Structural dynamics in the evolution of a bilobed protein scaffold

Giorgos Gouridis,a,b,c,1,2 Yusran A. Muthahari,b,c,d,1 Marijn de Boerb, Douglas A. Griffithb, Alexandra Tsirigotakib, Konstantinos Tassib, Niels Zijlstra, Ruxue Xuc,e, Nikolaos Eleftheriadisb,c, Yovin Sugijo, Martin Zacharias, Alexander Domlingb,c, Spyridoula Karamanouc, Charalampos Pozidis, Anastassios Economouc, and Thorben Cordesa,2,3

aPhysical and Synthetic Biology, Faculty of Biology, Ludwig Maximilians-Universität München, Planegg-Martinsried 82152, Germany; bMolecular Microscopy Research Group, Zernike Institute for Advanced Materials, University of Groningen, Groningen 9747 AG, The Netherlands; cLaboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven 3000, Belgium; dStructural Biology Division, Institute of Molecular Biology and Biotechnology, Heraklion-Crete 70013, Greece; eDrug Design Group, Department of Pharmacy, University of Groningen, Groningen 9713 AV, The Netherlands; and fTheoretical Biophysics, Physics Department, Technical University of Munich, Garching 85748, Germany

Published November 29, 2021.

Edited by Martin Gruebele, University of Illinois at Urbana–Champaign, Urbana, IL, and approved October 28, 2021 (received for review January 4, 2021)

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 remain poorly understood. Here, we investigated how proteins with a bilobed structure, composed of two related domains from the periplasmic-binding protein-like II domain family, have undergone divergent evolution, leading to adaptation of their structural dynamics. We performed a structural analysis on ~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, enable distinct structural dynamics, allowing the diversification of these proteins into transcription factors, enzymes, and extracytoplasmic transport-related proteins. Structural embellishments of the core created 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 evolution and are applicable to a large number of domain architectures.

Proteins drive and maintain all fundamental cellular processes (1) by interactions with small molecules and/or other biopolymers. Important mechanistic information on proteins is accessible via structural analysis of their functional cycle (2). While classical approaches rely on the interpretation of static structure snapshots, the visualization of structural dynamics (i.e., to follow the interconversion of distinct structural states at high spatial and temporal resolution) (3–7) has been recognized as an essential complement. The folding funnel model (8), rooted in the free-energy landscape theory (9–11), has by now become a widely accepted way to describe the energy landscape of a protein defines the lifetime of its structural states, ranging from nanoseconds (local flexibility) to seconds (large-scale motions). Transitions between the states are referred to as structural changes and are induced by interactions with ligands, posttranslational modifications (e.g., phosphorylation), or chemical events such as nucleotide hydrolysis. The coupling of the latter to structural changes enables proteins to perform a diverse range of functions.

Tier-0 dynamics were observed and characterized in various settings (e.g., in motor proteins, in which they are used in propelling movement along filaments) (16), in the transport of molecules or biopolymers across biological membranes (17–21), 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 enzymes also show varying degrees of structural dynamics (26), 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 catalysis (27) 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

Significance

Proteins conduct numerous complex biological functions by use of tailored structural dynamics. The molecular details of how these emerged from ancestral peptides remains mysterious. How does nature utilize the same repertoire of folds to diversify function? To shed light on this, we analyzed bilobed proteins with a common structural core, which is spread throughout the tree of life and is involved in diverse biological functions such as transcription, enzymatic catalysis, membrane transport, and signaling. We show here that the structural dynamics of the structural core differentiate predominantly via terminal additions during a long-period evolution. This diversifies substrate specificity and, ultimately, biological function.

Author contributions: G.G., Y.A.M., M.d.B., and T.C. designed research; G.G., Y.A.M., M.d.B., D.A.G., K.T., R.X., Y.S., M.Z., A.D., C.P., and A.E. contributed new reagents/analytic tools; G.G., Y.A.M., M.d.B., D.A.G., A.T., K.T., N.Z., R.X., Y.S., and S.K. performed research; G.G., Y.A.M., M.d.B., D.A.G., K.T., R.X., Y.S., M.Z., A.D., C.P., and A.E. contributed new reagents/analytic tools; G.G., Y.A.M., M.d.B., D.A.G., N.Z., A.T., N.E., and T.C. analyzed data; G.G., Y.A.M., D.A.G., and T.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1G.G. and Y.A.M. contributed equally to this work.
2To whom correspondence may be addressed. Email: g.gouridis@imbb.fORTH.gr or cordes@bio.lmu.de.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2026165118/-/DCSupplemental.

Published November 29, 2021.
shown. Revised alignments of bilobed proteins are shown in Dataset S3. The secondary structure elements forming the consensus Schematics are based on the crystal structures of the selected proteins. (e.g., HD1, first unique helix of class D proteins). An asterisk (*) marks a structural subclass (shown in detail in Dataset S3). Analyses of selected proteins suggests that functional domains are reused when creating new proteins and functions (42, 43). Analyses of selected domain groups (44, 45), and more-recent large-scale investigations (46–48) have shown that domains in such groups generally share a common structural core (40 to 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 (49). Such fusions occur frequently at the N or C termini (40), which has led to the belief that structural elements act as Lego bricks (47). Those are recombined in various ways for new functions to emerge during evolution (42, 49).

