor complex supports intrasubunit signaling mechanism. *Proc Natl Acad Sci U S A* **96**, 939-44 (1999).
103. Clifton, B.E. et al. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat Chem Biol* **14**, 542-547 (2018).
104. Bar-Even, A., Milo, R., Noor, E. & Tawfik, D.S. The Moderately Efficient Enzyme: Futile Encounters and Enzyme Floppiness. *Biochemistry* **54**, 4969-77 (2015).
105. Campbell, E. et al. The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol* **12**, 944-950 (2016).
106. Jiang, Y.L. et al. Coordinating carbon and nitrogen metabolic signaling through the cyanobacterial global repressor NdhR. *Proc Natl Acad Sci U S A* **115**, 403-408 (2018).
107. Jo, I. et al. Structural details of the OxyR peroxide-sensing mechanism. *Proc Natl Acad Sci U S A* **112**, 6443-8 (2015).
108. Monferrer, D. et al. Structural studies on the full-length LysR-type regulator TsaR from *Comamonas testosteroni* T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Mol Microbiol* **75**, 1199-214 (2010).
109. Muraoka, S. et al. Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *J Mol Biol* **328**, 555-66 (2003).
110. Pedre, B. et al. Structural snapshots of OxyR reveal the peroxidatic mechanism of H₂O₂ sensing. *Proc Natl Acad Sci U S A* **115**, E11623-E11632 (2018).
111. Sainsbury, S. et al. The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Res* **37**, 4545-58 (2009).
112. Kim, Y. et al. Crystal structure of the ligand-binding domain of a LysR-type transcriptional regulator: transcriptional activation via a rotary switch. *Mol Microbiol* **110**, 550-561 (2018).
113. Baskaran, K., Duarte, J.M., Biyani, N., Bliven, S. & Capitani, G. A PDB-wide, evolution-based assessment of protein-protein interfaces. *BMC Struct Biol* **14**, 22 (2014).
114. Bliven, S., Lafita, A., Parker, A., Capitani, G. & Duarte, J.M. Automated evaluation of quaternary structures from protein crystals. *PLoS Comput Biol* **14**, e1006104 (2018).
115. Nymeyer, H., Garcia, A.E. & Onuchic, J.N. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc Natl Acad Sci U S A* **95**, 5921-8 (1998).

MATERIALS AND METHODS

Structural analysis & Phylogenetic trees

Proteins harbouring the *cherry-core* (CC) structure were identified using the “Structure similarity search function” of the PDB web server.^{116,117} Protein structures were aligned using the PDBeFold server.¹¹⁸ Additionally, protein structures were downloaded and superimposed using Pymol (Molecular Graphics System, Version 1.6 Schrödinger, LLC) in order to verify and correct manually the results obtained via the PDBeFold server.

The structure-based phylogenetic trees as presented in SuppFile. 2, 6 and summarized in Fig. 1C were created for proteins with known structures using the “all against all structure comparison” feature of the DALI server.¹¹⁹ In the summarized sequence- and structure-based phylogenetic trees as presented in Fig. 1C and Fig. S1; the branch of each class has an average length of the branches of all proteins belonging to the very same class obtained from trees presented in SuppFile. 2, 6.

Sequence-based phylogenetic trees as presented in SuppFile. 1 and summarized in Figure S1 were computed using the MultiSeq software¹²⁰ embedded within the Visual Molecular Dynamics (VMD 1.9.2) software package.¹²¹ Briefly, these trees were based on the percent of sequence identity (PID) of polypeptide chains after the structure-based sequence alignment. Given the fact that the CC proteins (CCPs) share little amount of sequence conservation, MultiSeq was used to optimize the sequence alignment based on the structural alignment template. The alignments of MultiSeq have all been manually inspected.

The sequence-based extended phylogenetic tree as presented in SuppFile. 1 and summarized in Fig. S1 was made using sequence information of proteins with unknown structural information. We used the 53 CCPs with known structures to retrieve homologous protein sequences by inspecting the UniProtKB/Swiss-Prot (swissprot) database with Protein Blast.¹²² Our cut-off for sequence coverage was 70%, which yielded ~600 proteins that allowed to construct an extended sequence-based phylogenetic tree. In all analyses server's default settings were used unless specified.

Conformational states of the *cherry-core*

Domain movements were detected via the DynDom protein domain motion analysis server^{123,124} for all the proteins that have high resolution structural informational of both structural states available. Interactions between the C-tail and the domains were analysed by submitting the PDB ID codes corresponding to *apo* and *holo* states (Table. S2) in the Protein Interactions Calculator (PIC) webserver.¹²⁵ Similarly, the interactions between the N-terminal dimerization helix and the connecting-loop to the CC of class A proteins were retrieved and presented in SuppFile. 7. Standard cut-off distances for all interactions were used. The interactions were also manually inspected using Pymol (Molecular Graphics System, Version 1.6 Schrödinger, LLC). Cartoon protein representations, superimposition of structures and distance calculations between residues were accomplished using Pymol (Molecular Graphics System, Version 1.6 Schrödinger, LLC).

Gene isolation, protein expression and purification

His₁₀SBD2 (Class B), OpuACHis₆ (Class C), MalEHIS₆ (Class G) were expressed and purified as previously described.¹²⁶⁻¹²⁸ CC of *cynR* (Class A) gene (UniProtKB-P27111) was isolated from the genome of *Escherichia coli* DH5 α (F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ 80dlacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(rK-mK+), λ -). The primers were designed to exclude the HTH lysR-type DNA binding domain (1-58 amino acid); thus to include DNA sequence coding the protein chain 59-299 amino acid. Primers introduced *Nde*I and *Bam*HI restriction sites, and the gene product was sub-cloned into the pET16 vector (Novagen, EMD Millipore). For (full length CynR, the same procedure was adopted by including amino acids 1-58. CmpA (class E) gene (UniProtKB-Q55460) was isolated from the genome

of *Synechocystis* sp. (strain PCC 6803). The primers were designed to exclude signal peptide (1-35 amino acid); thus, to include DNA sequence coding the protein chain 36-452 amino acid.

Intergenic region of CynR operon was isolated from the genome of *Escherichia coli* BL21-DE3 (F⁻ ompT gal dcm lon hsdSB(rB⁻-mB⁻) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB⁺]K-12(λ S)) using a set of primers, i.e., the forward primers - X379 (5'- CCA TGT TCA GCC ACG GCA AGA AAA TAA TTG ATA TG -3') and the reverse - X380 (5'- CTG GAA TTT AAG GAA TCC ATC AAT AAT CTC TTT CAC CG -3'). The amplification of the intergenic region results in the DNA fragment that includes 191 bp, which subsequently was used as the binding partner of CynR for further experiments.

Protein derivatives having the cysteine mutations or the point mutations indicated into brackets throughout the manuscript were constructed using QuickChange mutagenesis¹²⁹ and Megaprimer PCR mutagenesis¹³⁰ protocols.

His₁₀CynR (CC) was over-expressed in BL21 DE3 cells (F⁻ ompT gal dcm lon hsdSB(rB⁻-mB⁻) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB⁺]K-12(λ S)). Cells harboring plasmid expressing the protein were grown in Terrific Broth medium (30°C; OD_{600 nm}=0.5) and protein overexpression was induced by IPTG (0.10 mM; growth at 16°C for 15 hours).

His₁₀CmpA was over-expressed in BL21 pLysS DE3 cells (F⁻, ompT, hsdS_B (rB⁻, mB⁻), dcm, gal, λ (DE3), pLysS, Cmr^r). Cells harboring plasmid expressing the protein were grown (30°C; OD_{600 nm}=0.5) and protein overexpression (growth at 16°C for 15 hours) was induced by IPTG (0.1 mM). For all proteins growth was accomplished in Luria Bertani medium, except for CynR in Terrific Broth.

