MATERIAL AND METHOD

Structural analysis & Phylogenetic trees

Proteins harbouring the *cherry-core* (CC) structure were identified using the “Structure similarity search function” of the PDB web server (1, 2). Protein structures were aligned using the PDBeFold server (3). 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 *Dataset* S2, 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 (4). In the summarized sequence- and structure-based phylogenetic trees as presented in Fig. 1C and *SI appendix*, 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 *Dataset* S2, 6.

Sequence-based phylogenetic trees as presented in *Dataset* S1 and summarized in *SI appendix*, Fig. S1 were computed using the MultiSeq software (5) embedded within the Visual Molecular Dynamics (VMD 1.9.2) software package (6). 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 *Dataset* S1 and summarized in *SI appendix*, 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 (7). Our cut-off for sequence coverage was 70%, which yielded ~600 proteins that allowed to construct an extended sequence-based phylogenetic tree. In all analysis, servers’ default settings were used unless specified.

Structural states of the *cherry-core*

Domain movements were detected via the DynDom protein domain motion analysis server (8, 9) for all the proteins that have high resolution 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 (*SI appendix*, Table. S2) in the Protein Interactions Calculator (PIC) webserver (10). 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 *Dataset* S7. 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

His_{10}SBD2 (Class B), $OpuACHis_6$ (Class C), $MalEHis_6$ (Class G) were expressed and purified as previously described (11-13). The *CC* of *cyn*R (Class A) gene (UniProtKB-P27111) was isolated from the genome of Escherichia coli DH5α (F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 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 H*I 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*) – ATCC 27184D-5. The primers were designed to exclude the signal peptide (1-35 amino acid); thus, to include the DNA sequence coding the protein chain 36-452 amino acid. PhnD (class D) gene (UniProtKB-Q1R3F7) 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(λS)). The primers were designed to exclude the signal peptide (1-26 amino acid); thus, to include DNA sequence coding the protein chain 27-338 amino acid. nFbp (class F) gene (UniProtKB-Q50964) was isolated from the genome of *Neisseria gonorrhoeae* – ATCC 700825D-5. The primers were designed to exclude the signal peptide (1-22 amino acid); thus, to include DNA sequence coding the protein chain 23-331 amino acid. (pro)*mal*E gene (UniProt:P0AEX9) was isolated from the genome of *Escherichia coli* K12. The primers were designed to include the signal peptide (1-26 amino acids). Primers introduced *Nde*I and *Hind*III restrictions sites and the gene product was sub-cloned into the pET20b vector (Merck).

The intergenic region of the 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(λS)) 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 191 bp DNA fragment, which 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 mutagenesis(14) and Megaprimer PCR mutagenesis(15) protocols.

His$_{10}$CynR (*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(λS)). Cells harbouring the plasmid expressing the protein were grown in Terrific Broth medium (30ºC; OD$_{600\ nm}$=0.5) and protein overexpression was induced by IPTG (0.10 mM; growth at 16ºC for 15 hours).

His$_{10}$PhnD and His$_{10}$nFbp were 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(λS)). Cells harbouring plasmids expressing the proteins were grown in LB medium (37ºC; OD$_{600\ nm}$=0.8) and protein overexpression was induced by IPTG (0.30 mM; for 4 hours).

His$_{10}$CmpA was overexpressed in BL21 pLysS DE3 cells (F–, *omp*T, *hsd*S$_{B}$ (r$_{B}$–, m$_{B}$–), *dcm*, *gal*, λ(DE3), pLysS, Cm^{r}). Cells harbouring the plasmid expressing the protein were grown (30ºC; OD$_{600\ nm}$=0.5) and protein overexpression (growth at 16ºC for 15 hours) was induced by IPTG (0.1 mM). For all proteins, growth was accomplished in Luria Bertani medium, except for CynR in which Terrific Broth was used.

proMalEHis$_{6}$ 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(λS)). Cells harbouring plasmids expressing the protein were grown in LB medium (37ºC; OD$_{600\ nm}$=0.8) and protein overexpression was induced by IPTG (1 mM; for 4 hours). 30 min prior to IPTG induction, 4mM Sodium Azide was added to the culture to inhibit SecA-dependent secretion that prevents cleavage of the signal peptide (16)

Proteins were purified as follows: after lysis by the cell disruptor (30.000 psi; 2 rounds), soluble material (50.000 g; 30 min; 4 ºC; 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; 1 M KCl, 10% glycerol; 10 mM Imidazole; 1 mM DTT; and 50 mM Tris-HCl, pH=8; 50 mM 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-Sartorious; Amicon-Millipore), dialyzed (50 mM Tris-HCl, pH=8; 50 mM KCl, 50% glycerol; 10 mM DTT), aliquoted and stored at -80ºC.

For untagged CynR (Full length): after lysis by cell disruptor (30.000 psi; 2 rounds), soluble material (50.000 g; 30 min; 4 ºC; Sorval) was loaded (50 mM Tris-HCl, pH = 8.3; 5 mM 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, 0,5 mM EDTA and 1 mM DTT), aliquoted and stored at -20ºC.

Purification of proMalEHis$_6$ from inclusion bodies was accomplished as described previously (17). Briefly, lysed samples (30.000 psi; 2 rounds) were centrifuged (50.000 g; 30 min; 4 ºC; Sorval) to separate the inclusion bodies. The pellet was solubilized (50 mM Tris-HCl, pH = 8; 100 mM KCl, 10% glycerol; 10 mM Imidazole; 1 mM DTT; 2 mM PMSF; 8 M Urea) with the aid of a Dounce homogenized and centrifuged (50.000 g; 30 min; 20 ºC; Sorval). The urea-solubilized pellet was diluted was diluted with (50 mM Tris-HCl, pH = 8; 100 mM KCl, 10% glycerol; 10 mM Imidazole; 1 mM DTT; 2 mM PMSF) to 6 M Urea before applying it on a Ni-NTA resin (Qiagen). Bound proteins were washed (50 mM Tris-HCl, pH = 8; 50 mM KCl, 10% glycerol; 2 mM Imidazole; 1 mM DTT; 6 M Urea and 50 mM Tris-HCl, pH=8; 50 mM KCl, 10% glycerol; 5 mM Imidazole; 1 mM DTT; 6 M Urea sequentially) and then eluted (50 mM Tris-HCl, pH=8; 50 mM KCl, 5% glycerol; 300 mM Imidazole; 1 mM DTT; 6M Urea). Protein fractions were pooled (supplemented with 5 mM EDTA; 50 mM DTT), concentrated (Vivacell-Sartorious; Amicon-Millipore), dialyzed (50 mM Tris-HCl, pH=8; 50 mM KCl, 10% glycerol; 10 mM DTT; 6 M Urea), aliquoted and stored at -80ºC. To prevent signal peptide cleavage all purification and dialysis buffers were supplemented with 5 mM $MgCl_2$. Prior to labelling, refolding was accomplished on the Ni-NTA resin (Qiagen) in 50 mM Tris-HCl, pH=8; 50 mM KCl, 1% glycerol; 5 mM Imidazole; 1 mM DTT by reducing the concentration of Urea progressively (6M, 5M, 4M, 3M, 2M, 1M) via 5 column volume washes. After refolding, proMalEHis$_6$ was handled as MalEHis$_6$.