Much less is known on the role played by structural dynamics in the evolution of protein function (50–53). 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 (54, 55). The “avant-garde” evolvability theory proposes the existence of a highly promiscuous primordial protein structure. It is hypothesized that for the emergence of the native state, the ability to evolve (evolvability) was traded for ligand-functional specificity (56, 57). These observations were experimentally verified (58) and agree with the proposal that changes in structural dynamics serve as a mechanism for the evolution of specialist enzymes from promiscuous generalists (59), though discordant examples are also observed in evolution (60). Furthermore, 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.
(54–56, 61). Remarkably, this short-period evolution can be faithfully reproduced in vitro using directed evolutionary approaches based on consecutive rounds of single point mutations (54, 61). 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 (62). However, it still remains unclear how, during longer periods of evolution, a primordial core structure evolves to modulate or diverge Tier-0 and/or quaternary dynamics, “generating” 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, 55), having (1) a conserved core structure, (2) large sequence diversity and related functional divergence, and (3) occurrence in all kingdoms of life. The selected structural core is composed of two Rossmann-like domains, which are believed to be among the most ancient architectures (63) and that are connected by a single β-sheet (Fig. 1B). We analyzed ~600 proteins with this structural core that contain two related domains from the periplasmic-binding protein (PBP)–like II domain homologs (ECOD: X-, H- and T-groups PBP-like II). We show here that different members of these homologous proteins diverged with respect to domains or secondary structure elements, predominantly at their termini (Fig. 1 C and D).

Using a combination of structural analysis and biophysical interrogation methods, we demonstrate that such structural embellishments confer multi-Tier structural dynamics that diversify the function of the core structure to yield transcription factors, enzymes, or extracytoplasmic transport-related proteins. To understand both the mechanistic distinction of these proteins and the evolutionary trajectories, we used single-molecule Förster resonance energy transfer (smFRET) (64) and Hydrogen–Deuterium Exchange Mass Spectrometry (HDX-MS) (65). smFRET allows to monitor Tier-0 dynamics with a temporal resolution down to microseconds and subnanometer spatial resolution at the single-molecule level, even for highly heterogeneous structural ensembles (12, 66–68). 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) (69). Using this combination of techniques, we show how the C-terminal extensions modulate Tier-0 dynamics in 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 (LTTRs). The remarkable modularity of these proteins permits us to confirm and expand the evolvability theory for the primordial core structure during a 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 three layers (α/β/α), articulated around a central β-sheet hinge (Fig. 1B). This focus meant that not all multidomain architectures harboring the PBP-like II domains were included. To investigate the structural dynamics of 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 and D), members of which are found throughout all kingdoms of life, in which some are even present in viruses (SI Appendix, Fig. S1A and Datasets S1 and S2). They have a consensus structure (Fig. 1B, Fig. 1D, Top, and Dataset S3), which we have dubbed the cherry-core (hereafter CC, and proteins harboring this core, cherry-core proteins, CCPs) because of the bilobed structures’ resemblance to a cherry (SI Appendix, Fig. S2A), that contains two PBP-like II domains (70). According to ECOD database, the two domains of CCPs belong to the same X-, H-, and T-groups (PBP-like II; Dataset S3), supporting their common ancestry (38). In the consensus structure (Fig. 1C and 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 (SI Appendix, Fig. S2B).

Most of the selected proteins have distinct segments linked N-terminally to the CC (Fig. 1C and D). 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 (71–73). Most proteins of classes B through D, and F and G contain N-terminal localization signals for export via the general secretion (Sec) or the twin-arginine translocation (Tat) pathway (74–77).

The presence of these signal peptides was obtained from UniProtKB (78) and verified manually by inspecting all 600 protein sequences with PRED-TAT, prediction of twin-arginine and secretory signal peptides (79). Class E proteins are predominantly cytosolic and lack an N-terminal signal peptide. In addition, classes B through 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 C, and importantly, it has a similar placement in the three-dimensional architecture of the proteins. Other C-tails are unique to a specific class, such as the Helical-tails G (HG1 and 2), present only in class G proteins.

Function, Ligand Specificity, and Structural States of the Selected Bilobed Proteins. The ligand specificity and function of the CCPs, as documented in UniProtKB (78), were found to correlate with the assigned structural class (SI Appendix, Fig. S1A). Class A proteins are bacterial transcription factors of the LTTR family. Class E proteins are predominantly eukaryotic single-turnover enzymes. The majority of the remainder (class B through D, and F and G) are found in prokaryotes and are associated with the translocator domains of ABC transporters or with the membrane-embedded domains of chemoreceptors, in which they mediate unidirectional solute transport and signal transduction, respectively.

To investigate what role the distinct C-tails of these proteins might have played in structural dynamics and the evolution of new functions, we examined CCPs for which high-resolution structures of unliganded (apo) and liganded (holo) states were available (Fig. 2 and SI Appendix, Table S1). Interestingly, class A and E proteins, in most cases, exhibit nearly identical apo and holo structures. They also display the widest variety of substrates with little chemical structure similarity (SI Appendix, 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 and SI Appendix, Table S1). For the solute-binding proteins, this mode of substrate binding has been termed the “Venus-Fly Trap mechanism” (80). These proteins recognize ligands with a specific pharmacophore (81): amino acids, ethanolamines, phosphonates, iron–phosphate complexes, and carbohydrates are recognized by classes B, C, D, F, and G, respectively (SI Appendix, Fig. S3). Another striking difference is that proteins in classes B through G are monomeric, whereas those in class A are oligomeric (see Fig. 4C).
smFRET to Monitor Large Domain Motions (Tier-0) in Bilobed Proteins.

