Structural Basis for the Recognition of the Bacterial Tyrosine Kinase, Wzc by its Cognate Tyrosine Phosphatase, Wzb

Sébastien Alphonse[1,†], Imane Djemil[2,†], Andrea Piserchio[1] and Ranajeet Ghose[1,2,3,4*]

[1]Department of Chemistry and Biochemistry, The City College of New York, New York, NY 10031.

PhD Programs in [2]Biochemistry, [3]Chemistry and [4]Physics, The Graduate Center of CUNY, New York, NY 10016.

[†]These authors contributed equally to this work.

*Address correspondence to Ranajeet Ghose – rghose@ccny.cuny.edu

Keywords: Bacterial tyrosine kinase, low molecular weight protein tyrosine phosphatase, solution NMR, relaxation dispersion, paramagnetic relaxation enhancement.

ABSTRACT

BY-kinases comprise a family of protein tyrosine kinases that are structurally distinct from their functional counterparts in eukaryotes and are highly conserved across the bacterial kingdom. BY-kinases act in concert with their counteracting phosphatases to regulate a variety of cellular processes, most notably the synthesis and export of polysaccharides involved in biofilm and capsule biogenesis. Biochemical data suggest that BY-kinase function involves the cyclic assembly and disassembly of oligomeric states coupled to the overall phosphorylation levels of a C-terminal tyrosine cluster. This process is driven by the opposing effects of intermolecular auto-phosphorylation, and dephosphorylation catalyzed by tyrosine phosphatases. In the absence of structural insight into the interactions between a BY-kinase and its phosphatase partner in atomic detail, the precise mechanism of this regulatory process has remained poorly defined. To address this gap in knowledge, we have determined the structure of the transiently assembled complex between the catalytic core of the *Escherichia coli* (K-12) BY-kinase, Wzc, and its counteracting low molecular weight protein tyrosine phosphatase (LMW-PTP), Wzb, using solution NMR-techniques. Unambiguous distance restraints from paramagnetic relaxation effects were supplemented with ambiguous interaction restraints from static spectral perturbations and transient chemical shift changes inferred from relaxation dispersion measurements and used in a computational docking protocol for structure determination. This structure, that represents the first atomic view of the mode of interaction between a BY-kinase/phosphatase pair, and indeed of any LMW-PTP, prokaryotic or eukaryotic, and its kinase substrate, provides mechanistic insight into the phosphorylation-coupled assembly/disassembly process proposed to drive BY-kinase function.

SIGNIFICANCE

This study provides the first view in atomic detail of the recognition of a bacterial tyrosine kinase (BY-kinase) by its counteracting phosphatase. BY-kinase function has been proposed to entail the cyclic assembly/disassembly of an oligomer coupled to the phosphorylation level of a C-terminal tyrosine cluster (Y-cluster). The structure of BY-kinase/phosphatase complex highlights use of the same conserved element on the kinase for both oligomerization and phosphatase docking. This mode of interaction prevents phosphatase engagement and dephosphorylation before exceeding a critical threshold of Y-cluster phosphorylation necessary for oligomer disassembly. Reassembly occurs once the Y-cluster is sufficiently dephosphorylated by the phosphatase that is then excluded, thereby repriming the cycle. This phosphorylation-coupled assembly/disassembly process likely serves to temporally coordinate intra- and extra-cytoplasmic events.

The BY-kinase (*b*acterial t*y*rosine kinase) family of protein tyrosine kinases (PTKs) are highly conserved within the bacterial domain and regulate a variety of physiological processes (1-5). Most notable among these processes is the synthesis and export of the polysaccharide constituents of biofilms and capsules (2, 4) that are cytoprotective under environmental stress and contribute to the virulence of several pathogenic species (6, 7). In Gram-negative bacteria, BY-kinases are anchored to the inner membrane through a two-pass transmembrane domain (TMD) that links a periplasmic domain (PD) to a C-terminal cytoplasmic catalytic domain (CD) (*SI Appendix* Fig. S1A) (8). Through the association of the PD with the periplasmic component of the outer membrane transporter, Wza (9), BY-kinases, form part of a platform (10) that spans both the outer and inner membranes. This assembly enables BY-kinases to influence two distinct processes, the synthesis, and the export of polysaccharides, by interactions within both the cytoplasmic and the periplasmic compartments.

BY-kinase CDs (8, 11-13) differ significantly in sequence (14) and structure (15) from eukaryotic protein tyrosine kinases (ePTKs). Rather, they contain sequence elements and display structural folds characteristic of the P-loop ATPases (16) (*SI Appendix* Fig. S1A). BY-kinase function has been linked to the auto-phosphorylation of a cluster of tyrosine residues (Y-cluster) located at the C-terminus of the CD (*SI Appendix* Fig. S1A) (17, 18) and its subsequent dephosphorylation by a counteracting protein tyrosine phosphatase (PTP) (19, 20). Unlike in many ePTKs where activity is related to the phosphorylation states of specific regulatory residues (21), BY-kinase function appears to be driven by cycling between states where the Y-cluster is highly phosphorylated (Y-high) and those where the level of Y-cluster phosphorylation is greatly reduced (Y-low) (18, 22) through the catalytic activity of PTPs (23). The Y-low states assemble in an octameric ring to enable efficient auto-phosphorylation of the Y-cluster in an intermolecular fashion (8, 12). Successive rounds of phosphorylation and generation of Y-high states results in ring disassembly; the ring reassembles after multiple rounds of PTP-mediated dephosphorylation and recreation of the Y-low states. This phosphorylation-linked assembly/disassembly cycle, that may be conceptualized as a supra-molecular conformational change, is central to the biological function of BY-kinases (23). This highlights the functional importance of maintaining the appropriate balance between self-interactions, to enable *in trans* auto-phosphorylation in BY-kinases, and those with a PTP, to facilitate dephosphorylation. While the atomic details of the self-interactions in BY-kinases have been highlighted by several structures (8, 12, 13), no such structural information is available for the interactions of the BY-kinases with their counteracting PTPs.

In prior studies, we had utilized NMR spectral perturbations to analyze the interactions between the CD of the *Escherichia coli* (K-12) BY-kinase, Wzc (Wzc$_{\mathrm{CD}}$), and its cognate low molecular weight PTP (LMW-PTP), Wzb (*SI Appendix* Fig. S1B) (24). These studies suggested that a conserved motif, ^{508}Glu–[Xxx]$_2$–Arg–[Xxx]$_2$–Arg^{514}, on the Wzc$_{\mathrm{CD}}$ $\alpha 2$ helix (*SI Appendix* Fig. S1A), necessary for oligomerization (12), to also be involved in binding Wzb (24). Since those studies relied solely on NMR spectral perturbations, they did not resolve atomic details of the interactions that define the Wzc$_{\mathrm{CD}}$/Wzb interface. Here we utilize solution NMR methods to derive a variety of structural restraints to generate an atomic model of the Wzb•Wzc$_{\mathrm{CD}}$ complex that is validated through functional studies *in vitro*. This structure represents the first view of the interaction between a BY-kinase and its cognate PTP, and indeed of any eukaryotic or prokaryotic LMW-PTP with its kinase partner, in atomic detail, and provides critical insight into the role of this interaction in driving BY-kinase function.

RESULTS

We have previously shown that the disordered C-terminal tail of $\mathrm{Wzc}_{\mathrm{CD}}$, that includes the Y-cluster (residues Tyr705-Lys720), is dispensable for its interaction with Wzb (24). A non-hydrolyzable phospho-tyrosine mimic was also found to make minimal contributions to the binding free energy (24). Therefore, we utilized a construct, $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$, encoding the catalytic core of Wzc (Ser447-Ala704) for our structural studies. We relied on two-dimensional line-shape analyses of resonances (25) in methyl ^{13}C,^{1}H HMQC spectra of ILV-labeled Wzb in the presence of increasing amounts of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ (and its variants; described below) to quantify their mutual affinities. In the case of wild-type $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$, this approach yields a K_D of 8.6 ± 0.6 μM (Table 1) comparable with that obtained from ITC (10.7 ± 1.3 μM, *SI Appendix* Fig. S2), and that previously obtained using surface plasmon resonance measurements (SPR, 5.1 ± 0.2 μM) (24).

We had previously noted extensive line broadening in ^{15}N, ^{1}H TROSY spectra of Wzb in the presence of an equimolar amount of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ (24) suggesting substantial exchange contributions attributable to the binding/unbinding process. Indeed, these effects prevented the use of spectral perturbations to precisely identify the $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$-binding residues on Wzb. To identify Wzb "seed" residues that facilitate $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ recognition, we analyzed the transient chemical shift changes induced on ^{15}N,^{2}H-labeled Wzb by the presence of sub-stoichiometric amounts of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ that manifest as exchange contributions to transverse relaxation (R_{ex}) in CPMG-based ^{15}N relaxation dispersion measurements (26). While Wzb does not exhibit significant exchange in the absence of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ (*SI Appendix* Fig. S3A), specific regions display exchange effects in the presence of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$, and the corresponding R_{ex} values are enhanced by increasing concentrations of the latter (and the fraction of Wzb within the Wzb•$\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ complex). That these R_{ex} values are induced by binding/unbinding processes involving $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ is reinforced by the fact that the limiting relaxation rates, $R_{2,eff}(\infty)$ (inferred from the corresponding values at the highest field used, 1000 Hz) in the fast pulsing regime for almost all residues (*SI Appendix* Fig. S4 for representative examples) are linearly dependent on the bound fraction (f_B).