Protein labelling

Protein labelling was accomplished as described previously (11). 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), FbpA (1SI1, 1SI0), PhnD (3S4U, 3QUJ) and MalE (1OMP, 1ANF); *SI Appendix*, 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, FbpA, PhnD, SBD1 and MalE were labelled with Alexa Fluor 555 and Alexa 647 (ThermoFischer) or AF555 and AF647 (Jena Bioscience). Histidine tagged proteins were immobilized on a Ni^{2+}-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, FbpA, PhnD and MalE; 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 material (18). 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 μM) was prepared in the dialysis buffer and was stepwise injected (0.5 μl) into the reaction cell containing 12 μM 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 50 μM K_2CO_3 and 1 μM EDTA. ITC experiments of CmpA were performed using MicroCal iTC200 (Malvern). The Calcium solution (500 μM) was prepared from $CaCl_2$ powder, which was subsequently diluted in the dialysis buffer and was injected (2 μl) into the reaction cell containing 40 μM of CmpA. All experiments were carried out at 280 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, FbpA, PhnD and MalE; 50 mM KPi, pH=7.4; 50 mM KCl for SBD2 and OpuAC) supplemented with additional reagents as stated in the text. For apo experiment of FbpA, 100mM of sodium citrate is added to chelate the trace of ferric. Beside 0.5 mM of ferric, excess amount of Phosphate ion (100mM) is added at saturating conditions for FbpA. The experiments were performed using a home-built confocal microscope similar to the setup described before (11) 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 µs and used for confocal excitation. Alternation between both excitation wavelengths was achieved by direct modulation of the two lasers. 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 µm above the interface of glass and water solution. Typical average laser powers were 30 µW at 532 nm (~30 kW/cm^2) and 15 µW at 637 nm (~15 kW/cm^2). 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 µm 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 search (19) using parameters M = 15, T = 500 µs 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 (20). 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 (11). Surface scanning was performed using a piezo stage with 100x100x20 µm 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 µW (~125 W/cm^2). Data acquisition was done with home-written Labview software (National Instruments). MalE was studied on standard functionalized cover-slides. Experiments were carried out in buffer (50 mM Tris-HCl, pH=7.4; 50 mM KCl) supplemented with 1 mM Trolox (22).

Time-traces were analysed by integrating the detected red and green photon streams in time-bins of 5 ms. 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) (23). For this an implementation of HMM was programmed in Matlab (MathWorks), based on previous studies (23). 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 (23). This was done with the forward-backward algorithm (24). Care was taken to avoid floating point underflow and was done as described (23). The most probable state-trajectory is then found from λ by using the Viterbi algorithm (25). 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 25ºC [5 μl of 28.25 μM 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 µl deuterated buffer [50 mM Tris-DCl pD 7.4, 50 mM KCl in D_2O (D_2O, 99.9% atom D; Euriso-top)] 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 uM and supplemented with 5 mM DTT. CynR (4 μl; 28.6 μM) was incubated at 25ºC for 30 min in the presence or absence of NaN_3, by addition of 1 μl of 0.5 M NaN_3 diluted in 50 mM Tris-HCl pH 7.4, 50 mM KCl. For the apo states, 1 μl of buffer was added instead. Deuterated buffer (95 μl; as for MalE) was added for 10, 100, (1000/ 2000 for the CC and CynR respectively), 10000 and 100000 (for the CC) sec at 25ºC, resulting in 95% v/v D_2O, 1.14 μM FL CynR, 5 mM NaN_3. 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 μM 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 μl; 32.85 μM protein; 5 mM DTT) was supplemented with 1.2 molar excess of DNA in H_2O (1.8 μl of 43.99 μM) and incubated at 25ºC 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 25ºC prior to addition of deuterated buffer (55.6 μl; as above) for 10, 1000, 10000 sec at 25ºC, 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 ($pD_{corrected}$ 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; 0.26μM CynR; 0.32μM 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; 0.32μM 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 (2ºC 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 µl/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 µl/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 μl/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 (*Dataset* S5-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 (*Dataset* S5-Tables F5B, F5C, F5D).

Statistical analysis: All HDX reactions were performed in triplicates, unless otherwise indicated (*Dataset* S5-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*., (26), described in Tsirigotaki *et al*., (27, 28). 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 ($1\text{-}p \geq 0.99$). An additional threshold was set at ±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 (*Dataset* S5-Table F5D and *Dataset* S4-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 *Dataset* S5-Tables F5C (for WT and MalE (M321K) apo states), *Dataset* S5-Table F5E (for WT MalE apo and holo states), *Dataset* S4-F4A (for DNA-bound CynR apo and holo states), *Dataset* S4-F4B (for DNA-free CynR apo and holo states) and *Dataset* S4-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 meting temperature $T_{m\ (apparent)}$, the changes in intrinsic tryptophan fluorescence emission of CynR (0.3 μM) as a function of increasing temperature (10 - 70 °C; ramping rate 0.8 °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 templates (29). 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 server (30). All models were reliable according to server thresholds (31-33).