To investigate structural dynamics of the CCPs, we first used smFRET (12, 66–68) to probe Tier-0 dynamics at near-physiological conditions in aqueous buffer at room temperature. One representative domain from each class was investigated: the effector binding domain (EBD) of CynR, representing the CC of full-length CynR (72) (class A), SBD2 (82) (class B), OpuAC (83) (class C), PhnD (84) (class D), CmpA (85) (class E), FbpA (86) (class F), and MalE (87) (class G). For these experiments, D1 and D2 were stochastically labeled with donor and acceptor fluorophores via cysteines that were substituted for nonconserved and surface-exposed residues, one in each domain (Fig. 3 A). Fluorophore labeling was selected based on the crystal structures to show large changes in separation between the apo and holo states. smFRET was performed by confocal microscopy with alternating laser excitation (ALEX) (66).

As predicted by our structural analysis (Fig. 2 and SI Appendix, Table S1) and in line with our previous observations (89), the FRET efficiency histograms and fitted distributions (Fig. 3 B–F) shifted toward higher FRET efficiency (E) values upon addition of saturating concentrations of ligand, for SBD2, OpuAC, PhnD, FbpA, and MalE. This indicates that in the apo state, the donor and acceptor dyes are further apart (Fig. 3 B–F, Upper, low FRET) compared to the holo state (Fig. 3 B–F, Lower, 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 [i.e., CynR (90) and CmpA (85), respectively] (Fig. 3 G and H), 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. 3 I and J). 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 (SI Appendix, Fig. SL4), for which the rigidity of their active site is a prerequisite for catalysis (SI Appendix, Fig. S4A) (91). As previously suggested by us and others, Tier-0 dynamics in periplasmic-binding proteins (classes B, C, D, F, and G) are utilized in the regulation of transport in ABC importers (Discussion). A remaining question is, however, how ligand binding in class A proteins triggers transcriptional processes without major structural changes related to ligand binding.

HDX-MS to Examine 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 links 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 EBD (Fig. 4B and SI Appendix, Fig. S4 B–E). Indeed, by using size exclusion chromatography with multi angle light scattering (SEC-MALS) we observed that (full-length) CynR is tetrameric, whereas its CC is dimeric (Fig. 4C).

To determine whether ligand binding to CynR resulted in structural changes that were not detectable along our selected reaction coordinate or that were too small or fast for smFRET (Fig. 3G), 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 (94, 95). 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 MS. The mass difference between hydrogen (1H) and deuterium (2H) results in a mass shift between nondeuterated and deuterated peptides that is a measure of the number of exchanged hydrogens (65, 96, 97).

For such investigations, the CC of CynR or full-length CynR were isotopically labeled (pD, 7.4; 25 °C) for different time periods (10 to 105 sec) either in free or DNA-bound states, and the peptic fragments were identified by MS (Fig. 4D and Dataset S4). For each peptide, the fraction of deuterium uptake relative to the maximum determined deuterium incorporation was calculated (Dataset S4). The data reveal a rigid character...
An Asymmetric C-tail Drives Large-Scale Tertiary (Tier-0) Structural Changes. From our results on the selected CCPs, only those with an asymmetric C-tail display Tier-0 dynamics (Figs. 2 and 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 states are class specific. Contact mapping of the interactions between the CC and C-tail using the protein interaction calculator web-server (99) showed that the number and characteristics of the interactions depends on the structural class (Fig. 2 and SI Appendix, Table S2). Strikingly, these interactions stabilize predominantly the open state of the CC, the only exception being for classes A (C-tail-less) 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 and D2 and thus cannot provide the required structural asymmetry needed to create a stable open state. The interactions of the C-tail to stabilize the open state involve the consensus CC-helices of D1 and D2 (Fig. 2 and SI Appendix, 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 and SI Appendix, Fig. S2).

In alternative phylogenetic trees, based on the protein sequence using either D1/D2-domains or the C-tail, the clustering remains similar (SI Appendix, Fig. S1 B–D and Datasets S1 and S2), indicating that D1/D2 (CC domains) and the C-tail of a specific class coevolved 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 State. To confirm the role of the C-tail interactions with D1 or D2 for stabilization of the open state, we manipulated structural states by FRET efficiency via stochastic labeling of D1 and D2 with donor and acceptor fluorophores. (B–H) Solution-based apparent FRET efficiency histograms in the absence (Top) or presence (Bottom) of saturating ligand concentrations for the indicated proteins. Gray bars are experimental data, and solid line is the fit. Centre position of Gaussian fits are given in SI Appendix, Table S5. (I and J) Binding isotherms of the caloric heat of reaction per injection, and the line is the fit.