As shown in Figs. 1A-E, both the number of Wzb residues that display significant R_{ex} values, and the values themselves, are enhanced with increasing f_B. The distribution of residues with substantial R_{ex} values plotted on the Wzb surface for $f_B = 0.37$ largely mirrors the spatial pattern of attenuations seen for the amide resonances of Wzb in the presence of an equimolar ratio of $Wzc_{CD\Delta C}$ (Fig. 1F) reported previously (24). At the lowest f_B value (0.05), F2, the catalytic C9 (on β1', see *SI Appendix* Fig. S1B), and R74 (α3') show statistically significant R_{ex} values > 5 s^{-1}. At $f_B = 0.24$, the R_{ex} values for F2, C9, R74, in addition to N75 and D77 (on the α3'-β3' loop), I79 (on β3'), and M82 (on the β3'-α4' loop) all exceed 15 s^{-1} (*SI Appendix* Fig. S3). Within this group of residues, F2, R74, N75 and D77 are all significantly solvent exposed and their substantial R_{ex} values suggest that they are proximal to, if not in direct contact, with $Wzc_{CD\Delta C}$ within the $Wzb•Wzc_{CD\Delta C}$ complex. These transient chemical shift changes (as reflected by the significant R_{ex} values) on Wzb were utilized (see *SI Appendix* Table S2 and the *SI Appendix* Materials and Methods for details) to generate ambiguous interaction restraints (AIRs) in our structure calculation protocol discussed below.

To obtain a direct measure of distances between the interacting partners, we relied on the measurement of paramagnetic relaxation enhancement (PRE) effects induced on Wzb by spin-labeled $Wzc_{CD\Delta C}$ variants. First, the two native cysteines on $Wzc_{CD\Delta C}$ were mutated to Ser (Cys544Ser/Cys563Ser) to generate a cysteine-less background ($Wzc_{CD\Delta C,\Delta Cys}$) to which non-native cysteines could be introduced, one at a time, for the subsequent attachment of a spin-label (*SI Appendix* Fig. S5). $Wzc_{CD\Delta C,\Delta Cys}$ expresses well and can be purified similarly as $Wzc_{CD\Delta C}$. The

^{15}N,^{1}H TROSY spectrum of this mutant displays well-dispersed resonances indicative of a folded protein and suggests the absence of any significant structural reorganization compared to wild-type Wzc$_{\mathrm{CD\Delta C}}$ (*SI Appendix* Fig. S6A) with chemical shift perturbations (average ± standard deviation over all resonances excluding those of residues within a 6 Å radius of the mutation sites) of 0.04 ± 0.03 ppm. That the structural changes caused by the double mutation are modest is further confirmed by an almost unchanged affinity of Wzc$_{\mathrm{CD\Delta C,\Delta Cys}}$ towards Wzb (Table 1).

Given the substantial number of Wzb resonances that are broadened in the presence of equimolar amounts of Wzc$_{\mathrm{CD\Delta C}}$, sub-stoichiometric concentrations of spin-labeled Wzc$_{\mathrm{CD\Delta C}}$ variants would be necessary for precise PRE measurements requiring thoughtful placement of the spin-labels. The spin-labels would have to be proximal to the α2-helix harboring the important ^{508}Glu–[Xxx]$_2$–Arg–[Xxx]$_2$–Arg^{514} motif for maximal effect while minimally perturbing native interactions. Additionally, they would have to be optimally distributed in space to provide sufficient non-redundant distance information for use in structure calculations. Five single cysteine variants (Val466Cys, Leu505Cys, Ser516Cys, Gln523Cys and Lys556Cys), distributed around the ^{508}Glu–[Xxx]$_2$–Arg–[Xxx]$_2$–Arg^{514} motif, were introduced into the Wzc$_{\mathrm{CD\Delta C,\Delta Cys}}$ background (*SI Appendix* Fig. S5). All variants express well and are properly folded as indicated by the corresponding ^{15}N,^{1}H TROSY spectra (*SI Appendix* Figs. S6B-F). The Wzb affinities of all variants, with the exception of Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Leu505Cys}}$ (see Table 1), remain largely unchanged in comparison to Wzc$_{\mathrm{CD\Delta C}}$. The ^{13}C,^{1}H HMQC spectra of Wzb display fast exchange in the presence of increasing amounts of Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Leu505Cys}}$ contrasting the slow exchange regime seen for Wzc$_{\mathrm{CD\Delta C}}$ and all other variants (*SI Appendix* Fig. S7). Given the ~5-fold reduction in affinity, and the likely related modification in its association dynamics, the Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Leu505Cys}}$ mutant was excluded from further analysis. The patterns of chemical shift perturbations of methyl resonances (*SI Appendix* Fig. S8) of Wzb in the presence of the $\text{Wzc}_{\text{CD}\Delta\text{C}}$ variants suggest no significant modification in the overall modes of interaction compared to the wild-type species.

To generate spin-labeled single cysteine variants of $\text{Wzc}_{\text{CD}\Delta\text{C}}$, we utilized a 3-(2-idoacetamido)-proxyl (IPSL) spin-label that covalently attaches to cysteine via a thioester bond. The efficiency of the spin labeling of ^{2}H-labeled $\text{Wzc}_{\text{CD}\Delta\text{C}}$ variants (for use in the NMR experiments) was confirmed by mass spectrometry; a successful labeling reaction is indicated by a mass increase of ~198 Da for the IPSL-modified species (see *SI Appendix* Fig. S9 for a representative example). Four IPSL-labeled variants (Val466Cys, Ser516Cys, Gln523Cys and Lys556Cys) were subsequently utilized to measure PREs induced on ^{2}H,^{15}N-labeled Wzb. The appropriate attachment of the spin-labels was confirmed by the pattern of signal attenuations seen in ^{15}N,^{1}H TROSY spectra of the spin-labeled $\text{Wzc}_{\text{CD}\Delta\text{C}}$ variants (*SI Appendix* Fig. S10). Analyses of the patterns of perturbations of the amide resonances of Wzb in the presence of the reduced forms of the IPSL-labeled $\text{Wzc}_{\text{CD}\Delta\text{C}}$ mutants confirm that this covalent modification does not substantially alter their modes of interaction with Wzb (*SI Appendix* Fig. S11).

Ratios of intensities of Wzb resonances in the presence of each of the four spin-labeled $\text{Wzc}_{\text{CD}\Delta\text{C}}$ variants in their paramagnetic (I_{para}) and diamagnetic (I_{dia}) states (*SI Appendix* Fig. S12; refer to *SI Appendix* Materials and Methods for experimental details) show distinct patterns suggesting that the measured PREs encode significant non-redundant distance information. Residues with statistically significant reduction in intensity ($I_{\text{para}}/I_{\text{dia}} < 0.7$) mapped onto the surface of Wzb (Fig. 2) are localized within well-defined patches implying that the transiently assembled complex is restricted to a single overall orientation, or a set of very similar orientations.

A set of 59 unambiguous distance restraints (listed in *SI Appendix* Table S3 and illustrated in *SI Appendix* Fig. S13) arising from the four spin-labeled sites on $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$: Val466Cys, Ser516Cys, Gln523Cys, and Lys556Cys, were supplemented with the AIRs (*SI Appendix* Table S2) obtained from the relaxation dispersion measurements on Wzb (described above), and our previously measured spectral perturbations on $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ (*SI Appendix* Fig. S14) (24) and utilized in docking calculations using the HADDOCK suite (27, 28) to obtain the structure of the Wzb•$\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ complex. The importance of the PRE-based unambiguous distance restraints is highlighted by the fact that their inclusion is crucial in obtaining a single dominant structural cluster ($\mathrm{Cluster}_{\mathrm{ALL,1}}$ and $\mathrm{Cluster}_{\mathrm{PRE,1}}$ with ~53% and ~59%, respectively, of the 400 refined structures; compare *SI Appendix* Tables S4, S5 with *SI Appendix* Table S6). The overall characteristics of the protein interfaces calculated with or without AIRs are quite similar (*SI Appendix* Fig. S15) but with the former set of calculations leading to more favorable HADDOCK scores. Therefore, unless otherwise stated, we focus exclusively on the structures that constitute $\mathrm{Cluster}_{\mathrm{ALL,1}}$ in the discussion below.

As previously noted, the α2 helix of $\mathrm{Wz}_{\mathrm{CD\Delta C}}$, and the invariant Glu508/Arg511/Arg514 triad therein, has been predicted to play a crucial role in the recognition of Wzb (24). A statistical analysis of the 212 structures that constitute $\mathrm{Cluster}_{\mathrm{ALL,1}}$ suggests that $\mathrm{Wz}_{\mathrm{CD\Delta C}}$ residues that comprise the interface (defined as those that lie within 4 Å of any Wzb residue; *SI Appendix* Fig. S16) includes all three residues of the triad, together with the intervening Ser512 in all constituent structures. His518, Phe519 and Met522, all on α2, and Asn555 and Arg557 on the α3-β2 loop, are found at the interface in >80% models. The complementary interface on Wzb involves residues (found in >80% of the structures*, SI Appendix* Fig. S16) of the N-terminus (F2), the α3’-β3’ loop (D77), the α4’-β4’ loop (P95, R98, G99, K100), the β4’ strand (M102), the D-loop (β4’-α5’loop;

H106, W107), and the C-terminus (E145, Q146, V147). These same residues are also part of the protein-protein interface in the other major clusters, $\text{Clusters}_{\text{ALL},2}$, $\text{Clusters}_{\text{ALL},3}$ and $\text{Clusters}_{\text{ALL},4}$ (*SI Appendix* Fig. S16).

The interaction results in the burial of relatively modest surface areas of 757Å^2 and 747.2 Å^2 for $\text{Wz}_{\text{CD}\Delta\text{C}}$ and Wzb, respectively, a likely explanation for their relatively weak interaction. Despite the presence of a significant number of polar residues at the interface, only 7 intermolecular hydrogen-bonds/salt-bridges are consistently detected, of which only two involve Wzb sidechains, D77 and R98 that forms salt-bridges with Asn555 and Glu508, respectively. The position corresponding to D77 is invariant in LMW-PTPs. The R98 position, though somewhat less conserved, is generally a basic (histidine or lysine) residue when substituted (*SI Appendix* Fig. S17). Hydrogen-bonds of the G99 and V147 mainchains with Arg514 and Asp637 sidechains, respectively, provide additional stability at the interface. These interactions are supplemented by those involving hydrophobic clusters on the two partners comprising H106 and W107 on Wzb and His518, Phe519 and Met522 on $\text{Wz}_{\text{CD}\Delta\text{C}}$. In LMW-PTPs, the position corresponding to H106 is somewhat less conserved though it is a histidine in large number of cases; the W107 position is largely conserved or contains an aromatic or a large hydrophobic residue in case of deviations (*SI Appendix* Fig. S17). Key interfacial interactions are illustrated in Fig. 3.