Proteins were purified as follows: after lysis by cell disruptor (30.000 psi; 2 rounds), soluble material (50.000 g; 30 min; 4 °C; Sorval) was loaded (50 mM Tris-HCl, pH = 8; 1 M KCl, 10% glycerol; 10 mM Imidazole; 1 mM DTT; 2 mM PMSF) on a Ni-NTA resin (Qiagen). Bound proteins were washed (50 mM Tris-HCl, pH = 8; 50 mM KCl, 10% glycerol; 10 mM Imidazole; 1 mM DTT; and 50 mM Tris-HCl, pH=8; 1 M KCl, 10% glycerol; 30 mM Imidazole; 1 mM DTT sequentially) and then eluted (50 mM Tris-HCl, pH=8; 50 mM KCl, 10% glycerol; 300 mM Imidazole; 1 mM DTT). Protein fractions were pooled (supplemented with 5 mM EDTA; 50 mM DTT), concentrated (Vivacell-Sartorius; Amicon-Millipore), dialyzed (50 mM Tris-HCl, pH=8; 1 M KCl, 50% glycerol; 10 mM DTT), aliquoted and stored at -80°C.

For untagged CynR (Full length): after lysis by cell disruptor (30.000 psi; 2 rounds), soluble material (50.000 g; 30 min; 4 °C; Sorval) was loaded (50 mM Tris-HCl, pH = 8.3; 5mM NaCl, 3 mM EDTA; 1 mM DTT; 2 mM PMSF) on a Q sepharose fast flow (GE Healthcare) and incubated over-night. Immobilized proteins were washed (50 mM Tris-HCl, pH 8.3, 135 mM NaCl, 3mM EDTA, 1mM DTT) and eluted (50 mM Tris-HCl, pH 8.3, 250 mM NaCl, 3mM EDTA, 1mM DTT). Protein fractions were pooled and concentrated (10.000 MWCO Amicon; Merck-Millipore) up to 5 mL and analyzed on a HiLoad 26/60 Superdex 200 (GE) (equilibrated with 50 mM Tris-HCl, pH 8.3, 5 mM NaCl, 0,1mM EDTA). Fraction were analyzed on SDS-PAGE and those containing pure CynR were collected and re-subjected to Size-exclusion chromatography to increase purity (>90%). Fractions were collected, dialyzed (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 50% glycerol, 0,5mM EDTA and 1 mM DTT), aliquoted and stored at -20°C.

Protein labelling

Protein labelling was accomplished as described previously.¹²⁶ In brief, cysteine positions were chosen based on the open and closed x-ray crystal structures of CynR (2HXR, 3HFU), OpuAC (3L6G, 3L6H), SBD2 (4KR5, 4KQP), CmpA (2I49, 2I4C) and MalE (1OMP, 1ANF); Table. S5. Stochastic labelling was performed with the maleimide derivative of dyes Cy3B (GE Healthcare) and ATTO647N (ATTO-TEC) for OpuAC and SBD2. CmpA, CC of CynR and MalE were labelled with Alexa555 and Alexa647 (ThermoFischer). Histidine tagged proteins were immobilized on a Ni²⁺-sepharose resin (GE Healthcare) to remove the 1 mM DTT (Dithiothreitol) used to keep all cysteine residues in a reduced state. The resin was incubated 2-8 hours at 4°C with 50 nmol of each fluorophore dissolved in the appropriate buffer (50 mM Tris-HCl, pH=7.4; 50 mM KCl for CynR (CC), CmpA and MAIE; 50 mM KPi, pH=7.4; 50 mM KCl for SBD2 and OpuAC) and subsequently washed to remove the majority of unbound fluorophores. The labelled protein was further

analysed by size-exclusion chromatography (Superdex 200, GE Healthcare) to enrich the double-labelled fraction and remove potential aggregation material¹³¹. For all proteins, labelling efficiency was higher than 80%.

Isothermal titration calorimetry (ITC)

Purified CynR was dialyzed overnight against 50 mM Tris-HCl, pH=7.4; 50 mM KCl. ITC experiments were carried out in a Nano ITC Low Volume (TA Instruments). The sodium azide stock solution (250 μ M) was prepared in the dialysis buffer and was stepwise injected (0.5 μ l) into the reaction cell containing 12 μ M CynR. All experiments were carried out at 298 K with a mixing rate of 250 r. p. m. Data were analysed with a single-binding site equation, provided by the NanoAnalyse software (TA Instruments).

Purified CmpA was dialyzed overnight against 1X PBS Buffer that has been supplemented with 50uM K_2CO_3 and 1uM EDTA. ITC experiments of CmpA were performed using MicroCal iTC200 (Malvern). The Calcium solution (500uM) was prepared from $CaCl_2$ powder, which subsequently diluted in the dialysis buffer and was injected (2 μ l) into the reaction cell containing 40 μ M of CmpA. All experiments were carried out at 297 K with a mixing rate of 750 rpm. Data were analysed with a single-binding site equation, provided by the provided MicroCal Analysis software (Malvern).

Solution-based smFRET and ALEX

ALEX experiments were carried out at 25-50 pM of double-labelled protein in the appropriate buffer (50 mM Tris-HCl, pH=7.4; 50 mM KCl for CynR, CmpA and MalE; 50 mM KPi, pH=7.4; 50 mM KCl for SBD2 and OpuAC) supplemented with additional reagents as stated in the text. The experiments were performed using a home-built confocal microscope similar to the setup described before¹²⁶ with minor modifications as outlined below. Briefly, two laser-diodes (Coherent Obis) with emission wavelength of 532 and 637 nm were modulated in periods of 50 μ s and used for confocal excitation. Alternation between both excitation wavelengths was achieved by direct modulation of the two laser. The beam of both lasers was coupled into a single-mode fiber (PM-S405-XP, Thorlabs) and collimated (MB06, Q-Optics/Linos) before entering a water immersion objective (60X, NA 1.2, UPlanSAPO 60XO, Olympus). The excitation spot was focused 20 μ m above the interface of glass and water solution. Typical average laser powers were 30 μ W at 532 nm ($\sim$ 30 kW/cm²) and 15 μ W at 637 nm ($\sim$ 15 kW/cm²). Excitation and emission light were separated by a dichroic beam splitter (zt532/642rpc, AHF Analysentechnik), mounted in an inverse microscope body (IX71, Olympus). Emitted light was focused onto a 50 μ m pinhole and spectrally separated (640DCXR, AHF Analysentechnik) onto two APDs (τ -spad, <50 dark-counts/s, Picoquant) with appropriate spectral filtering (donor channel: HC582/75; acceptor channel: Edge Basic 647LP; both AHF Analysentechnik). Photon arrival times were registered by an NI-Card (PXI-6602, National Instruments). A dual-colour burst search¹³² using parameters $M = 15$, $T = 500 \mu$ s and $L = 25$ was applied to identify bursts. Additional thresholding was applied to remove spurious changes in fluorescence intensity and selected for intense single-molecule bursts (total photons per burst > 150 unless otherwise mentioned). Binning the detected bursts into a 2D apparent FRET/S histogram (61 x 61 bins unless otherwise mentioned) allowed the selection of the donor and acceptor labelled molecules.¹³³ The selected apparent FRET histograms (61 x 1 bins unless otherwise mentioned) were fitted using a Gaussian function.

Single-molecule bursts from donor-only labelled CynR and MalE proteins were obtained by selecting the donor-only subpopulation ($S > 0.9$) from the 2D apparent FRET/S histogram. The total photon count per burst were normalised by its respective duration to obtain the photon count rate.