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 in 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 (Dataset S4). No statistically significant change in the deuterium uptake was observed (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 and Dataset S4). This includes Tier-0 dynamics (detected by smFRET and HDX-MS) and Tier-1 or quaternary dynamics (both detected by HDX-MS).
the relevant ones (Fig. 2 and SI Appendix, Table S2) in SBD2, SBD1, and MalE (Fig. 5 and SI Appendix, Figs. S5 and S6). The impact was tested via assessing structural states and ligand-binding affinities and 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. S4 and SI Appendix, Table S2). The mutation resulted in the appearance of a subpopulation of molecules (~20%) that were in the closed state in the absence of glutamine (Fig. 5C versus B). To rule out the possibility that this might have been due to an artifact introduced by the choice of fluorophores, a second pair was tested, which showed a comparable result (SI Appendix, Fig. S5 A and B). We also examined whether residual endogenous ligand might account for this subpopulation by performing smFRET measurements on diluted samples, but these experiments displayed subpopulations of a similar size (SI Appendix, Fig. SSC). A closed-unliganded conformation was also observed for SBD2 previously but with a much lower abundance (~1%). Detection of this small subpopulation required the use of confocal scanning microscopy (89), as populations <5% cannot be detected reliably with ALEX spectroscopy. 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 (SI Appendix, Fig. SSB). Destabilizing the open state is in line with the ~10-fold increase in glutaminel-binding affinity of SBD2 (L480A) as compared to SBD2 (Kd of 209 ± 64 nM and 1,990 ± 130 nM, respectively; Fig. 5D).

Determination of the dissociation constant (Kd) by smFRET measurements reports on the stability of the open state via $K_d \propto \exp(\beta \Delta)$, where $\beta = 1/k_B T$, $k_B$ is the Boltzmann constant, $T$ is the absolute temperature, and $\Delta = G_C - G_O$ is the conformational (free) energy difference, where $G_C$ and $G_O$ are the (free) energies of the closed and open structural states without ligand bound, respectively (100). Thus, destabilizing the open state will decrease $K_d$ and vice versa. From the percentages of open and closed in the absence of ligand, we obtain $\Delta = 1.4$ and $\Delta = 4.4$ k_BT for SBD2 (89) and SBD2 (L480A) (Fig. 5B–D and SI Appendix, Fig. S5 A–C), respectively. Based on these results, the $K_d$ value would be expected to decrease by ~16 folds due to the L480A mutation, in close agreement with the estimated ~10-fold difference. Similar trials on SBD1 did not show this trend (SI Appendix, Fig. S5 D–F), as the mutation had no 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 μM). Ultraviolet (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 were 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) and the CC of CynR (Lower). Thresholds used to define residues were used (SI Appendix, Material and Method), as described previously (98). Statistically significant differences would appear as black spheres (indicating that ΔD-uptake > 2×SD for a specific peptide), lying outside the 99% confidence threshold (1-P ≥ 0.99; indicated on y-axis) and outside the ±4 × pooled average SD cut-off (indicated on x-axis; value given on the right).

6 of 12 | PNAS
https://doi.org/10.1073/pnas.2026155118
Gouridis et al.
Structural dynamics in the evolution of a bilobed protein scaffold

Fig. 4. CynR tertiary and quaternary assemblies probed by MALS and HDX-MS. (A) Crystal structure of the CynR CC (PDB: 2HXR) with colored secondary structure elements discussed in the text that are critical for its quaternary dynamics (Top). A homology model of full-length CynR obtained from the SWISS-MODEL server with CbnR (PDB: 1IZ1) as a template (Bottom). The HTH domain is colored green, and the loop connecting it to the CC splitpea. W64 at the tip of the dimerization helix is indicated with an arrow. (B) Top) Schematic representation of the CC of CynR using the same color-coding as in A. (Middle) 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 the CynR protomers self-assemble to form dimers or, for the full-length protein, tetramers. For clarity, two of the protomers have been omitted. (Bottom) The tetrameric structure is superimposed on the measured mass (black cycles). (C) Absorbance 295 nm of CynR apo vs holo in the absence of ligand and DNA by HDX-MS. For the other in the extended conformation are shown at the top of the plane, whereas those that are extended are at the bottom. (D) UV traces of the chromatograms were superimposed on the measured mass (black cycles). (E) 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 were 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) and the CC of CynR (Lower). Thresholds used to define residues were used (SI Appendix, Material and Method), as described previously (98). Statistically significant differences would appear as black spheres (indicating that ΔD-uptake > 2×SD for a specific peptide), lying outside the 99% confidence threshold (1-P ≥ 0.99; indicated on y-axis) and outside the ±4 × pooled average SD cut-off (indicated on x-axis; value given on the right).
impact on the structural dynamics of SBD1 (I249A). The absence of a subpopulation of molecules in the closed state at apoprotein conditions (alike SBD2) is based on the inability to absence of a subpopulation of molecules in the closed state at impact on the structural dynamics of SBD1 (I249A). The

**Fig. 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) Dotted rectangles (Left) on the SBD2 (in A, PDB: 4KRS) structure 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) depicting interactions between the C-tail helix H1 and D2 in the indicated apo (open) or holo (closed) states. Distances (Å) between L480 and P419 are shown as black dotted lines. (B and C) Solution-based apparent FRET efficiency histograms of SBD2 (B) and SBD2 (480A0) (C) at different conditions as indicated. (D) Fraction of the closed state (high-FRET state) of SBD2 and the indicated derivative as a function of glutamine concentration. (E) As in panel (A) for the MalE (PDB: 1OMP) structure with the indicated secondary structure elements and critical contacts. (F and 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) of MalE and its derivatives as a function of maltose concentration. (I) Maltose release from MalE (M321K) over time determined by solution-based smFRET (reference detailed values in SI Appendix, Fig. S6). Data points (panels B and F) and gray bars (panels 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 Dataset S5). n = 3. (K and L) Deuterium uptake for the indicated MalE C-tail helices. SDs (SD) are shown in the deuterium uptake plots.