To test whether the structure of the Wzb•$\text{Wzc}_{\text{CD}\Delta\text{C}}$ complex (missing the Y-cluster) is compatible with Y-cluster dephosphorylation, we utilized the MODELLER suite (29) to attach the C-terminal tail fragment (Tyr705-Lys720) onto the docked structures (see *SI Appendix Materials* and Methods for details). Five of the six Y-cluster tyrosine residues (Tyr708, Tyr710, Tyr711, Tyr713 and Tyr715) can access the Wzb active-site in optimal fashion within the complex. The distances of the -OH moieties of these residues from the thiol group of C9 and the guanidino group of R15, both part of the catalytic C-[X]$_5$-R-[S/T] motif (30), are within (or very close to) their optimal distances of 3.3 and 2.6 Å (31, 32), respectively (*SI Appendix* Fig. S18). In contrast these distances deviate significantly for Tyr705, suggesting that the dephosphorylation of this residue, if it occurs at all, is expected to be highly inefficient. It is notable, Tyr705 is the only position on the Y-cluster that has been shown not be auto-phosphorylated in Wzc (17).

We have previously demonstrated that a Glu508Ala/Arg511Ala/Arg514Ala mutant, where the conserved residues of the ^{508}Glu–[Xxx]$_2$–Arg–[Xxx]$_2$–Arg^{514} motif of Wzc$_{CD}$ have been replaced by alanine, is severely compromised in its ability to interact with and be dephosphorylated by Wzb (24). Those results support the involvement of these residues in recognizing Wzb in line with our current structure. To test the complementary surface on Wzb, we mutated D77 and R98 (discussed above) individually to alanine. While the Wzb$_{R98A}$ mutant expresses well and can be purified to homogeneity, the expression of the Wzb$_{D77A}$ is sub-optimal suggesting that this mutation likely destabilizes the protein. We attributed this decreased stability to the loss of a salt-bridge (Fig. 3) between the sidechains of D77 and the proximal K100, that is also highly conserved (*SI Appendix* Fig. S17), in the mutant. To restore this salt-bridge we generated a charge reversal variant, Wzb$_{D77K/K100D}$. This variant is also expected to be compromised in its ability to interact with and dephosphorylate Wzc$_{CD}$. The Wzb$_{R98A}$ and Wzb$_{D77K/K100D}$ variants were compared with wild-type Wzb (WT) for their ability to dephosphorylate fully phosphorylated Wzc$_{CD}$. While Wzb$_{R98A}$ can fully dephosphorylate the Wzc$_{CD}$, it does so less somewhat less efficiently than the wild-type enzyme (WT:13.6 ± 2.3 s^{-1}, R98A: 9.1 ± 1.5 s^{-1}). Y-cluster dephosphorylation by Wzb$_{D77K/K100D}$ is severely compromised (1.8 ± 0.4 s^{-1}) and does not progress to completion (Fig. 4).

DISCUSSION

The structure Wzb•$\text{Wzc}_{\text{CD}\Delta\text{C}}$ complex determined here not only reveals key structural features that enable molecular recognition between these two species, but also provides a mechanistic framework to understand how the assembly/disassembly cycle correlates to Y-cluster phosphorylation states (8). As shown above, the $\alpha2$ helix of Wzc_{CD} plays a key role in the recognition of Wzb. Indeed, interactions involving this helix, both polar and hydrophobic, appear to be conserved within BY-kinase/LMW-PTP pairs by comparing predicted structural interfaces between several related BY-kinase/LMW-PTP pairs from Gram-negative species (see *SI Appendix* Fig S19). As has been mentioned before, the $\alpha2$ helix is also central to the formation of the oligomeric assembly (12) necessary to enable the active site of an enzyme-acting protomer to receive and phosphorylate the C-terminal Y-cluster of an adjoining substrate-acting subunit. Therefore, a competition for the $\alpha2$ helix site between the homotypic and heterotypic interactions involving Wzc_{CD} monomers and Wzb, respectively, is expected. Thus, the processes of phosphorylation and dephosphorylation would need to be temporally separated for maximal efficiency. We suggest that this is achieved through the intricate mechanism illustrated in Fig. 5. As mentioned before, Wzc_{CD} is the cytosolic domain of a protein that is embedded in the bacterial inner membrane (8) (*SI Appendix* Fig. S1A). Indeed, all BY-kinases need to form a membrane-associated species for activity (5). This arrangement results in the formation of an octamer through interactions between Wzc_{CD} monomers (8), supplemented by extra-cytoplasmic interactions, in the absence of Y-cluster phosphorylation. This membrane-bound assemblage greatly enhances the local concentration of Wzc_{CD} monomers and shields the $\alpha2$-site from diffusion-controlled access by the wholly cytosolic Wzb. This allows phosphorylation to proceed (likely in sequence from the C-terminus to the N-terminus (8) of the Y-cluster) without the counteracting effects of premature

Wzb-driven dephosphorylation. Once the Y-cluster is sufficiently phosphorylated, the homotypic affinity between $\mathrm{Wzc_{CD}}$ monomers is reduced due to strong electrostatic repulsion resulting from the significant accumulation of negative charge on the Y-cluster. When this repulsion exceeds a certain critical threshold, predicted to be at about 4 phosphorylated tyrosine residues based on functional data (8), the $\mathrm{Wzc_{CD}}$ monomers are pushed apart allowing Wzb to successfully outcompete the intra-$\mathrm{Wzc_{CD}}$ interactions and dock at the $\alpha2$-site to initiate dephosphorylation. After the dephosphorylation of a sufficient number of Y-cluster tyrosines, the repulsion between $\mathrm{Wzc_{CD}}$ monomers decreases below the critical threshold allowing their reassociation and the successful exclusion of Wzb, thereby reinitiating the cycle. It has also been suggested that oligomerization of $\mathrm{Wzc_{CD}}$ monomers is necessary to enable the exchange of product ADP for substrate ATP (33). Given that $\mathrm{Wzc_{CD}}$ possesses significant ATPase activity (34), oligomerization would have to precede ATP binding to prevent futile hydrolysis. Specific features within the $\alpha2$ helix have also been suggested to be coupled to that process (33). Thus, this helix appears to be a central hub in coordinating multiple temporally regulated processes, and represents, what appears to be, part of an intricate timing mechanism that likely serves to couple intra- and extra-cytoplasmic events necessary for polysaccharide synthesis and export. It is notable that the region of Wzc that we suggest plays a key role in driving the proposed temporal regulation shows broad conservation in orthologs across the bacterial kingdom (*SI Appendix* Fig. S20). While further studies are needed to test this “molecular stopwatch” hypothesis, and the role of BY-kinases therein, clues of its existence may be found several Gram-positive bacteria e.g., *Staphylococcus aureus* (35) and *Streptococcus pneumoniae* (36, 37) where critical cellular processes appear to coordinated through BY-kinase activity. Though direct experimental evidence of similar effects in the Gram-negatives is currently lacking, it has been shown that perturbations in the native Wzc in

E. coli K30 (Wzc_{K30}) uncouples the dual process of the synthesis of capsular polysaccharides and their export to the cell surface (38).

In addition to generating testable hypotheses about BY-kinase regulation, the current study also illustrates key structural features of molecular recognition involving an LMW-PTP for the very first time. Whether the engagement mode seen here is general for interactions involving LMW-PTPs is unclear. Based on the conservation of the LMW-PTP fold from bacteria to humans (39) and resulting diversity of interactions, one would expect this not to be the case. Indeed, given the absence of additional regulatory domains that decorate their high molecular weight counterparts (40), LMW-PTPs have likely evolved varied interaction modes based on their particular cellular context. For example, MptpA, one of the two LMW-PTPs encoded by *Mycobacterium tuberculosis*, that lacks BY-kinases, appears to utilize a different interaction mode to engage the atypical tyrosine kinase, PtkA (41). Nevertheless, for cognate BY-kinase/LMW-PTPs pairs, we expect that the interactions involving the α2 helix of the kinase and the corresponding elements of the phosphatase are conserved and represent a general mode of molecular recognition. This common mode may be supplemented by additional interactions in different BY-kinase/LMW-PTP pairs. For example, the interaction between D77 and Asn555, deemed important in the present case appears not to be conserved in the other related pairs illustrated in *SI Appendix* Fig. S19. While the phosphatase residue (D77 in Wzb) is fully conserved, the corresponding residue on kinase (Asn555 in Wzc) is not conserved for the cases shown (also see *SI Appendix* Fig. S20). That it is possible to compensate for this supplementary interaction is suggested by the fact Wzb can complement Wzb_{K30} within the group 1 capsule production system of *E. coli* K30 (38). Indeed, reinforcement of a primary mode of recognition by additional interactions would not be unusual for PTPs. For example, the MAP kinase regulatory HePTP recognizes the related p38 and ERK kinases through a kinase interaction motif (KIM) but specificity towards p38 result from additional interactions involving a kinase specificity sequence (KIS) (42). However, more extensive pairwise structure-based analyses on *bona fide* functional BY-kinase/LMW-PTP partners across the bacterial kingdom are needed to parse general recognition modes from system-specific nuances. These bioinformatic/computational analyses would have to be validated through additional experimental structures most importantly for a representative pair from Gram-positive species to complement the present work.

The LMW-PTPs represent an ancient, but poorly studied, class of protein phosphatases (39), and molecular interactions involving them are suitable targets for the design of pharmaceuticals ranging from antimicrobials (43) to antidiabetics (44). Beyond insight into BY-kinase regulation, we expect that our current studies will provide a general approach to studying molecular recognition in a broader class of LMW-PTPs.