Scanning confocal microscopy

Confocal scanning experiments were performed using the same homebuilt confocal microscope as described before.¹²⁶ Surface scanning was performed using a piezo stage with 100x100x20 μ m scanning range in XYZ, respectively (P-517-3CD with E-725.3CDA, Physik Instrumente). The detector signal was registered using a Hydra Harp 400 picosecond event timer and a module for time-correlated single photon counting (both Picoquant). Data were recorded with constant 532-nm excitation at an intensity of 0.5 μ W

(~125 W/cm²) for SBD2 and MalE, but 1.5 μ W (~400 W/cm²) for OpuAC. Data acquisition was done with home-written Labview software (National Instruments). A flow-cell arrangement was used for studies of surface-immobilized OpuAC and SBD2; whereas MalE was studied on standard functionalized coverslips. Flow-cells were assembled using established protocols described before.^{126,134} All experiments of OpuAC and SBD2 were carried out in a degassed buffer (50 mM KPi, pH=7.4; 50 mM KCl) under oxygen-free conditions obtained utilizing an oxygen-scavenging system supplemented with 10 mM of aged Trolox as a photostabilizer.¹³⁵ For MalE, experiments were carried out in buffer (50 mM Tris-HCl, pH=7.4; 50 mM KCl) supplemented with 1 mM Trolox.¹³⁵

Time-traces were analysed by integrating the detected red and green photon streams in time-bins of 1.5 ms (OpuAC) or 5 ms (MalE and SBD2). Only traces lasting longer than 50 time-bins having on average more than 10 photons per time-bin were used for further analysis. The apparent FRET per time-bin was calculated by dividing the red photons by the total number of photons per time-bin. The most probable state-trajectory was identified by Hidden Markov Modeling (HMM)¹³⁶. For this an implementation of HMM was programmed in Matlab (MathWorks), based on previous studies.¹³⁶ We assume that the FRET time-trace (the observation sequence) can be considered as a HMM with only two states having a one-dimensional Gaussian-output distribution. The aim of a HMM is to infer both the states and state-transition probabilities given the observation sequence. The Gaussian output-distribution of a state i ($i = 1, 2$) is completely defined by two parameters: the average and the variance. The HMM algorithm finds all parameters λ , given only the observation sequence, that maximizes the likelihood function.¹³⁶ This was done with the forward-backward algorithm.¹³⁷ Care was taken to avoid floating point underflow and was done as described.¹³⁶ The most probable state-trajectory is then found from λ by using the Viterbi algorithm.¹³⁸ The time spend in each state is interfered from the most probable state-trajectory and is summarized into a histogram for all traces under the same condition. The histogram was fitted with a single exponential decay to obtain the closing rate and closed-state lifetime.

Hydrogen/Deuterium exchange Mass Spectrometry (HDX-MS)

Isotope labeling: MalE and derivative MalE (M321K) were incubated for 30 min at 25°C [5 μ l of 28.25 μ M protein stock in 50 mM Tris-HCl pH 7.4, 50 mM KCl to ensure saturation and to thermally equilibrate the samples. At 30 min of incubation, MalE was isotopically labeled with 93.75% v/v final D content at 25°C, by addition of 75 μ l deuterated buffer [50 mM Tris-DCI pD 7.4, 50 mM KCl in D₂O (D₂O, 99.9% atom D; Eurisotop)] for 10, 100, 1000, 10000 and 100000 sec. pD refers to the corrected value for the isotope effect. The HDX reaction was quenched at the defined time intervals by instant acidification [pD 2.5; formic acid; (Ultra-pure) from Merck KGaA]. The pre-chilled quenching solution contained urea (Urea-d₄; Sigma) to a final concentration of 1.6 M at quenching to increase the peptide coverage by mild denaturation. 50 pmoles of protein were injected for analysis.

The CC of CynR (92-298) and the tetrameric DNA-free CynR (1-298) (isolated by SEC) were concentrated to 28.6 μ M and supplemented with 5 mM DTT. CynR (4 μ l; 28.6 μ M) was incubated at 25°C for 30 min in the presence or absence of NaN₃, by addition of 1 μ l of 0.5 M NaN₃ diluted in 50 mM Tris-HCl pH 7.4, 50 mM KCl. For the apo states, 1 μ l of buffer was added instead. Deuterated buffer (95 μ l; as for MalE) was added for 10, 100, (1000/ 2000 for the CC and CynR respectively), 10000 and 100000 (for the CC) sec at 25°C, resulting in 95% v/v D₂O, 1.14 μ M FL CynR, 5 mM NaN₃. The isotope labeling reactions were quenched by acidification (pD_{corrected} 2.5) with pre-chilled quenching solution containing urea (Urea-d₄; Sigma) to a final urea concentration of 1.52 M (0.73 μ M protein). The samples were instantly frozen in liquid nitrogen and analyzed by instant thawing within two days of preparation. 36 pmoles of protein were injected for analysis.

For the analysis of the DNA-bound CynR, a different approach was followed to minimize the signal suppression during the mass spectrometric analysis induced by the addition of DNA. Specifically, FL CynR (2 μ l; 32.85 μ M protein; 5 mM DTT) was supplemented with 1.2 molar excess of DNA in H₂O (1.8 μ l of

43.99 μM) and incubated at 25°C for 5 min prior to addition of 0.6 μl of 0.5 M NaN_3 (in H_2O) for the liganded state, or with 0.6 μl of nanopure H_2O for the apo state. The dilution of Tris and KCl concentrations by addition of H_2O at this step was corrected by increased their concentrations in the protein stock. The samples were incubated for an additional 25 min at 25°C prior to addition of deuterated buffer (55.6 μl ; as above) for 10, 1000, 10000 sec at 25°C, resulting in 92.7% v/v D_2O , 1.1 μM CynR, 1.3 μM DNA, 5 mM NaN_3 . The isotope labeling reactions were quenched by acidification ($\text{pD}_{\text{corrected}}$ 2.5) with 4.2 excess of pre-chilled quenching solution containing deuterated urea to a final concentration of 1.52 M (250 μl total volume of quenched reaction; 0.26 μM CynR; 0.32 μM DNA) and loaded on centrifugal concentrators (Vivaspin 500; 10,000 MWCO PES, Sartorius). The quenched samples were concentrated to 60 μl by centrifugation at 20,000xg for 2 min (Sigma 1-16K). Due to the linear nature of DNA, it is not retained in the centrifugal membrane and therefore only the protein is concentrated (1.1 μM CynR; 0.32 μM DNA). The retained solution in the centrifugal apparatus was transferred to a pre-chilled tube and was frozen in liquid nitrogen.

Online proteolysis-LC-MS analysis: The quenched samples were injected into a nanoACQUITY UPLC System with HDX technology (Waters, UK), thermostated at the digestion and LC separation chambers at 20°C and 0.8°C (2°C for CynR) respectively. Proteolytic digestion (Enzymate BEH pepsin column, Waters) and peptide trapping/desalting (ACQUITY UPLC R BEH C18 VanGuard pre-column; 130 Å, 1.7 μm , 2.1 x 5 mm; Waters) were performed with 0.23% formic acid in H_2O (Solvent A) at 100 $\mu\text{l}/\text{min}$ for 3 min, online with peptic peptide separation (ACQUITY UPLC R BEH C18 analytical column; 130 Å, 1.7 μm , 1 x 100 mm; Waters) at 40 $\mu\text{l}/\text{min}$ using a 12 min (8 min for CynR) linear gradient from 5% to 50% Solvent B [ACN (Optima LC/MS grade; Fischer Scientific), 0.23% formic acid]. The eluate was analyzed online on a Synapt G2 ESI-Q-TOF instrument (Waters, UK) with a MassLynX interface (version 4.1 SCN870; Waters) for data collection. The source/TOF conditions were set as: resolution mode, capillary voltage 3.0 kV, sampling cone voltage 20 V, extraction cone voltage 3.6 V, source temperature 80°C, desolvation gas flow 500 L/h at 150°C. The deuterated samples were analyzed in MS acquisition mode (300-2000 Da range), while for peptide identification non-deuterated samples (treated as above but in protiated buffers) were analyzed in MS^E acquisition mode over the m/z range 100-2,000 Da, using a collision energy ramp from 10 to 30 V. Leucine Enkephalin (2 ng/ μl in 50% ACN, 0.1% formic acid; 5 $\mu\text{l}/\text{min}$) was co-infused in both acquisition modes for accurate mass measurements (reference mas: m/z 556.2771).