SUPPLEMENTARY FIGURES

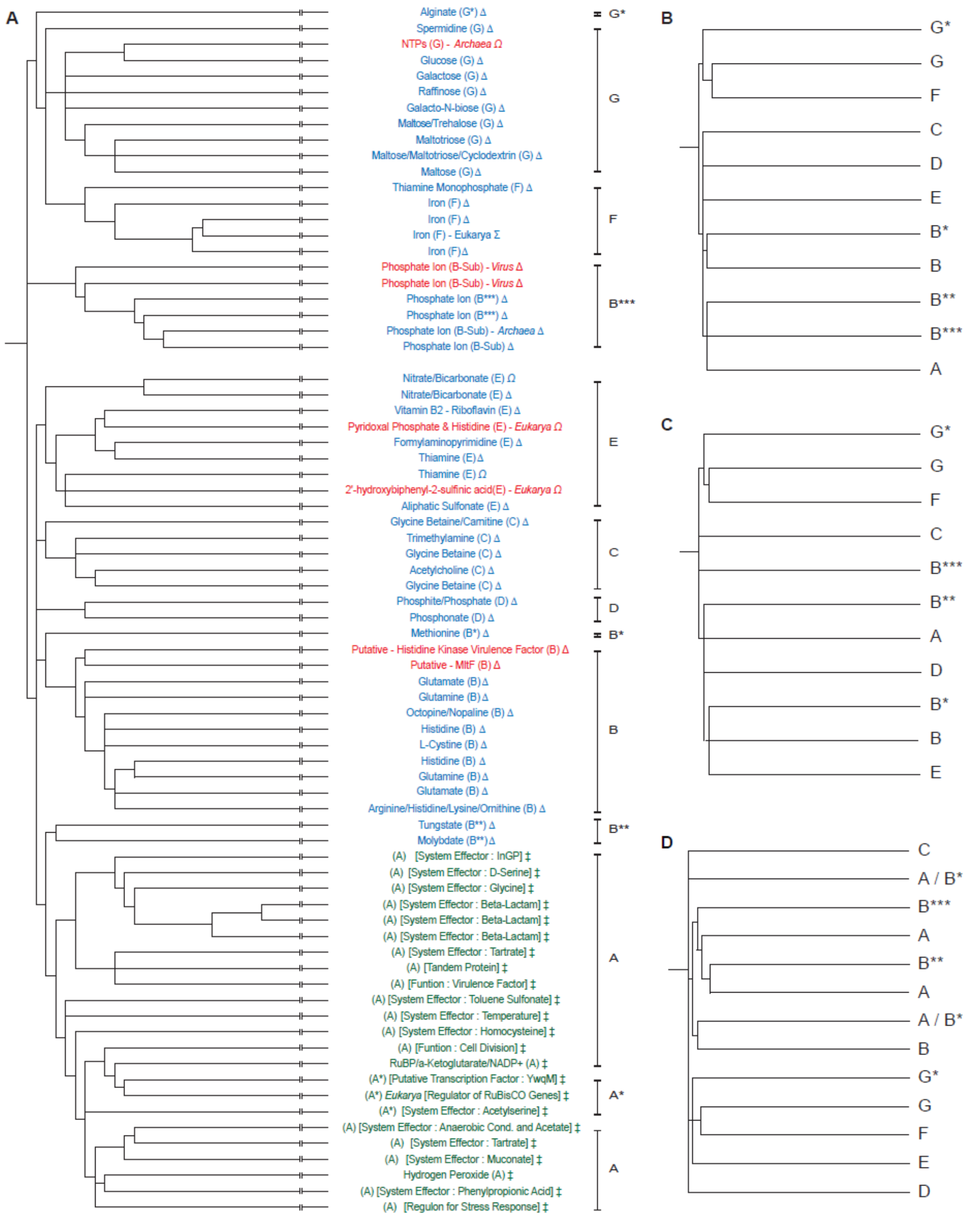

Figure S1. Phylogenetic Trees of *cherry-core* proteins (*CCPs*); related to Fig. 1C. (A) Summarized sequenced-based extended phylogenetic tree of the bilobed *CCPs*, grouped with respect to their ligands. Complete tree of ~ 600 proteins is shown in *Dataset* S1. Function was obtained from UniProtKB (34) 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 *SI Appendix*, Fig. S3. **(**B-D) Summarized (complete trees in *Dataset* S2) sequence-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 *Dataset* S3). 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 outcome is in line with the fact that the C-tail–domain interactions stabilize a specific (open) structural state (Fig. 2 and *SI Appendix*, 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 structure elements (see also Fig. 1D and *SI Appendix*, Fig. S2). On the other hand, the 'upper clade' includes *CCPs* with minor alterations/variability of their *CC*.

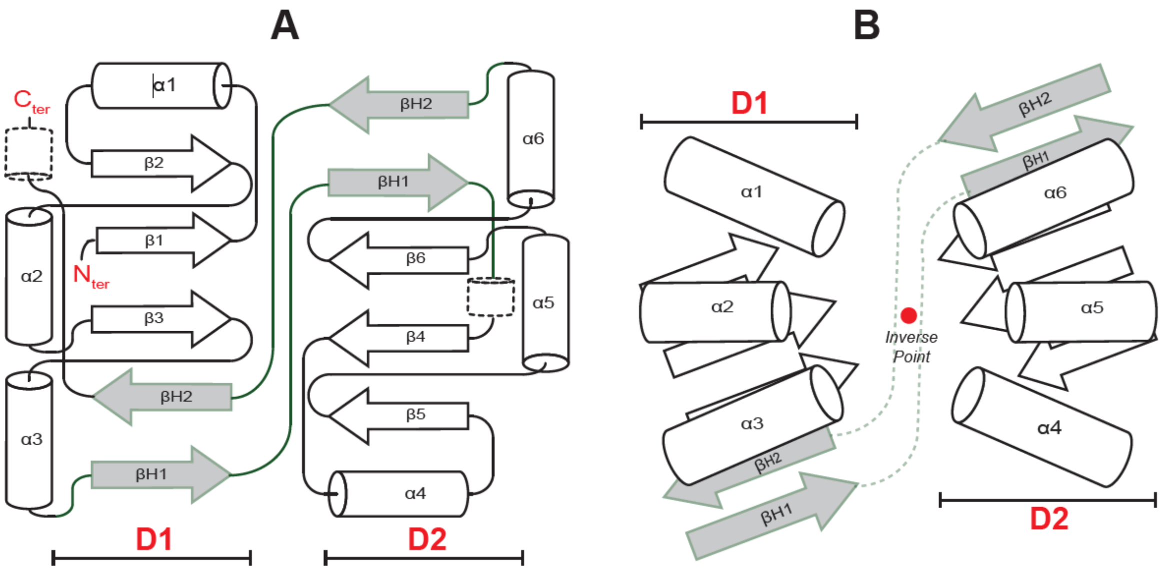

Figure S2. Structural Consensus of *cherry-core* proteins; related to Fig. 1B-D. (A) Consensus/model derived from secondary structure alignments of the *CC* of *CCPs* (Fig. 1D and *Dataset* S3**)**. (B) Schematics of the geometrical consensus for the 3D placement of secondary structure elements originated 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.