MATERIALS AND METHODS

Protein expression and purification

All constructs (and variants) of the catalytic domain of *E. coli* (K-12) Wzc (447-725, Wzc$_{\mathrm{CD}}$), its truncated form (lacking the C-terminal tail, 447-704, Wzc$_{\mathrm{CD\Delta C}}$), and full-length *E. coli* (K-12) Wzb were expressed and purified using protocols described at length in the SI Appendix Materials and Methods. A list of constructs used in this study are listed in *SI Appendix* Table S1.

NMR experiments

All NMR experiments were recorded at 25 °C using Bruker Avance III spectrometers operating at 600, 700 or 800 MHz, all equipped with cryogenic probes capable of applying pulsed-field gradients along the z-axis. Additional details of specific experiments and methodology may be found in the *SI Appendix* Materials and Methods.

Computational protocols

Structure calculations were performed using ambiguous interaction restraints (*SI Appendix* Table S2) from the measurement of transient chemical shifts (Wzb) and static spectral perturbations (Wzc$_{\mathrm{CD\Delta C}}$) in combination with PRE-based unambiguous restraints (*SI Appendix* Table S3) using the HADDOCK suite (27). Procedures used for the preparation of starting structures for Wzc$_{\mathrm{CD}}$ and Wzb, generation of experimental restraints, and computational protocols used are described in detail in the *SI Appendix* Materials and Methods.

Protein dephosphorylation assays

The ability of phosphorylated Wzc$_{\mathrm{CD}}$ to be dephosphorylated by wild-type Wzb and corresponding D77K/K100D and R98A mutants were tested by immunoblotting using the monoclonal anti-phosphotyrosine antibody, PY20 (dilution 1:5000, Invitrogen). Experimental details are provided in the *SI Appendix* Materials and Methods.

ACKNOWLEDGEMENTS

This work was supported by NSF grant MCB1937937. NMR measurements were carried out in the CUNY ASRC Biomolecular NMR facility and at the New York Structural Biology Center (NYSBC). The NMR facilities at the NYSBC are supported by NIH grants OD016432 and OD018509. The authors thank Dr. Kaushik Dutta (NYSBC) for help at the initial stages of this project, Dr. Christophe Grangeasse (MMSB, Lyon) and Mr. Fatlum Hajredini (CCNY) for useful discussions, Dr. James Aramini (CUNY-ASRC) for implementing some of the NMR experiments utilized in this study, and Dr. Brian Kloss (NYSBC) for the kind gift of the anti-phosphotyrosine antibody.

Author contributions

R. G. conceived the project; S.A. and I. D. performed all NMR experiments assisted by A. P. S. A. performed all computational calculations and functional studies. S. A. prepared a first draft of the paper and figures that were refined by R. G. with input from all authors.

Competing interests

The authors declare that they have no competing interests.

REFERENCES

1. J. D. Chao, D. Wong, Y. Av-Gay, Microbial protein-tyrosine kinases. *J. Biol. Chem.* **289**, 9463-9472 (2014).
2. C. Grangeasse, S. Nessler, I. Mijakovic, Bacterial tyrosine kinases: evolution, biological function and structural insights. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2640-2655 (2012).
3. L. Shi, A. Kobir, C. Jers, I. Mijakovic, Bacterial protein-tyrosine kinases. *Curr. Proteom.* **7**, 188-194 (2010).
4. D. C. Lee, Z. Jia, Emerging structural insights into bacterial tyrosine kinases. *Trends Biochem. Sci.* **34**, 351-357 (2009).
5. C. Grangeasse, A. J. Cozzone, J. Deutscher, I. Mijakovic, Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem. Sci.* **32**, 86-94 (2007).
6. C. Schwechheimer *et al.*, A tyrosine phosphoregulatory system controls exopolysaccharide biosynthesis and biofilm formation in *Vibrio cholerae*. *PLoS Pathog.* **16**, e1008745 (2020).
7. C. M. Ernst *et al.*, Adaptive evolution of virulence and persistence in carbapenem-resistant *Klebsiella pneumoniae*. *Nat. Med.* **26**, 705-711 (2020).
8. Y. Yang *et al.*, The molecular basis of regulation of bacterial capsule assembly by Wzc. *Nat. Commun.* **12**, 4349 (2021).
9. C. Dong *et al.*, Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature* **444**, 226-229 (2006).

10. R. F. Collins *et al.*, The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **104**, 2390-2395 (2007).

11. D. C. Lee, J. Zheng, Y. M. She, Z. Jia, Structure of *Escherichia coli* tyrosine kinase Etk reveals a novel activation mechanism. *EMBO J.* **27**, 1758-1766 (2008).

12. E. Bechet *et al.*, Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. *Mol. Microbiol.* **77**, 1315-1325 (2010).

13. V. Olivares-Illana *et al.*, Structural basis for the regulation mechanism of the tyrosine kinase CapB from *Staphylococcus aureus*. *PLoS Biol.* **6**, e143 (2008).

14. S. K. Hanks, T. Hunter, Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576-596 (1995).

15. S. W. Cowan-Jacob, Structural biology of protein tyrosine kinases. *Cell. Mol. Life Sci.* **63**, 2608-2625 (2006).

16. F. Jadeau *et al.*, Identification of the idiosyncratic bacterial protein tyrosine kinase (BY-kinase) family signature. *Bioinformatics* **24**, 2427-2430 (2008).

17. C. Grangeasse, P. Doublet, A. J. Cozzone, Tyrosine phosphorylation of protein kinase Wzc from *Escherichia coli* K12 occurs through a two-step process. *J. Biol. Chem.* **277**, 7127-7135 (2002).

18. A. Paiment, J. Hocking, C. Whitfield, Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzc, on its activity in assembly of group 1 capsules in *Escherichia coli. J. Bacteriol.* **184**, 6437-6447 (2002).

19. C. Grangeasse *et al.*, Functional characterization of the low-molecular-mass phosphotyrosine-protein phosphatase of *Acinetobacter johnsonii*. *J. Mol. Biol.* **278**, 339-347 (1998).

20. C. Vincent *et al.*, Cells of *Escherichia coli* contain a protein-tyrosine kinase, Wzc, and a phosphotyrosine-protein phosphatase, Wzb. *J. Bacteriol.* **181**, 3472-3477 (1999).

21. T. J. Boggon, M. J. Eck, Structure and regulation of Src family kinases. *Oncogene* **23**, 7918-7927 (2004).

22. B. Obadia *et al.*, Influence of tyrosine-kinase Wzc activity on colanic acid production in *Escherichia coli* K12 cells. *J. Mol. Biol.* **367**, 42-53 (2007).

23. E. Bechet *et al.*, Tyrosine-kinases in bacteria: from a matter of controversy to the status of key regulatory enzymes. *Amino Acids* **37**, 499-507 (2009).

24. D. B. Temel *et al.*, Regulatory interactions between a bacterial tyrosine kinase and its cognate phosphatase. *J. Biol. Chem.* **288**, 15212-15228 (2013).

25. C. A. Waudby, A. Ramos, L. D. Cabrita, J. Christodoulou, Two-dimensional NMR lineshape analysis. *Sci. Rep.* **6**, 24826 (2016).

26. J. P. Loria, M. Rance, A. G. Palmer, A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331-2332 (1999).

27. C. Dominguez, R. Boelens, A. M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737 (2003).

28. E. Karaca, A. S. Melquiond, S. J. de Vries, P. L. Kastritis, A. M. Bonvin, Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Mol. Cell. Proteom.* **9**, 1784-1794 (2010).

29. N. Eswar, D. Eramian, B. Webb, M. Y. Shen, A. Sali, Protein structure modeling with MODELLER. *Meth. Mol. Biol.* **426**, 145-159 (2008).

30. D. Barford, Z. Jia, N. K. Tonks, Protein tyrosine phosphatases take off. *Nat. Struct. Biol.* **2**, 1043-1053 (1995).

31. X. Wang, Q. Ma, Wzb of Vibrio vulnificus represents a new group of low-molecular-weight protein tyrosine phosphatases with a unique insertion in the W-loop. *J. Biol. Chem.* **296**, 100280 (2021).

32. G. Hagelueken, H. Huang, I. L. Mainprize, C. Whitfield, J. H. Naismith, Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly. *J. Mol. Biol.* **392**, 678-688 (2009).

33. F. Hajredini, A. Piserchio, R. Ghose, Long-range dynamic correlations regulate the catalytic activity of the bacterial tyrosine kinase Wzc. *Sci. Adv* **6**, eabd3718 (2020).

34. D. Soulat *et al.*, Tyrosine-kinase Wzc from *Escherichia coli* possesses an ATPase activity regulated by autophosphorylation. *FEMS Microbiol. Lett.* **274**, 252-259 (2007).

35. M. Rausch *et al.*, Coordination of capsule assembly and cell wall biosynthesis in *Staphylococcus aureus. Nat. Commun.* **10**, 1404 (2019).

36. R. Nakamoto *et al.*, The bacterial tyrosine kinase system CpsBCD governs the length of capsule polymers. *Proc. Natl. Acad. Sci. USA* **118**, e2103377118 (2021).

37. J. Nourikyan *et al.*, Autophosphorylation of the bacterial tyrosine-kinase CpsD connects capsule synthesis with the cell cycle in *Streptococcus pneumoniae*. *PLoS Genet.* **11**, e1005518 (2015).

38. A. N. Reid, C. Whitfield, functional analysis of conserved gene products involved in assembly of *Escherichia coli* capsules and exopolysaccharides: evidence for molecular recognition between Wza and Wzc for colanic acid biosynthesis. *J. Bacteriol.* **187**, 5470-5481 (2005).

39. A. Caselli *et al.*, Low molecular weight protein tyrosine phosphatase: Multifaceted functions of an evolutionarily conserved enzyme. *Biochim. Biophys. Acta* **1864**, 1339-1355 (2016).