Data analysis: For peptide identification, MS^E data were processed on the ProteinLynx Global Server (PLGS v3.0.1, Waters, UK), using a user-defined database containing MalE, MalE (M321K), the CC CynR and CynR sequences under the following criteria: digestion enzyme, non-specific; false discovery rate, 4%, minimum fragment ion matches/peptide and /protein, 3 and 7; minimum peptide matches/protein, 1; low and elevated energy thresholds, 150 and 25 counts; intensity threshold: 500 counts, reference mass correction window, 0.25 Da at 556.2771 Da/e. The identified peptides from two independent MS^E raw files were further filtered on the DynamX software (version 3.0, Waters) for DynamX score > 7.5 (7 for CynR), maximum MH^+ error of 5 ppm and minimum products/amino acid of 0.2. Only robustly identified peptides in both replicates were further processed, resulting in 94.3, 93.8%, 93.4% and 88.3% sequence coverage for MalE, MalE (M321K), the CC CynR and full-length CynR respectively. Deuterium uptake was determined using the DynamX software. For the comparison of the WT and MalE (M321K) at the mutated region, the peptide containing residues 321-337 was compared only at the level of D-uptake relative to the full deuteration of each state, after complete analysis of each protein with its individual controls, and not at the level of absolute D-uptake. For this, MalE (M321K) was analyzed using the peptide list created from the MalE (M321K) sequence, the MalE (M321K) non-deuterated and full deuterated controls (Table F5B). Additionally, for the comparison of MalE (M321K) at the apo state, statistical analysis was performed both on the data analyzed using the WT and MalE (M321K) sequence in order to include the mutated region (Tables F5B, F5C, F5D).

Statistical analysis: All HDX reactions were performed in triplicates, unless otherwise indicated (Table F5A, F4A, F4B, F4C). Statistical analysis of the significance of differences between the apo and liganded states, and between the WT and MalE (M321K) mutated sequence in the case of MalE, was achieved using a modified approach of Bennett *et al.*,¹³⁹ described in Tsirigotaki *et al.*,^{140,141} Briefly, two-tail paired t-tests, comparing the mean uptake, as absolute deuterium uptake values, of the two states for each peptide, were performed using R language and the significance threshold was set to 99% confidence ($1-p \geq 0.99$). An additional threshold was set at $\pm 4SD$ of average pooled standard deviations of both states. Finally, a third criterion was introduced to exclude false positives due to high SD outliers that are averaged out in the pooled SDs: The difference between the two states for each peptide must exceed twice the sum of SDs of the two states for the given peptide. Only differences that fulfill all three criteria were considered as statistically significant. Visualization of the statistical analysis in scatter plots was achieved by R language (Table F5D, F4A, F4B and F4C). For optimal realization of the extent of the ΔD -uptake regardless of the length of the identified peptides, the ΔD -uptake of each peptide was expressed relative to the experimentally determined full deuteration control (normalized ΔD -uptake). For higher robustness, only statistically significant differences that exceed the absolute value of 10% normalized ΔD -uptake are discussed. Detailed data sets including smaller, statistically significant differences are given in Tables F5C (for WT and MalE (M321K) apo states), F5E (for WT MalE apo and holo states) F4A (for DNA-bound CynR apo and holo states), F4B (for DNA-free CynR apo and holo states) and F4C (for the CC CynR apo and holo states).

Bulk Fluorescence measurements

Measurements were carried out in quartz cuvettes (1 ml; Hellma) with a Cary Eclipse fluorimeter supplemented with a four-position cuvette holder and a Peltier temperature controller (Varian). To determine an apparent melting temperature T_m (apparent), the changes in intrinsic tryptophan fluorescence emission of CynR (0.3 μM) as a function of increasing temperature (10 - 70 $^{\circ}C$; ramping rate 0.8 $^{\circ}C/min$; excitation 297 nm/emission 345 nm; slits at 2.5 and 20 nm; data acquisition interval = 0.5 min), in the presence or absence of NaN_3 (1 mM). All data were collected using Cary Eclipse software (Bio Package; Varian) and analyzed by nonlinear regression using Origin 5.0 (Microcal).

Homology modelling method

For establishing the three-dimensional structure and the different oligomeric states of CynR, its amino acid sequence has been modelled according to known structures having a substantial degree of sequence similarity that were used as templates¹⁴². The amino acid sequence of CynR was submitted to the SWISS-MODEL in the Automated Protein Modelling server provided by the Glaxo Smith Kline center (Geneva, Switzerland) using the standard settings of the server¹⁴³. All models were reliable according to server thresholds¹⁴⁴⁻¹⁴⁶.

REFERENCES

116. Prlic, A. et al. Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**, 2983-5 (2010).
117. Ye, Y. & Godzik, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19 Suppl 2**, ii246-55 (2003).
118. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68 (2004).
119. Holm, L. & Laakso, L.M. Dali server update. *Nucleic Acids Res* **44**, W351-5 (2016).
120. Roberts, E., Eargle, J., Wright, D. & Luthey-Schulten, Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**, 382 (2006).

121. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-8, 27-8 (1996).
122. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-10 (1990).
123. Poornam, G.P., Matsumoto, A., Ishida, H. & Hayward, S. A method for the analysis of domain movements in large biomolecular complexes. *Proteins* **76**, 201-12 (2009).
124. Taylor, D., Cawley, G. & Hayward, S. Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics* **30**, 3189-96 (2014).
125. Tina, K.G., Bhadra, R. & Srinivasan, N. PIC: Protein Interactions Calculator. *Nucleic Acids Res* **35**, W473-6 (2007).
126. Gouridis, G. et al. Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nat Struct Mol Biol* **22**, 57-64 (2015).
127. Wolters, J.C. et al. Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS One* **5**, e10361 (2010).
128. de Boer, M. et al. Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife* **8**(2019).
129. Bok, J.W. & Keller, N.P. Fast and easy method for construction of plasmid vectors using modified quick-change mutagenesis. *Methods Mol Biol* **944**, 163-74 (2012).
130. Vander Kooi, C.W. Megaprimer method for mutagenesis of DNA. *Methods Enzymol* **529**, 259-69 (2013).
131. Husada, F. et al. Watching conformational dynamics of ABC transporters with single-molecule tools. *Biochem Soc Trans* **43**, 1041-7 (2015).
132. Nir, E. et al. Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J Phys Chem B* **110**, 22103-24 (2006).
133. Kapanidis, A.N. et al. Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* **101**, 8936-41 (2004).
134. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat Methods* **5**, 507-16 (2008).
135. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the mechanism of Trolox as antiblinking and antibleaching reagent. *J Am Chem Soc* **131**, 5018-9 (2009).
136. Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition. in *Readings in speech recognition* (eds. Alex, W. & Kai-Fu, L.) 267-296 (Morgan Kaufmann Publishers Inc., 1990).
137. Baum, L.E. & Petrie, T. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. 1554-1563 (1966).
138. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* **13**, 260-269 (1967).
139. Bennett, M.J., Barakat, K., Huzil, J.T., Tuszynski, J. & Schriemer, D.C. Discovery and characterization of the laulimalide-microtubule binding mode by mass shift perturbation mapping. *Chem Biol* **17**, 725-34 (2010).
140. Tsirigotaki, A., Elzen, R.V., Veken, P.V., Lambeir, A.M. & Economou, A. Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry. *Sci Rep* **7**, 2456 (2017).
141. Tsirigotaki, A., Papanastasiou, M., Trelle, M.B., Jorgensen, T.J. & Economou, A. Analysis of Translocation-Competent Secretory Proteins by HDX-MS. *Methods Enzymol* **586**, 57-83 (2017).
142. Bordoli, L. et al. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**, 1-13 (2009).
143. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201 (2006).
144. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* **7**, 10480 (2017).