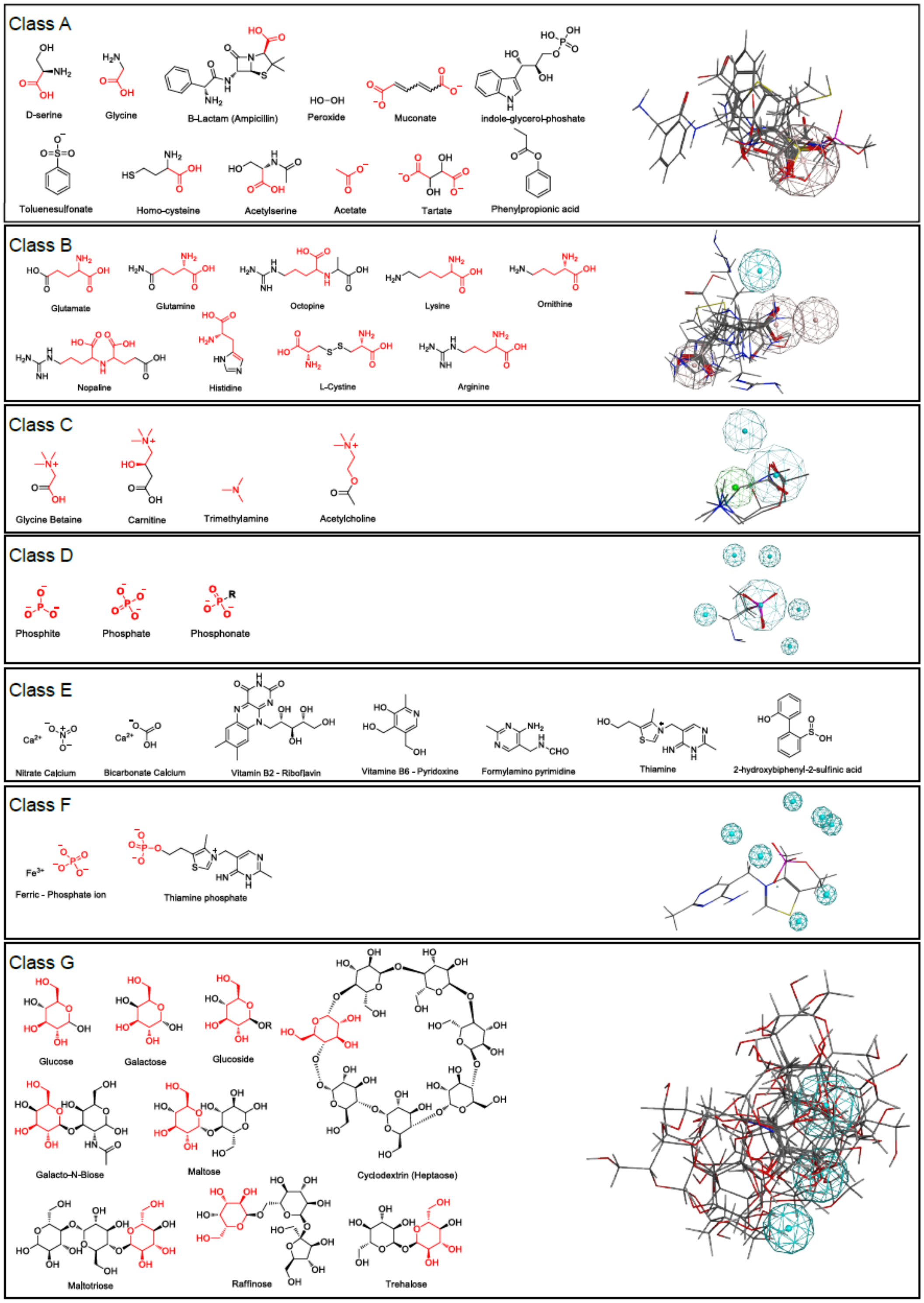

Figure S3. Common Chemical Structure and pharmacophore structure of the cherry-core protein ligands; related to Figure 2**.** The common chemical structure of all identified ligands for each of the *CCP* classes (A-G) is highlighted with red color. The ligands of class A, but mostly class E do not share 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).

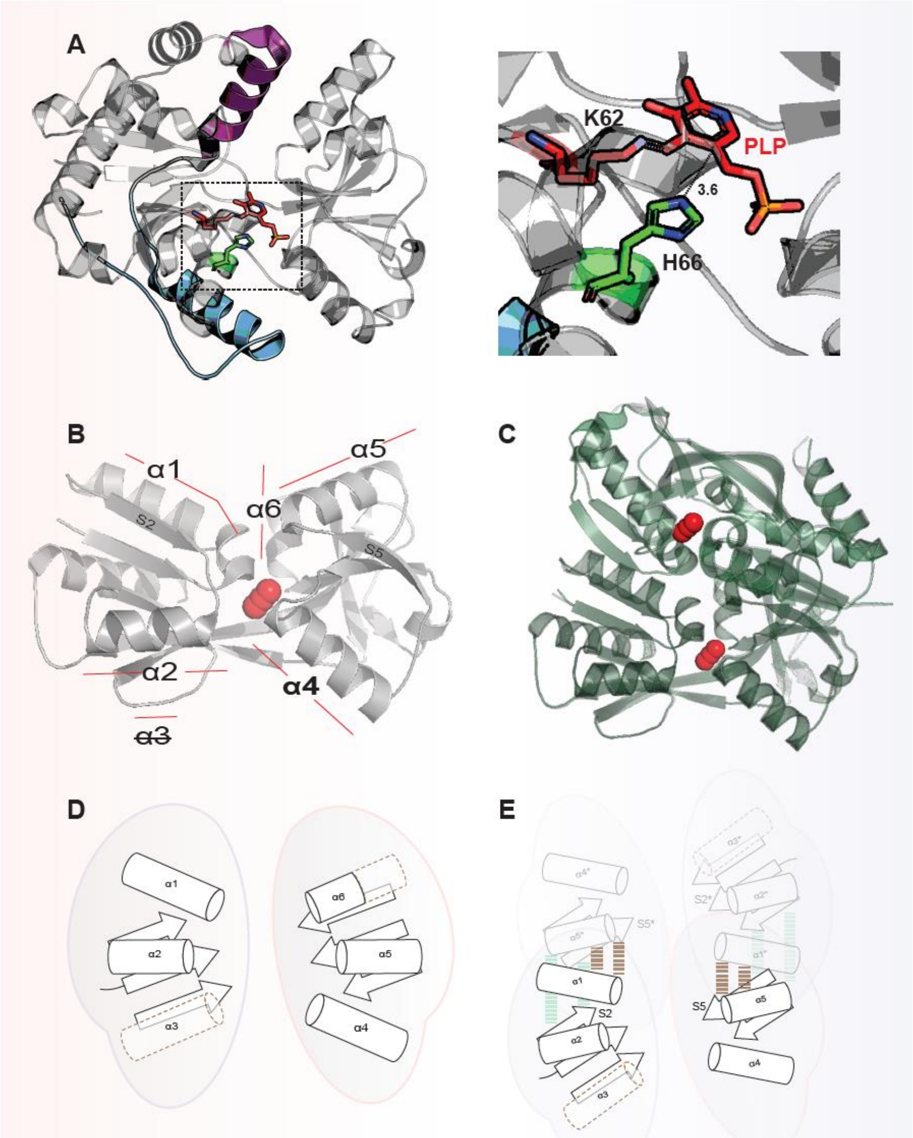

Figure S4. Structural analysis of the single-turnover enzyme pyrimidine synthase THI5p and of the oligomeric transcription factor CynR, related to Fig. 2 and Fig. 4. (A) Crystal structure (PDB:4ESX) of the single-turnover enzyme thiamin pyrimidine synthase (35) in complex with its ligand pyridoxal phosphate (red sticks). The 2 symmetric helices of the C-tail that stabilize the cleft in the closed state 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 asymetric 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 (*SI Appendix*, Fig. S2) α3 is missing from the *CC* of CynR and all class A proteins (Fig. 1D and *Dataset* S3). (E) Schematic representation of the dimeric *CC* of class A proteins. Interacting secondary structure elements are highlighted. α1 makes hydrophobic contacts with S5 and α5, while S2 hydrophobic contacts with α5 and hydrophobic contacts/hydrogen bonds with S5.