40. N. K. Tonks, Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell. Biol.* **7**, 833-846 (2006).

41. T. Stehle *et al.*, The apo-structure of the low molecular weight protein-tyrosine phosphatase A (MptpA) from *Mycobacterium tuberculosis* allows for better target-specific drug development. *J. Biol. Chem.* **287**, 34569-34582 (2012).

42. A. Piserchio *et al.*, Docking interactions of hematopoietic tyrosine phosphatase with MAP kinases ERK2 and p38α. *Biochemistry* **51**, 8047-8049 (2012).

43. L. Fanzani *et al.*, *Mycobacterium tuberculosis* low molecular weight phosphatases (MPtpA and MPtpB): from biological insight to inhibitors. *Curr. Med. Chem.* **22**, 3110-3132 (2015).

44. S. M. Stanford *et al.*, Diabetes reversal by inhibition of the low-molecular-weight tyrosine phosphatase. *Nat. Chem. Biol.* **13**, 624-632 (2017).

Table 1. Apparent binding affinities of various Wzc$_{\mathrm{C\Delta DC}}$ constructs for wild-type Wzb*

Construct	***K*$_{\mathrm{D}}$ (μM)**
Wzc$_{\mathrm{CD\Delta C}}$**	8.6 ± 0.6
Wzc$_{\mathrm{CD\Delta C,\Delta Cys}}$	15.2 ± 0.6
Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Val466Cys}}$	8.9 ± 0.3
Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Leu505Cys}}$	43.8 ± 1.3
Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Ser516Cys}}$	11.3 ± 0.7
Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Gln523Cys}}$	16.9 ± 0.5
Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Lys556Cys}}$	7.3 ± 0.4

*Determined through 2-dimensional line-shape analyses of ^{13}C, ^{1}H HMQC spectra assuming that the fast-relaxing component can be neglected.

**Determined by ITC: 10.7 ± 1.3 μM or SPR: 5.1 ± 0.2 μM

Figure Legends

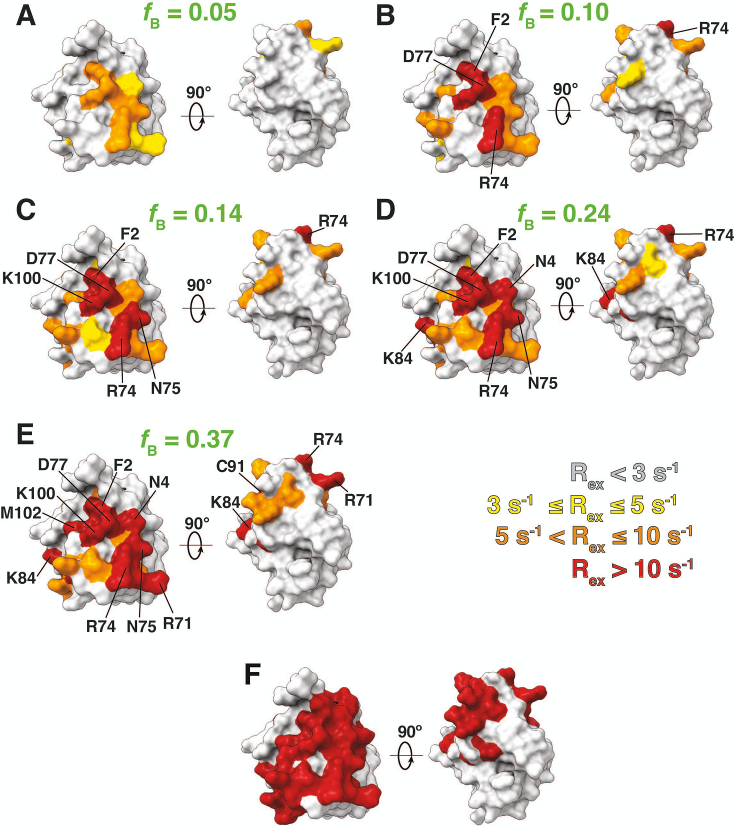
Fig. 1 R_{ex} values mapped onto the Wzb surface for various fractions of its $Wzc_{CD\Delta C}$-bound state. (**A-E**) $R_{ex} \leq 3\ s^{-1}$, grey; $3\ s^{-1} < R_{ex} \leq 5\ s^{-1}$, yellow; $5\ s^{-1} < R_{ex} \leq 10\ s^{-1}$, orange; $R_{ex} > 10\ s^{-1}$, red. (**F**) Wzb residues that show significant signal attenuation (> 90%) in the presence of equimolar amounts of $Wzc_{CD\Delta C}$ are mapped onto the Wzb surface.

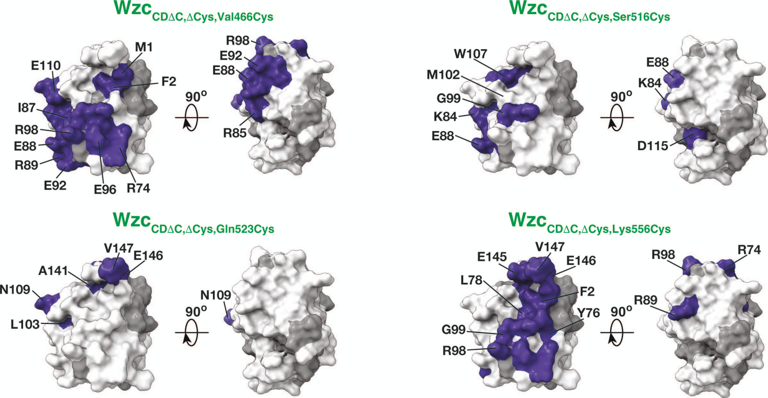
Fig. 2 PREs induced on Wzb in the presence of IPSL-labeled $Wzc_{CD\Delta C}$ variants. The ratios $I_{para}/I_{dia} < 0.7$ are mapped onto the Wzb surface (in blue) for the Val466Cys, Ser516Cys, Gln523Cys and Lys556Cys variants generated on the $Wzc_{CD\Delta C,\Delta Cys}$ background. Residues for which data could not be analyzed due to missing resonance assignments or spectral overlap are colored grey. Intensity ratios correspond to those measured for $^{2}H,^{15}N$-labeled Wzb in the presence of IPSL,^{2}H-labeled $Wzc_{CD\Delta C}$ variants in the following molar equivalents: $Wzc_{CD\Delta C,\Delta Cys,Val466Cys}$ (0.5), $Wzc_{CD\Delta C,\Delta Cys,Gln523Cys}$ (0.5), $Wzc_{CD\Delta C,\Delta Cys,Ser516Cys}$ (0.36), and $Wzc_{CD\Delta C,\Delta Cys,Lys556Cys}$ (0.34). Key perturbed residues are labeled.

Fig. 3 Key interaction at the protein-protein interface in the Wzb•$Wzc_{CD\Delta C}$ complex. The structure of the lowest energy model (from $Cluster_{ALL,1}$) is shown in surface (top) and ribbon (middle) representation. Interacting elements on each partner that comprise the interface are indicated. The bottom panel shows expansions of specific regions (indicated by the colored rectangles) illustrating key interactions that stabilize the interface. The cyan dashed lines indicate intermolecular hydrogen bonds; the green dashed line indicates the intramolecular D77-K100 salt-bridge discussed in the text.

Fig. 4 Dephosphorylation of Wzc_{CD} by Wzb and variants. Reaction mixtures containing fully phosphorylated Wzc_{CD} were incubated for various time intervals (0, 1, 3, 5 and 10 mins) with either wild-type Wzb or forms carrying double (D77K/K100D) or single (R98A) mutations at key sites at the protein-protein interface. Representative immunoblots for the time-courses are shown in **(A)** and the corresponding fits to extract apparent dephosphorylation rates (normalized to 100%) are shown in **(B)**. Also shown are the positions of the molecular weight markers (in kDa).

Fig. 5 Schematic illustration of the phosphorylation-coupled assembly/disassembly that drives BY-kinase function. At low levels of Y-cluster phosphorylation (Y-low assembled; represented as fully dephosphorylated for illustrative purposes), the cytoplasmic catalytic domain ($\mathrm{Wzc_{CD}}$) of Wzc (green) forms an octameric ring. This assembly allows each $\mathrm{Wzc_{CD}}$ (I_0) subunit to simultaneously act as an enzyme, accommodating the Y-cluster of the preceding subunit (I_{-1}) at its active site, and as a substrate, inserting its own Y-cluster into the active-site of the following subunit (I_{+1}). Oligomerization enables exchange of ATP•Mg^{2+} (dark red) for the bound ADP (purple), facilitating the progression to chemistry. At this stage, the diffusive binding of cytosolic Wzb (orange), is rendered inefficient due to the enhanced local concentration of $\mathrm{Wzc_{CD}}$ monomers in the context of the membrane-assembled species. Following several rounds of chemistry (phosphorylated tyrosine residues are indicated by blue spheres), and enhanced Y-cluster phosphorylation (Y-high assembled), the electrostatic repulsion (red arrows) increases above a critical threshold, making homotypic interactions between $\mathrm{Wzc_{CD}}$ monomers disfavored over heterotypic interactions with Wzb, that can engage the $\mathrm{Wzc_{CD}}$ monomers of the now disassembled ring (Y-high disassembled). After several rounds of Wzb-mediated dephosphorylation, the electrostatic repulsion is once again lowered below the critical threshold (Y-low disassembled) allowing the successful exclusion of Wzb, and regeneration of the oligomeric state of $\mathrm{Wzc_{CD}}$ (Y-low assembled), thus reinitiating the cycle.