For MalE, we constructed the derivatives MalE (M321A) (101) and MalE (M321K), with weakened interactions between the C-tail HG1 and D1, by disrupting the hydrophobic interaction of M321 with Y90 and F92 and the aromatic sulfur interactions with Y90 (Fig. 5E and SI Appendix, Table S2). For MalE (M321A), the stabilizing interactions of M321 are partially abolished. This resulted in an eightfold increase of maltose affinity (Kd from 2,400 ± 400 nM in MalE to 300 ± 50 nM in MalE [M321A] [Fig. 5H and SI Appendix, Fig. S6]). An even stronger effect was observed for MalE (M321K) with an affinity enhanced by 3,000-fold (Kd of 0.81 ± 0.15 nM; Fig. 5G versus F and Fig. 5H). Shilton and coworkers have also proposed the hydrophobic interactions of M321 as being important structural determinants of the open state (101), 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 MalE. Addition of 10 nM maltose allowed MalE (M321K) to occupy the closed state exclusively (SI Appendix, Fig. S6H). A total 20 μM unlabeled MalE (M321K) protein was subsequently added to scavenge maltose, which is stochastically released from the labeled protein. In a time-course experiment, the decrease in the population of closed, maltose-bound MalE (M321K) was then followed as a function of time (102) (Fig. 5I and SI Appendix, Fig. S6H). From these experiments, we established that the lifetime of the closed, maltose-bound conformation of MalE (M321K) was 122 ± 12 s. This value was 2,500-fold higher than the value determined for MalE (0.048 ± 0.010 s; SI Appendix, Fig. S6E). This result is consistent with the observed increase affinity of the derivatives for maltose.
The C-tail Local Dynamics Impact the Structural Transitions of the cherry-core. In the above-described MalE derivatives, the structural basis for the destabilization of the open state could not be addressed mechanistically. Our smFRET experiments report on Tier-0 dynamics that affect the proteins’ tertiary structure (Fig. 5). Considering that we observed differences in the Tier-0 dynamics, as a consequence of C-tail-D1/D2 destabilization, only for SBD2, we performed comparative HDX-MS (Dataset S5) to monitor changes in the secondary structure of MalE in comparison to MalE (M321K). For this, we determined the difference in deuterium uptake (ΔD) for each peptide between MalE and MalE (M321K) in their apo states (Fig. 5). Remarkably, comparative HDX-MS indicates that the differences between MalE and MalE (M321K) are localized almost exclusively at the C-tail and specifically in regions interacting with D1 (Dataset S5 and Fig. 5). The D-uptake of the C-tail helices H2/HG1 that are critical for stabilizing the open state (Fig. 5E and SI Appendix, Table S2) denotes that their rigidity was significantly reduced in the MalE derivative (Fig. 5 K and 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 MalE upon maltose binding (Dataset S5). 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 state, because these C-tail elements and the region containing contact residues were found to be more flexible and solvent exposed.

Taken together, we postulate that ligands attenuate the stability of the open state in the CCPs, adopting multiple allosteric models for signal propagation (103): either via Tier-0 (SBD2 [Fig. 5 C and D]) or Tier1/2 (MalE [Fig. 5 G, H, K, and L]) dynamics.

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 (104). However, similar structures can originate from divergent, convergent or parallel evolution (105). The most common “tricks” nature uses to vary a protein domain are the following: β-strand invasion/withdrawal, insertions/deletions/ substitutions of secondary structure elements, domain flip/s, and circular permutations (106, 107). 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 homologs. 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 conserved structural core and 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 (SI Appendix, Fig. S1). The identified proteins likely emerged from divergent long-term evolution from a common ancestor, which spreads throughout the tree of life. This common ancestor is seemingly represented by the consensus core structure (CC; SI Appendix, Fig. S2) and encountered within the type-II class of PBPs (70, 108). As proposed previously (70), the CC derived possibly from a gene duplication of a PBP-like II domain (ECOD; Dataset S3) connected by a β-sheet (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 extracytoplasmic monomeric protein that associates with the translocator domains of ABC transporters or with chemoreceptors (SI Appendix, Fig. S1 and Fig. S6). The two structural states are apo-open versus holo-closed and originate from Tier-0 dynamics, which are critical determinants of their biological function: the membrane-embedded partners can discriminate between open versus closed states to activate or inactivate a biological process (88, 109–112), 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. 5F) (the C-tail elements with either D2 [classes B, C, and D] or D1 [classes F and G] [Fig. 2]). Seemingly, entropic contributions (protein conformational entropy) bias the structural equilibrium toward the closed state. Given the fact that the holo is the closed state driven by ligand binding (89), we anticipate that the interactions of the ligand with the CC eleft allosterically induce order-to-disorder transitions to alter the structural equilibrium. Depending on the placement of the asymmetric C-tails 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 (Figs. 2 and 6E). We verified that each active site geometry can recognize a specific chemical structure (SI Appendix, Fig. S3), in full agreement with ligand binding triggering closing.

In contrast to the C-tail, the presence of the N-terminal signal peptide does not affect the Tier-0 states of MalE (MalE versus proMalE; SI Appendix, Table S5). However, to render protein trafficking to extracytoplasmic locations possible, the presence of the signal peptide is known to delay MalE folding (113, 114). By that, preproteins are allowed to be secreted by the protein translocase (115).