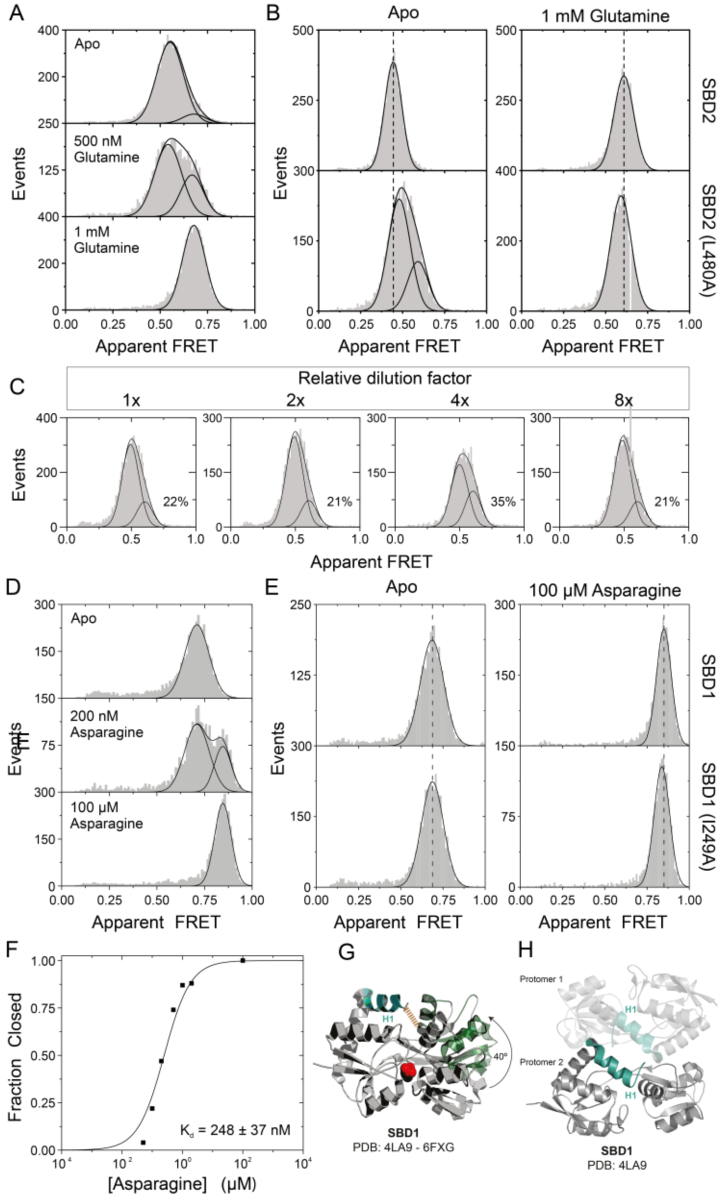

Figure S5. Structural dynamics of SBD2, SBD1, and derivatives, related to Fig. 5. (A) Apparent FRET efficiency histogram of freely diffusing SBD2 (L480A) labelled with the donor fluorophore Alexa Fluor 555 and acceptor fluorophore Alexa Fluor 647. The fraction of closed state is similar as found for SBD2 (L480A) labelled with the donor fluorophore Cy3B and the acceptor fluorophore Atto647N, as presented in Fig. 5C. (B) The apparent FRET efficiency peak 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. (D) Apparent FRET efficiency histograms of freely diffusing SBD1 (I249A) labelled with the donor fluorophore AF555 and acceptor fluorophore AF647. Here, no fraction closed is present in the absence of ligand. (E) The apparent FRET efficiency peak for the *apo* (open) state of SBD1 (I249A) is found to be identical as for SBD1, and for the *holo* (closed, 100 µM asparagine) state only a minor shift (~0.01) is found, which would suggest that the mutation I249A does not alter the energy landscape of SBD1. (F) Fraction of the closed state (high FRET state) of SBD1 (I249A) as a function of asparagine concentration. The K_d of 248 nM is similar to what has been found before for SBD1. (G) The *apo* and *holo* structures of SBD1 where superimposed as in Fig. 2. (H) The protomer arrangement in the open SBD1 structure (4LA9). The dimerization interface involves uniquely H1, as indicated. The inter-protomer contacts impact the intra-protein stabilizing open state interactions of SBD1 (compare SBD1 *vs* SBD2 contacts in SI *Appendix*, Table. S2).