A
fB = 0.05
90°
B
fB = 0.10
F2
D77
R74
90°
R74
C
fB = 0.14
F2
D77
K100
R74
N75
90°
R74
D
fB = 0.24
F2
D77
K100
N4
K84
R74
N75
90°
K84
R74
E
fB = 0.37
D77
F2
K100
M102
N4
K84
R74
N75
R71
90°
C91
K84
R74
R71
Rex < 3 s-1
3 s-1 ≤ Rex ≤ 5 s-1
5 s-1 < Rex ≤ 10 s-1
Rex > 10 s-1
F
90°

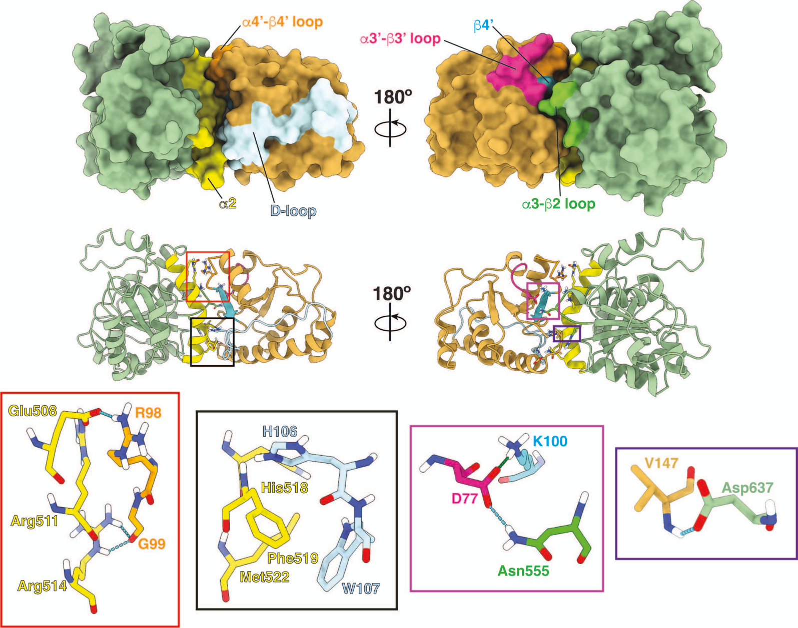


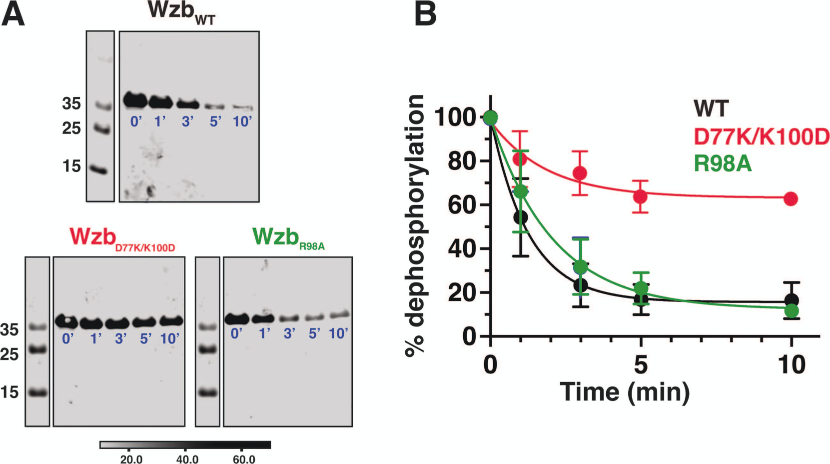
Wzc$_{CD\Delta C,\Delta Cys,Val466Cys}$

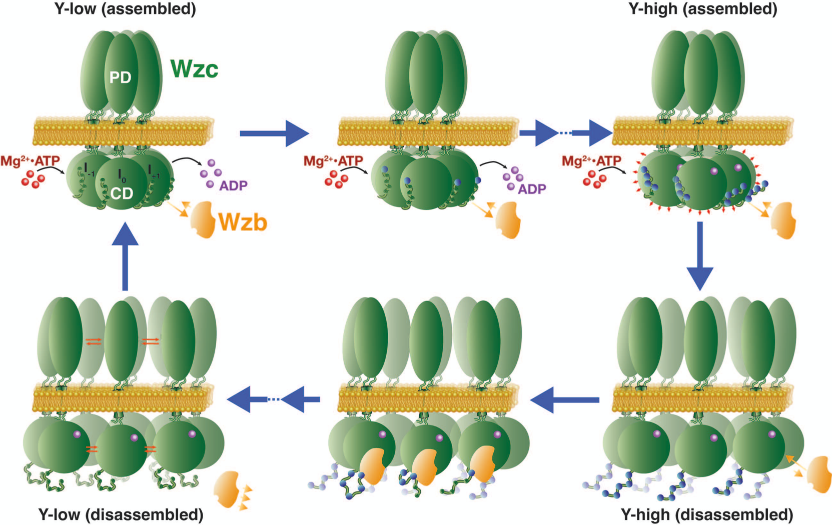
Wzc$_{CD\Delta C,\Delta Cys,Ser516Cys}$

Wzc$_{CD\Delta C,\Delta Cys,Gln523Cys}$

Wzc$_{CD\Delta C,\Delta Cys,Lys556Cys}$


α4'-β4' loop
α2
D-loop
180°
α3'-β3' loop
β4'
α3-β2 loop
180°
Glu508
R98
Arg511
G99
Arg514
H106
His518
Phe519
Met522
W107
K100
D77
Asn555
V147
Asp637


A
Wzb_{WT}
35
25
15
0'
1'
3'
5'
10'
$Wzb_{D77K/K100D}$
Wzb_{R98A}
20.0
40.0
60.0
B
WT
D77K/K100D
R98A
% dephosphorylation
100
80
60
40
20
0
Time (min)
5
10


Y-low (assembled)
PD
Wzc
Mg2+•ATP
I-1
I0
I+1
CD
ADP
Wzb
Mg2+•ATP
ADP
Y-high (assembled)
Mg2+•ATP
Y-high (disassembled)
Y-low (disassembled)

SI APPENDIX

for

Structural Basis for the Recognition of the Bacterial Tyrosine Kinase, Wzc by its Cognate Tyrosine Phosphatase, Wzb

MATERIALS AND METHODS

Protein expression and purification

A list of constructs used in this study and the corresponding primers are listed in Table S1. All mutations were carried out using QuikChange mutagenesis kits (Agilent) with either *wzb* or *wzc*$_{\mathrm{CD}}$ as templates and confirmed by sequencing. To express the various constructs, *E. coli* One shot BL21 Star (DE3) cells (Thermofisher) were transformed with plasmids harboring either the intact catalytic domain of *E. coli* (K-12) Wzc (447-725, Wzc$_{\mathrm{CD}}$), or its truncated form (lacking the C-terminal tail, 447-704, Wzc$_{\mathrm{CD\Delta C}}$) or variants thereof, and grown overnight on ampicillin agar plates at 37 °C.

To express unlabeled protein, freshly transformed colonies were grown overnight at 37 °C in 50 mL of LB supplemented with ampicillin. The overnight culture was used to inoculate 1L of ampicillin containing LB and incubated at 37 °C with shaking at 225 rpm; when the absorbance (A_{600}) reached 0.8, protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expression continued at 37 °C for 4 h. Cells were harvested by centrifugation at 4000 rpm for 30 mins using a (Fiberlite F12-6 x 500 LEX rotor (Thermofisher Scientific). Cell pellets were stored at -80 °C.

For the overexpression of proteins in D_2O-based media, freshly transformed colonies were initially grown overnight at 37 °C in 20 mL of H_2O-based M9 media supplemented with ampicillin followed by a gradual D_2O adaptation of the overnight culture by progressively increasing the D_2O content of the M9 (20 mL) medium (10%, 30%, 50%, 75%, 90%, and 100%). After reaching 100% D_2O, the cells were inoculated into 40 mL of fresh M9 medium (100% D_2O, ampicillin) for overnight growth. The overnight culture was transferred into 1L of M9 medium (100% D_2O, ampicillin) containing 1 g/L of NH_4Cl (or $^{15}NH_4Cl$ when^{15}N-labeling was required) and ^{12}C, ^{2}H-labeled glucose (Cambridge Isotope Laboratories). The cell culture was incubated at 37 °C with shaking at 225 rpm. When the absorbance (A_{600}) reached 0.7, protein expression was induced by adding 0.4 mM IPTG. For ILV-labeling, the cells were grown at 37 °C under gentle agitation at 250 rpm until the A_{600} reached 0.6, at which point 100 mg/L of methyl-^{13}C, 3,3-D_2 α-keto-butyric acid and 250 mg/L of 3-3'-dimethyl-$^{13}C_2$, 3-D α-keto-isovaleric acid (Cambridge Isotope Laboratories) were added to the medium and growth was continued at 37 °C for an additional hour prior to induction with 0.4 mM IPTG. The incubator temperature was reduced to 16 °C, and cells were grown for 20 to 24 h and harvested by centrifugation. Cell pellets were stored at -80 °C.

The cell pellets were thawed and resuspended in 40 mL of lysis buffer containing 50 mM phosphate (pH 8.0), 300 mM NaCl, 10% glycerol, 10 mM β-mercaptoethanol (BME), 10 mM imidazole, and one protease inhibitor tablet (Roche). All subsequent purification steps were performed at 4 °C. The cell suspension was lysed by sonication and centrifuged at 15000 rpm for 30 min using a Fiberlite F21-8 x 50y rotor (Thermofisher) to remove cell debris. The soluble fraction was incubated with nickel beads (Ni-NTA) (GoldBio), pre-equilibrated with lysis buffer for 1 h at 4 °C under gentle rotation, the slurry was then poured into a Econo-Pac column (Bio-Rad). The column flow-through was discarded and beads were extensively washed, first with lysis buffer, and then with thrice the column volume of wash buffer containing 50 mM phosphate (pH 8.0), 300 mM NaCl, 10% glycerol, 10 mM BME and 20 mM imidazole. The bound proteins were eluted with 40 mL of elution buffer comprising 50 mM phosphate (pH 8.0), 300 mM NaCl, 10% glycerol, 10 mM BME and 200 mM imidazole. Immediately after elution, Wzc$_{\mathrm{CD}}$ (or variants) samples were incubated overnight at 4 °C in the presence of 1 mM ATP and 1 mM $MgCl_2$ per 40 μM of purified protein. After elution, Wzb samples were stored at 4 °C. All proteins were further purified by gel filtration using either Hiload 16/60 Superdex 75 GL columns (GE Healthcare Biosciences) for all Wzb constructs, or Hiload 16/60 Superdex 200 PG columns for all Wzc$_{\mathrm{CD}}$ constructs. In both cases, the columns were pre-equilibrated with NMR buffer containing 50 mM phosphate (pH 6.2), 50 mM NaCl and 25 mM DTT. The samples used to record methyl ^{13}C, ^{1}H HMQC spectra were extensively buffer exchanged into the corresponding D_2O-based NMR buffer using spin columns and adjusted for pH changes.