145. Bienert, S. et al. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **45**, D313-D319 (2017).
146. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).

SUPPORTING INFORMATION

Figure S1: Phylogenetic Trees of *cherry-core* proteins, related to Figure 1. (A) Summarized sequenced-based extended phylogenetic tree of the bilobed *cherry-core* proteins (CCPs), grouped with respect to their ligands. Complete tree of ~ 600 proteins is shown in SuppFile. 1. Function was obtained from UniProtKB¹ and is color-coded (Green: transcription factors; Red: enzymes; Blue: substrate binding proteins). Localization is denoted by a symbol (‡: DNA-bound; Δ: extracellular; Ω: cytoplasm; Σ: chloroplast). Ligands or system effectors are indicated and chemical structure is presented in Fig. S3. **(B-D)** Summarized (complete trees in SuppFile. 2) sequenced-based phylogenetic trees of the 53 identified CCPs with known structures for the entire polypeptide chain **(B)**, for D1, D2 **(C)**, and for the C-tail only **(D)**. An asterisk (*) marks structural subclasses (see SuppFile. 3). A similar clustering is observed in all above trees and comparable to the structure-based phylogenetic tree (Fig. 1C). This indicates that proteins having a class-specific rigid domain (D1 and D2) have evolved to have the class-specific C-tail. This

SUPPORTING INFORMATION

outcome is in line with the fact that the C-tail–domain interactions stabilize a define structural state (Fig. 2 and Table. S2)

Two major clades are encountered in the structure-based phylogenetic trees (Fig. 1C). The ‘lower clade’, which includes classes F and G, manifest substantial alterations of the CC, represented primarily by insertions of secondary structural elements (see also Fig. 1D and Fig. S2). Our secondary structural alignments indicated that the most pronounced alterations in classes F, G involve: the addition of α -helix 3 (α_3), α -helix 4 (α_4) and the deletion of β -sheet 5 (β_5). On the other hand, the ‘upper clade’ includes CCPs with minor alterations/variability of their CC. In this study, we selected to analyse representative examples to describe faithfully the phylogenetic relationships: From the lower clade, MalE of Class G has been selected as it fulfils the following criteria: (i) belongs to the class with the most complex C-tail (Fig. 1C, D) of the CCPs and (ii) includes some of the most well-studied proteins with abundant structural information available in the Protein Data Bank. OpuAC of Class C has been selected as it belongs to the first diversified branch of the ‘upper clade’. Moreover, SBD2 of Class B has been picked regarding its long evolutionary distance with the previously selected ‘upper clade’ cousin (i.e. OpuAC). CynR (Class A) and CmpA (Class E) have also been analysed experimentally as the only ones with no Tier-0 dynamics (Fig. 2 and Table. S1).

Figure S2: Structural Consensus of *cherry-core* proteins, related to Figure 1B-D. (A) Consensus/model derived from secondary structure alignments of the CC of CCPs (Fig. 1D and SuppFile. 3). **(B)** Schematics of the geometrical consensus for the 3D placement of secondary structure elements originated from the 3D Structural analysis (manually inspected by Pymol). The unique placement of such elements in the 3D space provides a remarkable degree of symmetry of D1 & D2 of the CCPs forming the binding site.

SUPPORTING INFORMATION

Figure S3. Common Chemical Structure and pharmacophore structure of the *cherry-core* protein ligands, related to Figure 2. The common chemical structure of all identified ligands for each of the different CCP classes (A-G) is highlighted with red color. The ligands of class A, but mostly class E do not share a substantial degree of a chemical structure. In addition, pharmacophore analysis, after flexible alignment (by using the MOE software) of the ligands for every class, reveals the common interactions which are represented with brown (hydrogen bond donor-acceptor, HBD-A), blue (hydrogen bond acceptor, HBA) or green (lipophilic) spheres. Specifically, class A shares 1 interaction (HBD-A); class B, 4 interactions (3 HBD-A and 1 HBA); class C, 3 interactions (2 HBA and 1 lipophilic); class D, 6 interactions (HBA); class E, 0 interactions; class F, 7 interactions (HBA) and class G, 3 interactions (HBA).

SUPPORTING INFORMATION

Figure S4: Structural Analysis of the oligomeric transcription factor CynR and of the single-turnover enzyme pyrimidine synthase THI5p, related to Figure 2, 4. (A) Crystal structure (PDB:4ESX) of the single-turnover enzyme thiamin pyrimidine synthase² in complex with its ligand pyridoxal phosphate (red sticks). The 2 symmetric helices of the C-tail that stabilize the cleft in the closed structure and render it rigid, are colored as in Fig. 2. The extreme rigidity required is evidenced by the fact that pyridoxal phosphate needs to bind to the cleft of the pyrimidine synthase of *C. albicans* with a very specific geometry (provided also by the covalent interaction with K62, forming a Schiff base) with respect to the side chain of H66 (green sticks). This allows the “remarkable” chemistry to occur, as stated by the authors of the corresponding paper. The reaction consists of the excision of the side chain of H66 to produce thiamine pyrimidine. **(B)** Crystal structure of the *holo* state of the CC of CynR (PDB: 3HFU) with azide (indicated with red spheres) in the binding cleft. For clarity, the second protomer present within the asymmetric unit, was omitted. **(C)** The *holo* state of the CC CynR (PDB: 3HFU, grey color) was superimposed with the *apo* CC CynR (PDB: 2HXR, green color). The dimeric assembly was classified to be biologically relevant by EPPIC. Both the protomer structure, and the dimeric assembly do not undergo significant structural changes and are perfectly super-imposable. **(D)** Schematic representation of the CC of CynR. The conserved CC helix (Fig. S2) $\alpha 3$ is missing from the CC of CynR and all class A proteins (Fig. 1D, SuppFile. 3). **(E)** Schematic representation of the dimeric CC of class A proteins. Interacting secondary structural elements are highlighted. $\alpha 1$ makes hydrophobic contacts with S5 and $\alpha 5$, while S2 hydrophobic contacts with $\alpha 5$ and hydrophobic contacts/hydrogen bonds with S5.

SUPPORTING INFORMATION

Figure S5. Conformational dynamics of SBD2 and derivatives, related to Figure 5 (A-D). **(A)** Apparent FRET efficiency histogram of freely diffusing SBD2 (L480A) labelled with the donor fluorophore Alexa Fluor 555 and acceptor fluorophore Alexa Fluor 647. The fraction of closed state is similar as found for SBD2 (L480A) labelled with the donor fluorophore Cy3B and the acceptor fluorophore Atto647N, as presented in Fig. 5C. **(B)** The apparent FRET efficiency for both the *apo* (open) and *holo* (closed) state of SBD2 (L480A) is found to differ slightly from SBD2, which would suggest that the energy landscape is altered. **(C)** To exclude that the presence of endogenous ligand is responsible for the fraction closed observed, FRET efficiency histograms were measured under different sample concentrations. By diluting the sample, the ligand concentration would decrease, and hence the fraction of closed observed should decrease if it was due to ligand contamination. These results show no change in fraction closed over almost a 10-fold increase in dilution, indicating that the fraction closed is not caused by endogenous ligand.