The evolvability theory of Tawfik and coworkers (29, 57) explains the functional promiscuity encountered within the different structural classes. We selected two closely related proteins in our phylegetic tree in class B (Fig. 1C) to illustrate this (Dataset S2, marked red asterisk). SBD1 (Protein Data Bank [PDB]: 4LA9) mediates the unidirectional transport of two substrates (glutamate and asparagine), whereas SBD2 (PDB: 4KKS) transports one of them (glutamine), though it captures it with higher affinity (88). SBD1 can be transformed to bind glutamine with higher affinity like SBD2 by mutating three amino acids (82). Clearly, functional promiscuity is traded seemingly with ligand specificity (62). 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 (SI Appendix, Fig. S1) would gain tertiary structural variability by acquiring four more helical elements at their C termini (H2, HG1, H4, and HG2; Fig. 1C and D) (Fig. 6). This would potentially allow to switch them to class G to bind and mediate transport of carbohydrates (SI Appendix, Fig. S1A). According to another evolutionary trajectory, class B proteins, by acquiring helices H3, HE1, and 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 dehydratase (116), 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-ligated structural states 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 represents a constraint for catalytic activities (117, 118). 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 (SI Appendix, Fig. S1). Those proteins are present in a unique (apo and holo) closed state (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 extracytoplasmic side (like the classes B, C, D, F, and G; described in the previous paragraph and in such a case also possess an N-terminal signal peptide) but are also tethered covalently to the translocator ATPase motor domains (93). As the switching (open to closed) behavior—to activate the transport cycle—is missing (88, 89, 109), the noncanonical 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 (SI Appendix, Fig. S4A) (91). Evidently, the rigidity required for the chemistry is so dramatic that the flexibility of a noncovalently (to the polypeptide chain) bound histidine to the clmf 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 (no C-tail; Fig. 4D) yields transcription factors of the LTTR family. Apparently, in such a case, Tier-0 dynamics of the CC are not required for function (Fig. 3G). The rigidity of the CC (Fig. 4D) is evidently required for the stability of the quaternary assemblies, as those have large cavities and holes (SI Appendix, Fig. S7) (119–124).

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, and 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 EBD) of a full-length transcription factor of the LTTR family with an additional winged HTH-type DNA-binding domain (Fig. 4A and 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 (72, 125). These quaternary dynamics were suggested to activate transcription by a transition from a bent DNA (transcription OFF) to an un bent state (transcription ON) (72, 125). Our smFRET and HDX results (Figs. 3G and 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 potential quaternary dynamics, we analyzed the available high-resolution structures of the oligomeric assemblies belonging to the LysR transcription factors (119–124) to identify the evolutionary relevant ones by Evolutionary Protein–Protein Interface Classifier (EPIC) (126, 127). Subsequently, we modeled the CynR sequence after these structures in the LTTR family. Apparently, in such a case, Tier-0 dynamics of the CC are not required for function (Fig. 3G). The rigidity of the CC (Fig. 4D) is evidently required for the stability of the quaternary assemblies, as those have large cavities and holes (SI Appendix, Fig. S7) (119–124).

Gouridis et al.
Structural dynamics in the evolution of a bilobed protein scaffold
PNAS | 9 of 12
https://doi.org/10.1073/pnas.2026165118
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 $T_{m(app)}$ at $\sim 27^\circ \text{C}$ in both cases (SI Appendix, Table S3).

To understand the signal propagation originating from ligand binding, we inspected the available structural information (SI Appendix, Fig. S7). In agreement with the experimental findings (Figs. 3G and 4E), our structural analysis indicated that D1 and D2 motions are not required (SI Appendix, Table S4) for the oligomeric assemblies of the same transcription factor (OxyR) to differ dramatically in order to trigger structural changes on DNA (SI Appendix, Fig. S7 E–G). What varies between those assemblies is the order–disorder of secondary structure elements within D2 of the CC ($S$ till $\alpha_5$; Dataset S3). We anticipate that such changes in the flexibility of secondary structure elements are induced by ligand binding, like in the case of MalE (Fig. 5 J–L).

We conclude that signal propagation in CCPs comprising large D1/D2 rearrangements (Fig. 2; classes B, C, D, E, and G) driven by the C-tail/D1 and D2 interactions involves bending/unbending of the “spring-like” hinge (108). On the other side, propagation in class A is initiated by small/localized rearrangements of second- ary structure element within the CC somehow transmitted to the N-terminal HTH1 domain leading to global quaternary structural changes (SI Appendix, Figs. S7 and 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. Fig. 6 summarizes our proposed evolutionary path taken by the CC proteins over long time periods. The terminal modules impact and finetune the multi-Tier structural dynamics as can be described by the folding funnel model (8). The CCPs with asymmetric C-tails displaying Tier-0 dynamics (i.e., open and closed states) have the two characteristic wells in the funnel separated by a large energetic barrier (Fig. 6B and SI Appendix, Fig. S8C). For that reason, such proteins are found predominantly in the lowest–energetic level open state (apo energetic funnel; SI Appendix, Fig. S8C), with infrequent transitions to the closed one (Fig. S8B). Only in the SBD1/2 proteins (class B), $\sim 1\%$ occurrence of a closed state has been experimentally observed (89). By destabilizing the open state of SBD2, we obtained a $\sim 20\%$ occurrence of a closed state (Fig. 5C). As in the holo state energetic funnel, the lowest energy state is the closed one (SI Appendix, Fig. S8C); addition of the ligand shifts the equilibrium toward the closed state. Since the dissociation constant ($K_D$) derives from the difference between the lowest energetic levels of the apo and holo funnels (SI Appendix, Fig. S8), destabilization of the C-tail (Fig. 5 D and H) that directs the open state of the apo funnel at a higher energetic level leads to an increased affinity.