Generation of spin-labeled variants of Wzc$_{\mathrm{CD\Delta C}}$

For the spin-labeling conjugation reactions, the mono-cysteine variants of Wzc$_{\mathrm{CD\Delta C}}$ (see Table S1) were incubated overnight in a pre-labeling buffer containing 50 mM sodium phosphate (pH 8.1), 100 mM NaCl and 25 mM DTT. The reducing agent, DTT, was then removed by passing the proteins through PD-10 desalting columns (GE Healthcare Biosciences) pre-equilibrated with spin-labeling buffer containing 50 mM sodium phosphate (pH 8.1) and 100 mM NaCl. The protein samples were then treated with 10-fold molar excess of 3-(2-iodoacetamido)proxyl (IPSL) dissolved in acetonitrile, and the reaction was allowed to proceed for 1 h at 25 °C in the dark. The unreacted spin-label was then removed by passing the protein samples through a PD-10 column pre-equilibrated in NMR buffer. The efficiency of the spin-labeling reactions was confirmed by MALDI-TOF mass spectrometry measurements. The reactions were found to proceed to completion in all cases.

NMR spectroscopy

All NMR experiments were recorded at 25 °C using Bruker Avance III spectrometers operating at 600, 700 or 800 MHz, all equipped with cryogenic probes capable of applying pulsed-field gradients along the z-axis. All samples were prepared in NMR buffer containing 5% D_2O (when using H_2O-based buffers) to enable field locking. Data were processed using NMRPipe (1) and analyzed using NMRViewJ (2).

Methyl chemical shift perturbations and changes in resonance intensity induced on Wzb by Wzc$_{\mathrm{CD\Delta C}}$ (and variants) were derived from ^{13}C, ^{1}H SOFAST-HMQC (3) experiments recorded at 700 MHz. Experiments used 512 and 128 complex points in the direct and indirect dimensions, respectively, with corresponding spectral windows of 12.5 ppm and 17.25 ppm. A full set of titrations were performed using 40 μM of ^{15}N, ILV-labeled Wzb with uniformly ^{2}H-labeled Wzc$_{\mathrm{CD\Delta C}}$ variants at the following Wzb:Wzc$_{\mathrm{CD\Delta C}}$ molar ratios: 1.0:0.2, 1.0:0.4, 1.0:0.6, 1.0:0.8, 1.0:1.0, 1.0:1.5, 1.0:2.0, 1.0:3.0, 1.0:4.0, 1.0:6.0, 1.0:8.0, and 1.0:10.0. The signal intensities in each titration point were measured using NMRViewJ (2) and subsequently corrected for sample dilution effects. In each case, the final titration point (1:10.0 molar ratio of the Wzc$_{\mathrm{CD\Delta C}}$ variant) was used to determine the chemical shift perturbation values (Δδ, in ppm) using

$$\Delta\delta = \sqrt{\left(\delta_{\mathrm{H,ref}} - \delta_{\mathrm{H}}\right)^2 + \left(\frac{\delta_{\mathrm{C,ref}} - \delta_{\mathrm{C}}}{\sigma_{\mathrm{C}}}\right)^2} \tag{S1}$$

where $\delta_{\mathrm{H,ref}}$ ($\delta_{\mathrm{C,ref}}$) and δ_{H} (δ_{C}) correspond to the ^{1}H (^{13}C) chemical shifts for the apo and fully-saturated (1:10.0 molar ratio) states, respectively,. The heteronuclear shifts were normalized by the standard deviation (σ_{C}) obtained from the Biological Magnetic Resonance Bank (BMRB) for the specific methyl position. Analysis of resonance line-shapes was carried out using TITAN software (4) to determine the apparent binding affinity (K_D) in each case and subsequently utilized to calculate the corresponding bound-state fraction (f_B).

Relaxation dispersion measurements

^{15}N relaxation dispersion measurements were carried out using the relaxation-compensated Carr-Purcell-Meiboom-Gill (CPMG) sequence in its standard implementation (5). Dispersion data were acquired at 600 MHz using samples containing 200 μM of ^{15}N, ^{2}H-labeled Wzb in the presence of 0 μM, 10 μM, 20 μM, 30 μM, 50 μM or 80 μM ^{2}H-labeled Wzc$_{CD\Delta C}$ (corresponding to molar ratios of 1:0, 1:0.05, 1:0.10, 1:0.15, 1:0.25 and 1:0.4) in NMR buffer. All datasets utilized 512 and 100 complex points in the ^{1}H and ^{15}N dimensions, respectively, with corresponding spectral windows of 13 ppm and 30 ppm, utilizing recycling delays of 1.5 s. For all experiments, data were acquired with a constant relaxation time (T_{CPMG}) of 60 ms using 9 sets of delays between the π-pulses corresponding to effective lock fields (ν_{CPMG}) of 33.33, 66.67, 100, 200, 300, 500, 800 and 1000 Hz. An experiment that excludes the CPMG period was used to obtain reference intensities (I_{ref}) and the corresponding errors (Δ_{ref}). The effective relaxation rate, $R_{2,eff}(\nu_{CPMG})$ and the corresponding error, $\Delta R_{2,eff}(\nu_{CPMG})$, were obtained using:

$$R_{2,eff}(\nu_{CPMG}) = -\frac{1}{T_{CPMG}}\ln\left(\frac{I_{CPMG}}{I_{ref}}\right) \tag{S2a}$$

$$\Delta R_{2,eff}(\nu_{CPMG}) = \frac{1}{T_{CPMG}}\sqrt{\left(\frac{\Delta_{ref}}{I_{ref}}\right)^2 + \left(\frac{\Delta_{CPMG}}{I_{CPMG}}\right)^2} \tag{S2b}$$

I_{CPMG} are the peak intensities for a given field-strength (ν_{CPMG}) and Δ_{CPMG} are the corresponding errors. The exchange contribution (R_{ex}) to the relaxation rates were obtained for those resonances ascertained to display exchange by visual inspection of the corresponding dispersion curves using

$$R_{ex} = R_{2,eff}(33.33) - R_{2,eff}(1000) \tag{S3}$$

Generation of ambiguous interaction restraints

Ambiguous interaction restraints (AIRs) for use in the data-driven docking calculations (described below) were obtained for Wzc$_{CD\Delta C}$ and Wzb. For Wzc$_{CD\Delta C}$, active residues were defined as those that were solvent exposed and displayed average attenuations of ~90% at a Wzc$_{CD\Delta C}$:Wzb molar ratio of 1:0.25. These include Arg511 and Arg514, that form part of the conserved motif implicated in Wzb binding as demonstrated through our previous mutational analyses (6). Passive residues were defined as solvent exposed residues within 6.5 Å of the active set and which were attenuated by at least 50%. Residues were defined as solvent exposed if they displayed at least 20% of relative solvent accessibility (RSA) determined using the program NACCESS. For Wzb, active residues were selected as those with significant exchange contributions ($R_{ex} > 10$ s^{-1} at an $f_B = 0.24$). Passive residues were defined as solvent exposed residues within 6.5 Å of the active set and which show R_{ex} values > 5 s^{-1} at $f_B = 0.24$. The AIRs used for the structure calculations are listed in Table S2.

Measurement of paramagnetic relaxation enhancement

Samples used for the paramagnetic relaxation enhancement (PRE) measurements contained 200 μM of ^{2}H,^{15}N-labeled Wzb. ^{2}H-labeled Wzc$_{CD\Delta C,\Delta Cys}$ variants in their spin-labeled forms were used in the following molar ratios: 1:0.5 for Val466Cys and Gln523Cys, 1:0.36 for

Ser516Cys, and 1:0.34 for Lys556Cys. For all spin-labeled variants, one sample each of the paramagnetic or the diamagnetic (in which the spin-label was reduced by incubation with 1 mM ascorbic acid for 2 h) species, were prepared. All measurements were performed at 25 °C at 600 MHz. ^{1}H,^{15}N-TROSY spectra of Wzb were acquired using 1024 and 256 complex points in the direct and indirect dimensions, respectively, with corresponding spectral widths of 14.02 ppm and 30.5 ppm. A recycle delay between scans of 4 s was used to insure adequate magnetization recovery for both the diamagnetic and paramagnetic states.