SUPPORTING INFORMATION

Figure S6. Conformational dynamics of MalE and derivatives, related to Figure 5 E-I. (A) Apparent FRET efficiency histogram of MalE, obtained from the solution-based smFRET and ALEX measurements under conditions as indicated. (B) Fluorescence trajectories of MalE under different conditions as indicated; donor (green) and acceptor (red) photon counts are binned with 5 ms. The top panel shows calculated apparent FRET efficiency (blue) with the most probable state-trajectory of HMM (black). (C, D) Waiting time distribution of the low (C) and high FRET state (D) as obtained from the most probable state-trajectory of the HMM of all molecules per condition. Grey bars are the binned data and the solid line is the fit. (E) Mean closing rate (black) and closed-state lifetime (purple) as function of maltose concentration obtained from the fits in (C) and (D). Error bars indicate the 95% confidence interval of the fit obtained from the fit in (C) and (D). (F, G): Apparent FRET efficiency histogram of MalE (M321A) (F) and MalE (M321K) (G). (H) Time dependence of apparent FRET efficiency histograms of MalE (M321K) in the presence of 10 nM maltose and 20 μ M unlabeled MalE (M321K) added to scavenge all maltose. First 10 nM maltose was added to saturate the protein (top panel). Subsequently $\sim$ 20 μ M unlabeled protein was added to scavenge all ligands that are released from the labelled protein. Protein conformation was probed at indicated time points. Grey bars are the data and solid line is the fit. (I, J) Fluorescence trajectories of MalE (M321A) (I) and MalE (M321K) (J) obtained from surface immobilized molecules in the *apo* ($\sim$ 20 μ M unlabeled protein) or 1 mM maltose conditions. Fluorescence trajectories of MalE derivatives are plotted similarly to (B).

SUPPORTING INFORMATION

Figure S7: Model of Possible Oligomeric States of CynR, related to Figure 4. The CynR sequence (UniprotKB: P27111) was modelled from the indicated PDB structures that were used as a template, by the SWISS-MODEL server. The DNA in all structures was modelled from the structure of the HTH-DNA binding domain of CbnR in complex with the RBS of the *cbnA* promoter³. (A-D) Each tetramer is a dimer of dimers. The dimerization interface is formed by the long α -helix connecting the CC to the DNA binding domain (see also Fig. 4A, B). The α -helical coil-coil is stabilized always (i.e. in all known high-resolution structures of the HTH-motif transcription factors) by

SUPPORTING INFORMATION

hydrophobic interactions. In one of the protomers, the helix is present in a compact configuration (i.e. it bends toward the CC), while in the other it is extended. The compact dimers are always represented by cold colors (blue-cyan), whereas the extended with warmer colors (green, yellow). The tetramerization interface is provided by two antiparallel CCs (Fig. S4B-E), one of them in the compact and the other in the extended configuration. By superimposing the CC of all protomers, we notice differences in the arrangement of the dimerization helix with respect to the CC between the compact or the extended protomers. This is clearly evident in the contact map (SuppFile. 7) of the protomers depicting the interactions between the dimerization helix and the CC, as such interactions vary even between the protomers having the same arrangement (either compact or extended). The dimerization helices are in the same plane, forming an intelligible rectangle, indicated in the 3D representation column. The two compact dimers are in one side on the rectangular plane, whereas the other two extended on the other (**3D Representation Column**). In some assemblies (4X6G and 6G1B), the quaternary structure is stabilized via interactions between secondary structure elements belonging to the compact dimers. Such elements ($\alpha 4$ and CH2, see SuppFile. 3) are colored with pink and magenta color respectively in all assemblies. The DNA is always present on the side of the two extended protomers. (**E, F**): In these oligomers, all protomers are in the compact configuration, thus colored with cold colors. The DNA runs parallel to the intelligible rectangle denoting the plane of the dimerization helices. (**G**) Here two of the protomers are in the compact and two in the extended configuration alike A-D. The difference is that in the compact protomers the loop connecting the dimerization helix to the CC is rotated 180° with respect to the compact protomers of **A-D**, placing the DNA binding domain below the CC. In this case all CCs are in the same plane with that of the dimerization helices. (**H**) Here, all protomers are in the compact configuration, alike the compact protomers of panel G. With such an arrangement, an octameric assembly is formed. It has been verified that the octameric form is indeed the predominant form by native mass spectrometry and analytical ultracentrifugation.⁴

Strikingly, the quaternary assemblies presented in panels **E, F**, and **G** derive from the same protein at different conditions (*apo* vs. *liganded*, see also Discussion for details)⁵. Such structures indicate that the quaternary assembly that dictates DNA bending depends on the orientation of the dimerization helix with respect to the CC. No large domain motions of the D1-D2 of the protomers are observed in such quaternary assemblies (Table. S4). The differences between the protomers of these assemblies involve minor rearrangements of secondary structure elements and partial loss of secondary structure (in SuppFile. 3 denoted with a black rectangle). Moreover, such high-resolution data highlight the remarkable flexibility of the transcription factors, as the same protomer can obtain multiple conformations: two distinct compact configurations with the DNA binding domain either above or below the CC and even an extended configuration. There is even additional variability within the apparently similar compact and extended configurations (SuppFile. 7). Moreover, the prominent octameric population found for 6G1B⁴ (panel **G**) indicates that all of the protomers could also obtain the compact form of 3HHG (panel **H**). Only the compact form of 3HHG satisfies the geometrical criteria to obtain an octameric assembly.

This remarkable flexibility is also substantiated by inspecting the contact map of the CynR models based on the different oligomeric assemblies (SuppFile. 7). Such interaction map indicates that the dimerization helix could obtain multiple arrangements with respect to the CC of CynR in the different protomers, yielding multiple quaternary arrangements (**A-H**). This agrees with our HDX-MS results demonstrating that the dimerization helix and the DNA binding domain are extremely flexible (Fig. 4D).

The 3HHG octameric assembly is unique, as all other known LysR transcription factors are tetramers. Our phylogenetic tree indicated that this structure separates from all the rest in the first clade (SuppFile. 6). Moreover, the phylogenetic tree based uniquely on the CCs provides the same results. This means that the “information” for obtaining an octamer (i.e. all protomers in the compact orientation) resides within the CC. By inspecting the structure-based sequence alignment, we noticed that 3HHG has two differences compared to all other CCPs (SuppFile. 3): the additional helix CH5 connecting Hinge- β -sheet 2 (β H2) and CH6 and it lacks helix $\alpha 4$ that we represent with pink color and stabilizes tetrameric assemblies (see also Fig. 4B). The helix CH6 is placed just below the connecting-loop and fixes it in the compact configuration with the DNA binding domain below the CC. Helix $\alpha 4$ that stabilizes tetrameric assemblies is not required for formation of the octamer, and probably its absence facilitates octamerization as it destabilizes tetramerization.