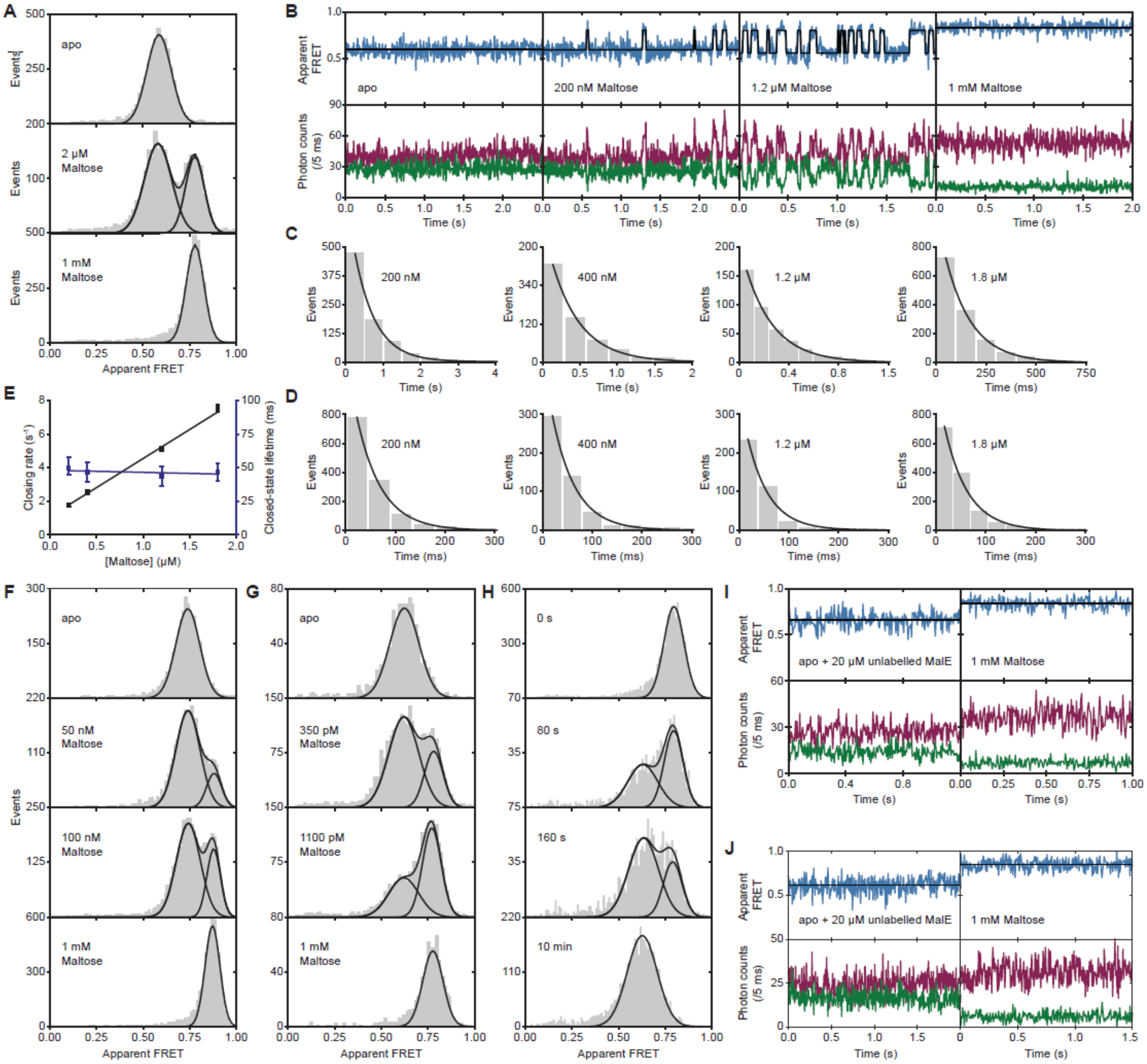

Figure S6. Structural dynamics of MalE and derivatives; related to Fig. 5 E-I. (A) Apparent FRET efficiency histogram of MalE, obtained from the solution-based smFRET and ALEX measurements under conditions as indicated. (B) Fluorescence trajectories of MalE under different conditions as indicated; donor (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 µM unlabeled MalE (M321K) added to scavenge all maltose. First 10 nM maltose was added to saturate the protein (top panel). Subsequently ~20 µM 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* (~20 µM unlabeled protein) or 1 mM maltose conditions. Fluorescence trajectories of MalE derivatives are plotted similarly to (B).

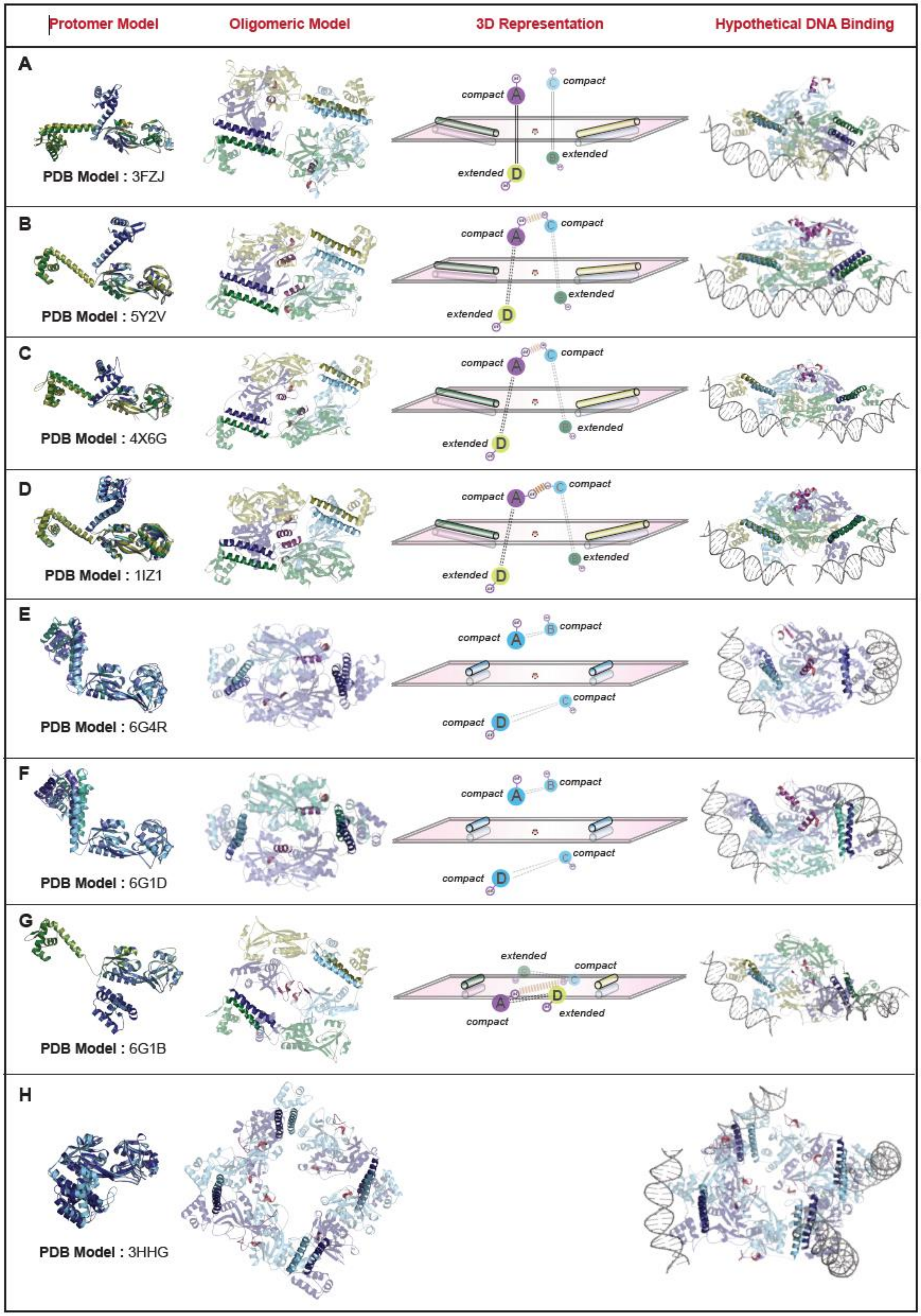

Figure S7. Models of Oligomeric States of CynR, related to Fig. 4. The CynR sequence (UniprotKB: P27111) was modelled from the indicated PDB structures that were used as a template, by the SWISS-MODEL server. The DNA in all structures was modelled from the structure of the HTH-DNA binding domain of CbnR in complex with the RBS of the cbnA promoter (36). **(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 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* (*SI Appendix*, 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 (*Dataset* S7) 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 (α4 and CH2, see *Dataset* S3) 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 bellow 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 (37). 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) (38). 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 is such quaternary assemblies (*SI Appendix*, Table. S4). The differences between the protomers of these assemblies involve minor rearrangements of secondary structure elements and partial loss of secondary structure (in *Dataset* S3 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 (*Dataset* S7). Moreover, the prominent octameric population found for 6G1B (37) (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 (*Dataset* S7). 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 (*Dataset* S6). 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* (*SI Dataset* S3): the additional helix CH5 connecting Hinge-β-sheet 2 (βH2) and CH6 and it lacks helix α4 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 fixates it in the compact configuration with the DNA binding domain below the *CC*. Helix α4 that stabilizes tetrameric assemblies is not required for formation of the octamer, and probably its absence facilitates octamerization as it destabilizes tetramerization.