Alternatively, CCPs 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 (SI Appendix, Fig. S8D).

The CCPs with N-terminal domain additions (Fig. 6D) have multiple energetic funnels corresponding to oligomerization states (SI Appendix, 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 (SI Appendix, Fig. S8E). 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 state), thus allowing sampling of multiple structural states (SI Appendix, Fig. S8A). Such a structural variability would lead to increased substrate promiscuity, however, exploiting extremely weak interactions (i.e., low binding affinities; SI Appendix, Fig. S8B).

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

Materials and Methods

Detailed materials and methods are included in SI Appendix, Material and Method. All proteins were expressed in *Escherichia coli* cells (BL21 DE3 or BL21 pLysS DE3) and grew in Luria Bertani or Terrific Broth media. Purification was based on affinity (Ni-NTA (Qiagen)); anion exchange [Q Sepharose (GE Healthcare)], and/or size-exclusion [Hi Load 26/60 Superdex 200 (GE Healthcare)] chromatography. Samples used for analysis (IAT, smFRET, HDX-MS, etc.) were freely available software or servers: Structure similarity search-PDB webserver (128, 129), PDBSelect (130), DalI server (131), multidimensional QR factorization of multiple sequence and structure alignments (132); visual molecular dynamics 1.9.2 software package (133), Protein blast (134), DynDom domain motion server (135, 136), Protein Interaction calculator webserver (99), EPIC server (126, 127), ECOD database (38), ConSurf-DB server (137, 138), SWISS-MODEL server (139), and PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). smFRET experiments were performed with a custom-made confocal ALEX microscope that has been previously described in the literature (22, 89) and HDX-MS by a nanomCOUTITY Ultra Performance Liquid Chromatography System with HDX technology (Waters, United Kingdom). Data and statistical analyses were performed by established procedures described in detail in the literature.

Data Availability. Study data are included in the article, supporting information or SI Appendix and Datasets.

Acknowledgments. This work was financed by an Netherlands Organisation for Scientific Research (NWO Veni grant 722.012.012 to G.G.), an ERC StG (2012-2016) (European Research Council SIG No. 638936, SM-IMPORT to T.C.), Deutsche Forschungsgemeinschaft within GRK2062 (project CO3 to T.C.), SFB863 (project A10 to M.Z. and A13 to T.C.), Research Foundation Flanders (FWO CARBS #G0C6814N to G.G. and A.E.). G.G. also acknowledges a fellowship from the European Molecular Biology Organization (EMBO long-term fellowship ALF 47-2012 to G.G.) and financial support by the Zernike Institute for Advanced Materials and the Rega foundation. Y.A.M. was supported by the Indonesian Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan, Republik Indonesia, LPDP RI PhD scholarship). G.G. and Y.A.M. acknowledge financial support from the IMBB-FORTH (start-up grant to G.G.). N.Z. acknowledges an Alexander von Humboldt postdoctoral fellowship. N.E. acknowledges a fellowship from the Marie Skłodowska Curie Action (MSCA SoE FWO (195872), an ion exchange [Q Sepharose (GE Healthcare)], and/or size-exclusion [Hi Load 26/60 Superdex 200 (GE Healthcare)]). We thank Eltan Lerner for useful comments and critical reading of the manuscript, Monique Wiertsma for help with smFRET experiments, and Florence Husada for support with IAT experiments.

1. G. M. Cooper, R. E. Hausman, The Cell: A Molecular Approach (Sinauer Associates, Sunderland, MA, Washington, D.C., ed. 6, 2013).
2. I. D. Campbell, Timeline: The march of structural biology. Nat. Rev. Mol. Cell Biol. 3, 377–381 (2002).
3. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, J. G. Gunsalus, Dynamics of ligand binding to myoglobin. Biochemistry 14, 5355–5373 (1975).
4. H. Frauenfelder, B. H. McMahon, P. W. Fenimore, Myoglobin: The hydrogen atom of biology and a paradigm of complexity. Proc. Natl. Acad. Sci. U.S.A. 100, 8615–8617 (2003).
5. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. Science 254, 1598–1603 (1991).
6. K. Nienhaus, P. Deng, J. M. Kriegl, G. U. Nienhaus, Structural dynamics of myoglobin: Effect of internal cavities on ligand migration and binding. Biochemistry 42, 9642–9658 (2003).
7. K. Okazaki, N. Koga, S. Takada, J. N. Onuchic, P. G. Wolynes, Multiple-basin energy landscapes for large- amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. 103, 11844–11849 (2006).

Gouridis et al.
Structural dynamics in the evolution of a bifolobed protein scaffold

10 of 12 | PNAS https://doi.org/10.1073/pnas.2026151158
100. M. de Boer, G. Gouridis, Y. A. Muthahari, T. Cordes, Single-molecule observation of structural dynamics in the evolution of a bilobed protein scaffold. *Biochemistry* 43, 9395–9401 (2004).
101. K. Illergård, D. H. Ardell, A. Elofsson, Structure is three to ten times more conserved than secondary structure with quantitative mass spectrometry. *J. Mol. Biol.* 287, 1131–1133 (1999).
102. C. G. Wermuth, C. R. Ganellin, P. Lindberg, L. A. Mitscher, Glossary of terms used in medicinal chemistry (UPAC Recommendations 1998).  