SUPPORTING INFORMATION

SUPPORTING INFORMATION

Figure S8: Hypothetical Energetic Funnels for the better visualization of the effects of Modularity on the Evolvability, related to Figure 6. (A) According to the view of generalists⁶, an ancestral structure composed of two interacting type-II PBP domains can acquire many distinct structural states depending on the arrangement of the two domains relative to each other. The energetic funnel of such an assembly will have many isoenergetic wells and manifest increased ruggedness, thus a high degree of ligand promiscuity (see also text for details). (B) Evolution of a generic 'primordial' two domain protein becoming a specialized *cherry-core*: When the two domains are linked with a β -sheet, a fix-length polypeptide chain emerges and a specific well (indicated with red arrow) is selected. However, it remains elusive if the evolution process is facilitated by utilizing either: (i) the positive selection (direct deepening of the well) or (ii) negative selection (destabilization of non-beneficial or unfunctional structures; indirectly deepen the well *relative* to other wells, becoming shallow). In the case of negative selection, the consecutive rounds of neutral mutations followed by plastic adaptations might indeed destabilized the wells of the non-functional structural states. The process results in the *relative* deepening and smoothening of the native state well (i.e. closed state, compared to the other wells) conferring ligand specificity (see discussion for details). The aforementioned structural arrangement represents the conserved CC structure. Solely the CC without embellishments is not encountered in extant proteins. Probably because the function of such a structure could be solely binding. As the CC *still* has the "remnants" of the highly-promiscuous ancestral two type-II PBP domains, we anticipate that this structure would exhibit a relatively high ligand promiscuity compared to the evolved CCPs. (C-E) Addition of the modular parts paves the way for evolution to sample a specific well on the funnel. If the presence of the module is beneficial for evolving the new function, nature may then introduce additional mutation to integrate the module into the core. Adopting the concept of protein evolvability, rounds of neutral mutations might be introduced synergistically on the module and the core and thus specify the well of the beneficial promiscuous structure. (C) A modular part represented by an asymmetric C-tail directs the evolutionary pressure to deepen one other specific well. Again, rounds of Neutral mutations coupled with plastic adaptations can greatly deepen this well, beyond the initial CC well (closed state), to create the open state well. This well is deep/smooth having increased ligand specificity. A distinct chemical environment represented by the presence of the ligand generates the *holo* energetic funnel. In the later one, the well corresponding to the closed state is the deeper. (D) A module consisting of a symmetric C-tail, directs evolution to deepen the closed state well and minimize the ruggedness of the funnel. This yields a very robust protein, providing the required stereochemistry necessary for specific reactions (see Fig. S4A). (E) Addition of an N-terminal domain having two modules (dimerization helix and DNA binding motif) widens the energetic funnel. In such a case a single closed well represents the CC, but additional ones derive from the flexible N-terminal domain. The CC is linked by a loop to the flexible domain, providing multiple wells in the energetic funnel (C1-C7). Oligomerization of such monomers will deepen specific wells in the oligomer funnel, similarly to the substrate in the *holo* funnel.

SUPPORTING INFORMATION

Table. S1. Large-scale Domain Motions (Tier-0 states) observed by comparing the *apo* and *holo* crystal structures. The domain movement comparison has been performed by the DynDom server using the indicated available crystal structures on the Protein Data Bank.

	<i>Apo</i> Conformation (PDB ID code)	<i>Holo</i> Conformation (PDB ID code)	Moving Domain of Core Scaffold	Rotation Angle (deg)	Translation (Å)
Class A	2F7B	2F7C	D2	4	-0,4
	3KOT	4WKM	-	0	0
	4XWS	4X6G	-	0	0
	2HXR	3HFU	-	0	0
Class B	4KR5	4KQP	D2	52,1	0,6
	4P0I	4POW	D2	47,7	1,8
	3ZSF	2YLN	D2	60,1	-1,1
Class C	3L6G	3L6H	D2	35	-0,2
	1SW5	1SW2	D2	58,9	-0,3
	2REJ	2RIN	-	0	0
Class D	3S4U	3QUJ	D2	74,3	-0,5
	2X7P	2X7Q	D2	13,4	-0,3
Class E	2I49	2I4B	-	0	0
	2DE2	2DE3	-	0	0
Class F	1SI1	1SI0	D1	26,5	-0,2
	1O7T	1D9Y	D1	19,7	0,1
	1Y9U	2OWT	D1	25,4	-0,6
	2HEU	2I58	D1	25,5	0,3
Class G	1OMP	1ANF	D1	36	-0,6
	3OOA	3OO6	D1	37,9	0,3
	5DVF	5DVI	D1	18,9	-0,2
	1Y3Q	1Y3N	D1	39,1	0,3
	1J1N	1KWH	D1	29,7	0,4

SUPPORTING INFORMATION

Table S2. Stabilization Interactions of the Tier-0 *apo* and *holo* states. Only those interactions between the C-tail and D1/2 that specifically stabilize either the *apo* or *holo* state are shown. For each protein class the corresponding PDB ID code of the *apo* or *holo* structures is indicated. For class E, the only structures differing in the *apo* and *holo* states has been analyzed and shown. H (hydrophobic interaction, cut-off distance 5 Å), MSHB (main-chain side-chain hydrogen bond), I (ionic interaction, cut-off distance 6 Å), A (aromatic interaction, distance 4-7 Å), SSHB (side-chain side-chain hydrogen bond), MMHB (main-chain main-chain hydrogen bond), CP (cation-pi-interaction, cut-off distance 6 Å) and AS (aromatic-sulphur-interaction, cut-off distance 5 Å).

	<i>apo</i> conformation			<i>holo</i> conformation		
	Residue	Interaction	Location	Residues	Interaction	Location
Class A (2HXR - 3HFU)	-	-	-	-	-	-
Class B (4KR5 - 4KQP)	Leu480-Pro419	H	H1			
	Lys426-Asp477	MSHB	H1			
	Lys426-Lys589	MSHB	H1			
	Glu481-Lys426	I	H1			
	Asp483-Lys422	I	H1			
Class C (3L6G - 3L6H)	Tyr423-Tyr479	A	H1			
	Trp488Leu545	H	H3	Pro486-Met544	H	H3
	His487-Glu541	MSHB	H3			
	Glu541-His487	MSHB	H3			
	Met70-Leu254	H	HD1	Val73-Leu254	H	HD1
Class D (3S4U - 3QUJ)	Met70-Leu257	H	HD1	Leu132-Ile257	H	HD1
	Met70-Ala261	H	HD1	Leu132-Ala261	H	HD1
	Leu132-Phe263	H	HD1	Arg258-Asp74	SSHB	HD1
	Tyr136-Leu291	H	HD2	Phe131-Lys264	CP	HD1
	Tyr137-Pro256	H	HD1			
	Ala140-Phe263	H	HD1			
	Tyr137-Pro256	MSHB	HD1			
	Tyr137-Pro256	SSHB	HD2			
	Tyr137-Glu259	SSHB	HD1			
	Arg258-Met70	SSHB	HD1			
Class E (2X7P - 2X7Q)	Gln259-Tyr137	SSHB	HD1			
	Lys141-Asp295	I	HD2			
	Phe122-Val278	H	HE1-HE2	Tyr118-Val278	H	HE1-HE2
	Val123-Leu273	H	HE1	Tyr118-Ala276	H	HE1
	Val123-Ala276	H	HE1	Phe122-Leu275	H	HE1
Class F (1SI1 - 1SI0)	Thr167-Tyr237	SSHB	H1-H3	Tyr118-Ala276	MSHB	HE1
	Lys170-Asp236	SSHB	H1-H3	Lys170-Asp236	MSHB	H1-H3
	Lys169-Glu244	I	H4	Lys169-Glu240	I	H3
				Lys170-Aps241	I	H3
				Tyr237-Lys170	CP	H1-H3
Class G (1OMP - 1ANF)	Ile60-Tyr301	H	H5	Lys202-Leu256	MSHB	H1
	Met64-Tyr301	H	H5	Phe220-Trp284	A	H2
	Leu294-Gln87	MMHB	H5			
	Ser91-Glu295	MSHB	H5			
	Ser91-Glu295	SSHB	H5			
Class H (1OMP - 1ANF)	Phe220-Tyr269	A	H1-H2			
	Tyr301-Met64	AS	H5			
	Tyr90-Met321	H	H4	Ala63-Met330	H	H4-H5
	Phe92-Met321	H	H4	Tyr99-Pro334	H	H5
	Phe92-Ala324	H	H4	Trp232-Pro298	H	H1-H2
	Tyr99-Ile333	H	H4-H5	Gly68-Asn332	MMHB	H4-H5
	Ala168-Pro331	H	H4-H5	His64-Met330	MSHB	H4-H5
	Ala168-Met336	H	H5	Asn332-Ala96	MSHB	H4-H5
	Ala168-Glu328	MSHB	H4-H5	Asp65-Met330	SSHB	H4-H5
	Asp184-Gln365	MSHB	H5	Trp340-Asp65	SSHB	H5
Class I (1OMP - 1ANF)	Trp158-Glu328	SSHB	H4-H5			
	Tyr90-Met321	AS	H4			

SUPPORTING INFORMATION

Table S3. Apparent Melting Temperature Deduced from Intrinsic Trp Fluorescence of CynR. Intrinsic Trp fluorescence as a function of temperature (10°C-70°C) yielded a main T_m (apparent) and a secondary one characteristic of the presence of azide.