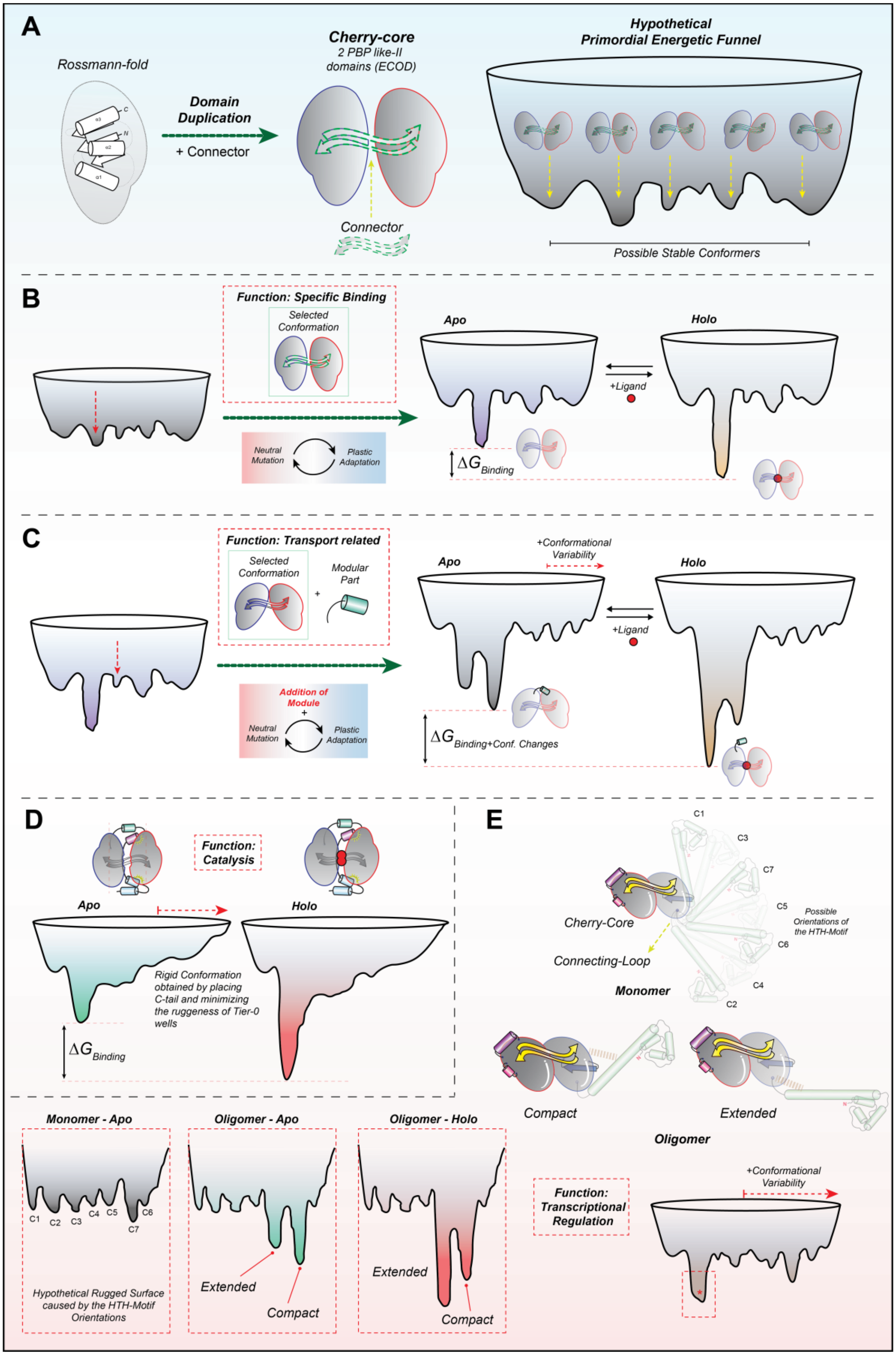
A
Rossmann-fold
Domain Duplication
+ Connector
Cherry-core
2 PBP like-II domains (ECOD)
Connector
Hypothetical Primordial Energetic Funnel
Possible Stable Conformers
B
Function: Specific Binding
Selected Conformation
Neutral Mutation
Plastic Adaptation
Apo
Holo
+Ligand
ΔG Binding
C
Function: Transport related
Selected Conformation
Modular Part
+
Addition of Module
+
Neutral Mutation
Plastic Adaptation
+Conformational Variability
Apo
Holo
+Ligand
ΔG Binding+Conf. Changes
D
Function: Catalysis
Apo
Holo
Rigid Conformation obtained by placing C-tail and minimizing the ruggeness of Tier-0 wells
ΔG Binding
E
C1
C3
C7
C5
C6
C4
C2
Cherry-Core
Connecting-Loop
Possible Orientations of the HTH-Motif
Monomer
Compact
Extended
Oligomer
Monomer - Apo
C1
C2
C3
C4
C5
C6
C7
Hypothetical Rugged Surface caused by the HTH-Motif Orientations
Oligomer - Apo
Extended
Compact
Oligomer - Holo
Extended
Compact
Function: Transcriptional Regulation
+Conformational Variability

Figure S8. Hypothetical Energetic Funnels for the visualization of the effects of Modularity on the Evolvability; related to Fig. 6. (A) According to the view of generalists (39), an ancestral structure composed of two Rossmann-like domains can acquire many distinct structural states depending on the arrangement of the two domains relative to each other. The exchange between the different conformers is governed by the properties of the connector. 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. 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 increased ligand promiscuity compared to the evolved *CCPs*. (C-E) Addition of the modular parts paves the way to select a second well on the funnel. If the presence of the module is beneficial for evolving the new function, nature may then introduce additional mutations 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 "select" the well of the beneficial structural state. (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 structurally "robust" protein, providing the required stereochemistry necessary for specific reactions (see *SI Appendix*, 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.

SUPPLEMENTARY TABLES

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

	***Apo* Conformation (PDB ID code)**	***Holo* 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
	4LA9	6FXG	D2	39,7	1,0
	3ZSF	2YLN	D2	60,1	-1,1
Class C	3L6G	3L6H	D2	35	-0,2
	1SW5	1SW2	D2	58,9	-0,3
Class D	3S4U	3QUJ	D2	74,3	-0,5
Class E	2X7P	2X7Q	D2	13,4	-0,3
	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

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).

	***apo* conformation**			***holo* conformation**		
	Residue	**Interaction**	**Location**	**Residues**	**Interaction**	**Location**
Class A (2HXR - 3HFU	-	-	-	-	-	-
Class B – SBD2 (4KR5 – 4KQP)	Leu480-Pro419	H	H1			
	Lys426-Asp477	MSHB	H1			
	Lys426-Lys589	MSHB	H1			
	Glu481-Lys426	I	H1			
	Asp483-Lys422	I	H1			
	Tyr423-Tyr479	A	H1			
Class B – SBD1 (4LA9 – 6FXG)	Ile249-Pro186	H	H1			
Class C (3L6G - 3L6H)	Trp488Leu545	H	H3	Pro486-Met544	H	H3
	His487-Glu541	MSHB	H3			
	Glu541-His487	MSHB	H3			
Class D (3S4U - 3QUJ)	Met70-Leu254	H	HD1	Val73-Leu254	H	HD1
	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			
	Gln259-Tyr137	SSHB	HD1			
	Lys141-Asp295	I	HD2			
Class E (2X7P – 2X7Q)	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
	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 F (1SI1 – 1SI0)	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			
	Phe220-Tyr269	A	H1-H2			
	Tyr301-Met64	AS	H5			
Class G (1OMP – 1ANF)	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
	Trp158-Glu328	SSHB	H4-H5			
	Tyr90-Met321	AS	H4			

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.

	Tm (^{0}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 *SI Appendix*, 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 Cα-Cα 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 Fig. 5. The PDB ID codes of proteins are reported in the SI Appendix, Methods section and SI Appendix Table. S1.