Primary source: *Pure Appl. Chem. 70*, 1129–1143 (1998).

103. F. Fulyani et al., Functional diversity of tandem substrate-binding domains in ABC transporters from pathogenic bacteria. Structure 21, 1879–1888 (2013).
104. J. C. Wolters et al., Lignid binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS One* 5, e10361 (2010).
105. I. Alicea et al., Structure of the *Escherichia coli* phosphonate binding protein PhnD and rationally optimized phosphonate biosensors. *J. Mol. Biol.* 414, 356–369 (2011).
106. N. V. Grishin, Fold change in evolution of protein structures. *Protein Sci.* 14, 1579–1593 (2005).
107. S. G. Peisajovich, L. Rockah, D. S. Tawfeik, Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet.* 38, 168–174 (2006).
108. G. H. Schepers, J. A. Lycklama A Nijeholt, B. Poolman, An updated structural classification of substrate-binding proteins. *FEBS Lett.* 590, 4393–4401 (2010).
109. J. Cui, A. L. Davidson, ABC solute importers in bacteria. *Essays Biochem.* 50, 85–99 (2011).
110. J. J. Falke, G. L. Hazelbauer, Oligomerization and membrane binding of the talin head domain. *Biopolymers* 43, 3608–3619 (1994).
111. C. F. Higgins, ABC transporters: From microorganisms to man. *Annu. Rev. Cell Biol.* 6, 87–113 (1992).
112. Y. Zhang et al., Model of substrate-binding proteincromerocomplex structure supports intrasubunit signaling mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 96, 939–944 (1999).
113. K. Beena, J. B. Udgaonkar, R. Varadarajan, Effect of signal peptide on the stability and folding kinetics of maltose binding protein. *Biochemistry* 43, 3608–3619 (2004).
114. L. R. Randall, S. J. Hardy, Correlation of stoichiometry for enzyme, with quantitative mass spectrometry. *Biochemistry* 44, 1969–1977 (2005).
115. B. E. Clifton et al., Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. *Nat. Chem. Biol.* 14, 542–547 (2018).
116. A. Bar-Even, R. Milo, E. Noor, D. S. Tawfeik, The moderately efficient enzyme: Futile encounters and enzyme fitness. *Biochemistry* 54, 4969–4977 (2015).
117. E. Campbell et al., The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* 12, 944–950 (2016).
118. Y. L. Jiang et al., Coordinating carbon and nitrogen metabolic signaling through the cytoskeletal global repressor NdrH. *Proc. Natl. Acad. Sci. U.S.A.* 115, 403–408 (2018).
119. A. Tsirigotaki et al., Structural details of the OxyR peroxide-sensing mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6443–6448 (2015).
120. I. Jo et al., Structural details of the OxyR peroxide-sensing mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 112, 6443–6448 (2015).
121. D. Monferr et al., Structural studies on the full-length LysR-type regulator TssR from *Comamonas testonii* T-2 reveal a novel open conformation of the tetrameric LTTR fold. *Mol. Microbiol.* 75, 1199–1214 (2010).
122. S. Murakota 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, 555–566 (2003).
123. B. Pedre et al., Structural snapshots of OxyR reveal the peroxidatic mechanism of H2O2 sensing. *Proc. Natl. Acad. Sci. U.S.A.* 115, E11623–E11632 (2018).
124. E. Sainsbury et al., The structure of the CrgA protein from *Neisseria meningitidis reveals a new octameric assembly state for LysR-type transcriptional regulators. *Nucleic Acids Res.* 37, 4545–4558 (2009).
125. Y. Kim 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).
126. K. Baskaran, J. M. Duarte, N. Biyani, S. Bliven, G. Capitanii, A PD-B wide, evolution-based assessment of protein-protein interfaces. *BMC Struct. Biol.* 14, 22 (2014).
127. S. Bliven, A. Lafaia, A. Parker, G. Capitanii, J. M. Duarte, Automated evaluation of quaternary structures from protein crystals. *PLoS Comput. Biol.* 14, e1001604 (2018).
128. A. Prlic et al., Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26, 2983–2985 (2010).
129. Y. Ye, A. Godzik, Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19 (suppl. 2), i246–i253 (2003).
130. E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2256–2268 (2004).
131. L. Holm, L. Mauro, Dali server update. *Nucleic Acids Res.* 44, W351–W355 (2016).
132. E. Roberts, J. Eargle, D. Wright, Z. Luthey-Schulten, MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7, 392 (2006).
133. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38 (1996).
134. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410 (1990).
135. G. P. Evsinska, A. Matsumoto, H. Ishida, S. Hayward, A method for the analysis of domain movements in large biomolecular complexes. *Proteins* 76, 201–212 (2009).
136. D. Taylor, G. Cawley, S. Hayward, Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics* 30, 1319–1319 (2014).
137. A. Ben Chorin et al., ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci.* 29, 258–267 (2020).
138. O. Goldenberg, E. Erez, G. Nimrod, N. Ben-Tal, The ConSurf-DB: Pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res.* 37, D323–D327 (2009).
139. K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22, 195–201 (2006).

---

Gouridis et al.

Structural dynamics in the evolution of a bilobed protein scaffold

12 of 12 | PNAS doi:10.1073/pnas.2026151118

---

gni@cri.x.org, 20-03-2022