	T_m (°C)		
	Primary	Secondary	Azide
- DNA	54.6	N.A.	-
	55.3	26.9	+
+ DNA	48.1	N.A.	-
	48.3	26.5	+

Table S4. Tier-0 states of the indicated protomers within the tetrameric assemblies of the class A transcription factors. Analysis as described in Table. S1

	Conformation 1 (PDB ID code - Chain)	Conformation 2 (PDB ID code - Chain)	Moving Domain of Core Scaffold	Rotation Angle (deg)	Translation (Å)
Class A	6G1B-A	6G1B-B	-	0	0
	6G1D-A	6G1D-B	-	0	0
	6G1D-A	6G1D-C	-	0	0
	6G1D-A	6G1D-D	-	0	0
	6G4R-A	6G4R-B	-	0	0
	6G4R-A	6G4R-C	-	0	0
	6G4R-A	6G4R-D	-	0	0
	6G1B-A	6G1D-A	-	0	0
	6G1D-A	6G4R-A	-	0	0

Table S5: Apparent FRET Efficiencies and Ca-Ca Distances derived from crystal structures. The apparent FRET efficiencies are the center and the errors are the limits of the 95% confidence interval of the Gaussian fit in Fig. 3 and 5. The PDB ID codes of proteins are reported in the Methods section and Table. S1.

	Apparent FRET efficiency		Ca- Ca distance (Å)	
	Apo condition	Saturated condition	Open conformation	Closed conformation
CynR (C137 R233C)	0.703 ± 0.003	0.707 ± 0.003	41.8	
CmpA (T89C N166C)	0.439 ± 0.009	0.425 ± 0.010	48.2	
MalE (T36C S352C)	0.586 ± 0.002	0.780 ± 0.002	51.3	40.8
OpuAC (V360C V423C)	0.550 ± 0.003	0.696 ± 0.003	45.7	36.3
SBD2 (T369C S451C)	0.586 ± 0.002	0.736 ± 0.002	49.0	40.1

SUPPORTING INFORMATION

Legends of Tables Provided as Separate Files

SFL1. Supplementary File S1 – Extended Sequence Based Phylogenetic tree

Sequence-based extended phylogenetic tree based on percentage of sequence identity. Summary of this tree is presented in Fig. S1A. Every protein is represented by its PDB ID (known structures) or UniProt ID followed by its localisations identifier (SP, for presence of signal peptide; Lipidation, Glycosylation for lipidation and glycosylation signals respectively). If protein origin is not bacterial, origin (Archaea, Eukarya, Virus) is indicated.

SFL2. Supplementary File S2 – Full Phylogenetic Trees

Structure- and sequence-based phylogenetic trees of the 53 identified *cherry-core* proteins with known structures. Summary of this tree is presented in Fig. 1C and Fig. S1B-D. Trees are constructed after the entire polypeptide chain (**A**, **B**), the CC only (**C**) and the C-tail only (**D**). Every protein is represented by its PDB ID code.

SFL3. Supplementary File S3 – Secondary Structure Alignment Table

Sheet 1: Revised alignments of *cherry-core* proteins. The identified proteins were sorted in different classes (**A-G**) depending on their N- & C-terminal regions (see also Fig. 1C, D). The alignment of the secondary structure elements is presented (**H**, for α -helix; **S**, for β -sheets) followed by the corresponding residues (the number of residues forming the secondary structural element is indicated in parenthesis). Every protein is represented by its PDB ID code followed by its chain identifier. We represent equivalent secondary structure elements (occupying the same place in the 3D space) with the same uniform color throughout a column. We identified distinct protein classes **A-G** according to the C-terminal regions of their CC, which is formed by different secondary structure elements. Secondary structure elements that are joined (i.e. two α -helices becoming a long one) are marked by red letters. We indicated subclasses with asterisks. Subclasses represent proteins having minor differences in the secondary structure elements of their C-tails, with respect to all other proteins of the same class, e.g., A* has an additional terminal helix at the C-tail similar to H1, but positioned in a different location in the 3D space compared to A. Similarly, for B* which has an additional helix (CH5) and an extra small β -sheet compared to B. The class B** has two extremely short helices CH6 and H1, while sub-class B*** and G* have one additional helix at the C-tail with respect to other B and G class members, respectively.

Sheet 2: Classification of the proteins included in Sheet 1 according to the ECOD database.

SFL4. Supplementary File S4 – HDX MS Data for CynR

Excel file composed of 4 sheets containing 3 tables (F4A, F4B, F4C) and coverage maps. The sheets include detailed description of the conditions and analysis of the HDX-MS experiments.

SFL5. Supplementary File S5 – HDX MS Data for MalE

Excel file composed of 7 sheets containing 5 tables (F5A, F5B, F5C, F5D, F5E), difference plots and coverage maps. The sheets include detailed description of the conditions and analysis of the HDX-MS experiments.

SFL6. Supplementary File S7– Structure based phylogenetic trees of Class A Protein

Results as obtained from the Dali server (see Methods). The structure based phylogenetic tree of class A proteins with known oligomeric structures (see Fig. S7). Top part contains full length class A transcription factors (HTH motif/dimerization helix and CC) while bottom part, only the CC.

SFL7. Supplementary File S7 –Interaction of CynR connecting loop/dimerization helix with the CC. The *cynR* sequence was modelled after the known oligomeric high-resolution structures (see Fig. S7 and methods section). The contact interfaces between the dimerization helix and the CC derived from the structures (left part of the excel table) or from the corresponding models (right part of the excel table) were analyzed in the Protein Interaction calculator server, using standard server settings (alike

SUPPORTING INFORMATION

Table S2, see also Methods). The residues forming such interface belong to the connecting-loop or the dimerization helix (indicated as loop and helix respectively) and elements of the CC (indicated according to the SuppFile. 3). The nature of the interactions, hydrophobic, aromatic, Cation- π , hydrogen-bonds, Ionic are represented with blue, cyan, green, orange and red color respectively.

REFERENCES

1. UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515 (2019).
2. Lai, R.Y. et al. Thiamin pyrimidine biosynthesis in *Candida albicans* : a remarkable reaction between histidine and pyridoxal phosphate. *J Am Chem Soc* **134**, 9157-9 (2012).
3. Koentjoro, M.P., Adachi, N., Senda, M., Ogawa, N. & Senda, T. Crystal structure of the DNA-binding domain of the LysR-type transcriptional regulator CbnR in complex with a DNA fragment of the recognition-binding site in the promoter region. *FEBS J* **285**, 977-989 (2018).
4. Sainsbury, S. et al. The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic Acids Res* **37**, 4545-58 (2009).
5. Pedre, B. et al. Structural snapshots of OxyR reveal the peroxidatic mechanism of H₂O₂ sensing. *Proc Natl Acad Sci U S A* **115**, E11623-E11632 (2018).
6. Tawfik, D.S. & Gruic-Sovulj, I. How evolution shapes enzyme selectivity - lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *FEBS J* **287**, 1284-1305 (2020).