	Apparent FRET efficiency		Cα- Cα 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
SBD1 (T159C G87C)	0.612 ± 0.003	0.805 ± 0.002	44.8	34.2
SBD2 (T369C S451C)	0.586 ± 0.002	0.736 ± 0.002	49.0	40.1
nFbp (N184C Q237C)	0.429 ± 0.003	0.470 ± 0.001	52.8	47.3
PhnD (N8C P130C)	0.466 ± 0.004	0.647 ± 0.002	54.8	45.4
ProMalE(T36C S352C)	0.558 ± 0.007	0.762 ± 0.002	51.3	40.8

SUPPLEMENTARY DATASETS

Dataset S1. Extended Sequence Based Phylogenetic tree. Sequence-based extended phylogenetic tree based on percentage of sequence identity. Summary of this tree is presented in *SI Appendix*, Fig. S1A. Every protein is represented by its PDB ID (known structures) or UniProt ID followed by its localization 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.

Dataset S2. Complete 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 *SI Appendix*, 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.

Dataset 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. 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.

Dataset 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.

Dataset 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.

Dataset S6. Structure based phylogenetic trees of class A *CCPs*. Results as obtained from the Dali server (see *SI Appendix*, Material and Method). The structure based phylogenetic tree of class A proteins with known oligomeric structures (see *SI Appendix*, Fig. S7). Top part contains full length class A transcription factors (HTH motif/dimerization helix and CC) while bottom part, only the *CC*.

Dataset S7. Interaction of CynR connecting loop/dimerization helix with the CC. The *cynR* sequence was modelled after the known oligomeric high-resolution structures (see *SI Appendix*, Fig. S7 and Material and Method 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 *SI Appendix*, Table S2, see also *SI Appendix* Material and Method). 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 *Dataset* S3). The nature of the interactions, hydrophobic, aromatic, Cation-P_i, hydrogen-bonds, Ionic are represented with blue, cyan, green, orange and red color respectively.

SUPPLEMENTARY REFERENCES

1. A. Prlic *et al.*, Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* **26**, 2983-2985 (2010).
2. Y. Ye, A. Godzik, Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* **19 Suppl 2**, ii246-255 (2003).
3. 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).
4. L. Holm, L. M. Laakso, Dali server update. *Nucleic acids research* **44**, W351-355 (2016).
5. E. Roberts, J. Eargle, D. Wright, Z. Luthey-Schulten, MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC bioinformatics* **7**, 382 (2006).
6. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).
7. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
8. G. P. Poornam, A. Matsumoto, H. Ishida, S. Hayward, A method for the analysis of domain movements in large biomolecular complexes. *Proteins* **76**, 201-212 (2009).
9. D. Taylor, G. Cawley, S. Hayward, Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics* **30**, 3189-3196 (2014).
10. K. G. Tina, R. Bhadra, N. Srinivasan, PIC: Protein Interactions Calculator. *Nucleic acids research* **35**, W473-476 (2007).
11. G. Gouridis *et al.*, Conformational dynamics in substrate-binding domains influences transport in the ABC importer GlnPQ. *Nature structural & molecular biology* **22**, 57-64 (2015).
12. J. C. Wolters *et al.*, Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PloS one* **5**, e10361 (2010).
13. M. de Boer *et al.*, Conformational and dynamic plasticity in substrate-binding proteins underlies selective transport in ABC importers. *Elife* **8** (2019).
14. J. W. Bok, N. P. Keller, Fast and easy method for construction of plasmid vectors using modified quick-change mutagenesis. *Methods in molecular biology* **944**, 163-174 (2012).
15. C. W. Vander Kooi, Megaprimer method for mutagenesis of DNA. *Methods in enzymology* **529**, 259-269 (2013).
16. D. B. Oliver, R. J. Cabelli, K. M. Dolan, G. P. Jarosik, Azide-resistant mutants of Escherichia coli alter the SecA protein, an azide-sensitive component of the protein export machinery. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 8227-8231 (1990).
17. G. Gouridis, S. Karamanou, M. Koukaki, A. Economou, In vitro assays to analyze translocation of the model secretory preprotein alkaline phosphatase. *Methods in molecular biology* **619**, 157-172 (2010).
18. F. Husada *et al.*, Watching conformational dynamics of ABC transporters with single-molecule tools. *Biochemical Society transactions* **43**, 1041-1047 (2015).
19. E. Nir *et al.*, Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *The journal of physical chemistry. B* **110**, 22103-22124 (2006).
20. A. N. Kapanidis *et al.*, Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8936-8941 (2004).
21. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nature methods* **5**, 507-516 (2008).
22. T. Cordes, J. Vogelsang, P. Tinnefeld, On the mechanism of Trolox as antiblinking and antibleaching reagent. *Journal of the American Chemical Society* **131**, 5018-5019 (2009).
23. L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition" in Readings in speech recognition, W. Alex, L. Kai-Fu, Eds. (Morgan Kaufmann Publishers Inc., 1990), pp. 267-296.
24. L. E. Baum, T. Petrie, Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Annals of Mathematical Statistics* **37**, 1554-& (1966).
25. A. J. Viterbi, Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. *Ieee Transactions on Information Theory* **13**, 260-+ (1967).

26. M. J. Bennett, K. Barakat, J. T. Huzil, J. Tuszynski, D. C. Schriemer, Discovery and characterization of the laulimalide-microtubule binding mode by mass shift perturbation mapping. *Chem Biol* **17**, 725-734 (2010).
27. A. Tsirigotaki, R. V. Elzen, P. V. Veken, A. M. Lambeir, A. Economou, Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry. *Sci Rep* **7**, 2456 (2017).
28. A. Tsirigotaki, M. Papanastasiou, M. B. Trelle, T. J. Jorgensen, A. Economou, Analysis of Translocation-Competent Secretory Proteins by HDX-MS. *Methods Enzymol* **586**, 57-83 (2017).
29. L. Bordoli *et al.*, Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**, 1-13 (2009).
30. 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).
31. M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, T. Schwede, Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep* **7**, 10480 (2017).
32. S. Bienert *et al.*, The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res* **45**, D313-D319 (2017).
33. A. Waterhouse *et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
34. C. UniProt, UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* **47**, D506-D515 (2019).
35. R. Y. Lai *et al.*, Thiamin pyrimidine biosynthesis in Candida albicans : a remarkable reaction between histidine and pyridoxal phosphate. *Journal of the American Chemical Society* **134**, 9157-9159 (2012).
36. M. P. Koentjoro, N. Adachi, M. Senda, N. Ogawa, T. Senda, 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).
37. S. Sainsbury *et al.*, The structure of CrgA from Neisseria meningitidis reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic acids research* **37**, 4545-4558 (2009).
38. B. Pedre *et al.*, Structural snapshots of OxyR reveal the peroxidatic mechanism of H2O2 sensing. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E11623-E11632 (2018).
39. D. S. Tawfik, I. Gruic-Sovulj, How evolution shapes enzyme selectivity - lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *FEBS J* **287**, 1284-1305 (2020).