Generation of unambiguous distance restraints from PREs

The PREs induced on the amide ^{1}H positions by the spin-label were measured from peak intensities in a ^{15}N, ^{1}H TROSY experiment using

$$\frac{\mathrm{I_{para}}}{\mathrm{I_{dia}}} = \frac{\mathrm{R_{2,dia}}}{\mathrm{R_{2,dia}}+\mathrm{R_{2,para}}}\,\mathrm{e}^{-\mathrm{R_{2,para}T}} \qquad \text{(S4)}$$

where $\mathrm{I_{para}}$ (σ_{para}) and $\mathrm{I_{dia}}$ (σ_{dia}) represent the intensities (and the associated errors) for corresponding cross-peaks in the ^{1}H,^{15}N-TROSY spectra of Wzb in the presence of the oxidized and reduced forms, respectively, of the specific IPSL-labeled single cysteine variant generated on a cysteine-less background of $\mathrm{Wzc_{CD\Delta C}}$ ($\mathrm{Wzc_{CD\Delta C,\Delta Cys}}$); $\mathrm{R_{2,para}}$ and $\mathrm{R_{2,dia}}$ are the corresponding effective relaxation rates for the amide ^{1}H; T = 8 ms is the total evolution time. For residues for which the resonances were broadened beyond the threshold of detection in the paramagnetic spectrum, the $\mathrm{I_{para}}$ value was set to the noise level for the corresponding spectrum. The errors corresponding to the intensity ratios from Equation S4 were determined using:

$$\sigma_{\mathrm{ratio}} = \frac{\mathrm{I_{para}}}{\mathrm{I_{dia}}}\sqrt{\left(\frac{\sigma_{\mathrm{para}}}{\mathrm{I_{para}}}\right)^2 + \left(\frac{\sigma_{\mathrm{dia}}}{\mathrm{I_{dia}}}\right)^2} \qquad \text{(S5)}$$

For each amide ^{1}H, the $\mathrm{R_{2,dia}}$ was determined from the ^{1}H,^{15}N TROSY spectrum in the following way: the linewidth at half-height ($\Delta\nu_{1/2}$) in the ^{1}H dimension was obtained through a Lorentzian fit using NMRviewJ (2); the $\Delta\nu_{1/2}$ values were then used to calculate $\mathrm{R_{2,dia}} = \pi\ \Delta\nu_{1/2}$. The $\mathrm{R_{2,para}}$ values were then obtained using an iterative numerical protocol and Equation S4. The raw $\mathrm{R_{2,para}}$ values were adjusted for the bound fraction (f_{B}) using the relevant K_{D} values (listed in Table 1) for each variant to obtain $\mathrm{R_{2,para,corr}}=\mathrm{R_{2,para}}/f_{\mathrm{B}}$ (referred to as $\mathrm{R_{2,para}}$ for simplicity and then converted into intermolecular distances using:

$$\mathrm{r}_{SL} = \left[\frac{\mathrm{K}_{SL}}{\mathrm{R_{2,para,corr}}}\left(4\tau_{\mathrm{c}} + \frac{3\tau_{\mathrm{c}}}{1+\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}^2}\right)\right]^{1/6} \qquad \text{(S6)}$$

where $\mathrm{r_{SL}}$ is the distance between the unpaired electron on IPSL and a given amide ^{1}H on Wzb; ω_{H} is the ^{1}H Larmor frequency; the constant $\mathrm{K_{SL}} = 1.23 \times 10^{-32}\ \mathrm{cm^6\ s^{-2}}$. The rotational correlation time (τ_{c}) value was estimated as follows: An initial docking simulation was performed (as described below) using only the AIRs (see above). 200 structures generated by this calculation were grouped into 11 clusters each containing at least 4 structures. The lowest energy structure for each cluster was then utilized to obtain an estimate for τ_{c} using the HYDRONMR software suite (7). An average value (35.4 ± 0.8 ns) was then used in Equation S6 to convert the measured PREs into distances. It is important to note that altering the τ_c value has little influence on the quality of the determined docked ensemble given the relatively generous error bounds used for the distance restraints in the docking protocol. Additional bounds of ± 2 Å were added to the distances in all cases, as suggested by Battiste and Wagner (8), to compensate for the experimental uncertainties that arise from using a single time-point measurement. Further error bounds of ± 2 Å were introduced to compensate for the differences in the topologies of MTSL (modelled in the docking calculations, as described below) and IPSL (used for spin-labeling). The set of 59 unambiguous distance restraints obtained from the PRE measurements, defined between the nitroxyl oxygen of the MTSL (*in lieu* of IPSL) moieties on spin-labeled $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ variants and the amide ^{1}H of Wzb, are listed in Table S3.

Generation of starting structures for docking calculations

The NMR structural ensemble of Wzb (PDB: 2FEK) (9) was used for the docking calculations. The structural ensemble of $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ used for the calculation was generated using a single chain extracted from the octameric crystal structure (PDB: 3LA6) (10) for which the missing RK-cluster was modelled as described in Hajredini et al. (11) and subjected to a short 1 ns classical MD using the GROMACS suite (12) and standard protocols. A set of 20 structures were extracted from the generated ensemble at 50 ps time intervals. This exercise was necessary to provide some degree of variability for the loops (including the disordered regions of the RK-cluster) rather than use a single conformation represented by the crystallographic form. The sites of interest, Val466, Ser516, Gln523 and Lys556, were mutated and the MTSL moieties (CYM; in lieu of IPSL) were introduced using tools available in the HADDOCK suite (13). For Wzb, the entire NMR ensemble (20 structures; PDB: 2FEK) (9) was used for the docking calculations.

Data-driven docking

Structure calculations were performed using all ambiguous and unambiguous restraints using the HADDOCK suite (13). Docking simulations consisted of three consecutive stages. In the first, rigid body docking and energy minimization stage, a total of 10,000 structures of the complex were calculated allowing each of structures from the Wzb and $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ ensembles to explore a sufficiently broad landscape of initial orientations. At the end of this stage, 400 structures with the lowest energy scores were selected for the simulated annealing (second) and water refinement (third) stages. Atoms close to the interface (if found below the pre-defined threshold of ~6 Å) on both $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ and Wzb, were allowed to move. The spin-labels (at positions 466, 516, 523 and 556 on $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$) were kept fully flexible during this stage of molecular docking. After the final stage of the docking simulation, 381 structures were grouped into 4 major clusters, each comprising of at least 10 individual structures, representing about 95.3% of the 400 water-refined models generated after the final refinement stage. The statistics are listed in Table S4. The most relevant cluster ($\mathrm{Cluster}_{\mathrm{ALL,1}}$), as defined by the most favorable HADDOCK score (HS, -89.0 ± 1.5), contained ~56% (212 of 381) of the clustered structures, with an average RMSD value of 1.6 ± 1.0 Å relative to the lowest energy structure.

To test the influence of the PRE-based unambiguous distance restraints on the calculated structural ensemble of the Wzb•$\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ complex, a docking simulation was carried out using identical parameters as above but discarding all AIRs. In this case, 371 of the 400 structures (92.8%) generated after the final water refinement stage could be grouped into 4 major clusters of at least 10 structures each (statistics in Table S5). The cluster ($\mathrm{Cluster}_{\mathrm{PRE,1}}$) with the most favorable HADDOCK score (-72.6 ±1.1), remained the most populated cluster containing ~64% of the (237 of 371) clustered structures. The overall statistics for the simulations with or without the AIRs were quite similar (compare Tables S4 and S5), however inclusion of the AIRs suggested an improvement in the models as reflected in more favorable HADDOCK scores (-89.0 versus -72.6), lower RMSDs (1.6 versus 2.0 Å), and a larger buried surface area (1664.8 versus 1354.8 $Å^2$) comparing the dominant clusters in the two cases.

HADDOCK calculations were also performed using only the AIRs and the same parameters as above (statistics shown in Table S6). However, the absence of unambiguous distance restraints limited convergence. After the final water refinement stage, only ~60% of generated structural models could be grouped into clusters comprising at 10 models in this case. Further, no clear dominant cluster could be ascertained as for the cases described above. For example, clusters 1, 2 and 3 encompassing 59, 54 and 38 structures, respectively showed similar HADDOCK scores of -119.2, -120.6 and -123.9. Further, the models obtained without unambiguous restraints satisfied only ~40% of the restraints obtained from the PRE data.

Modelling the C-terminal Y-cluster at the Wzb active-site

Using the lowest energy structure of the Wzb•$Wzc_{CDΔC}$ complex obtained from the docking calculations (using all restraints) above, the C-terminal tail of $Wzc_{CDΔC}$ containing the Y-cluster was introduced into the Wzb active-site (to generate the Wzb•Wzc_{CD} complex) using the MODELLER suite (14). Distance restraints between the Tyr -OH group for each of Tyr705, Tyr708, Tyr710, Tyr711, Tyr713 and Tyr715, and the Wzb active-site were introduced based on the structure of *S. cerevisiae* LMW-PTP, LTP1, bound to p-nitrophenyl phosphate (PDB ID: 1D1P) (15). A total of 10 structural models were generated for each of the 6 cases.

Protein dephosphorylation assays

Phosphorylated Wzc_{CD} (prepared by incubation with ATP/Mg^{2+} as described above) and all Wzb constructs (wild-type and the following mutants: D77K/K110D and R98A) were purified in a buffer containing 50 mM Tris (pH 7.5), 100 mM NaCl, 10% glycerol, 1 mM EDTA and 2 mM Tris(2-carboxyethyl)phosphine (TCEP, Sigma). Dephosphorylation reactions were performed using equimolar amounts of phosphorylated Wzc_{CD} and Wzb (or a mutant thereof) to target final concentrations of 10 μM followed by 2-fold dilutions of the assay mixtures with dephosphorylation buffer comprising 100 mM sodium citrate (pH 6.5), 1 mM EDTA and 2 mM BME. The reaction mixture was incubated at 37 °C for 0, 1, 2, 5, 10, 20 and 30 mins. The reaction was quenched by diluting the reaction mixture into SDS-PAGE loading buffer and boiling at 97 °C for 5 min. The reaction products were analyzed by immunoblot using the monoclonal anti-phosphotyrosine antibody, PY20 (dilution 1:5000, Invitrogen) as the primary antibody and Alexa Fluor 680 goat anti-mouse (dilution 1:10000, ThermoFisher) as the secondary antibody. Membranes were visualized using a Li-Cor Odyssey Infrared Imaging System and quantified using ImageJ (16). All measurements were performed in triplicate and the rate of dephosphorylation (k_{dephos}) was obtained from fitting the experimental data to Equation S7.

$$\mathrm{S(t)} = Se^{-k_r t} + S_0 \qquad \text{(S7)}$$

Where S represents the maximum signal amplitude and S_0 represents the basal signal determined using $Wzc_{CDΔC}$ as control.

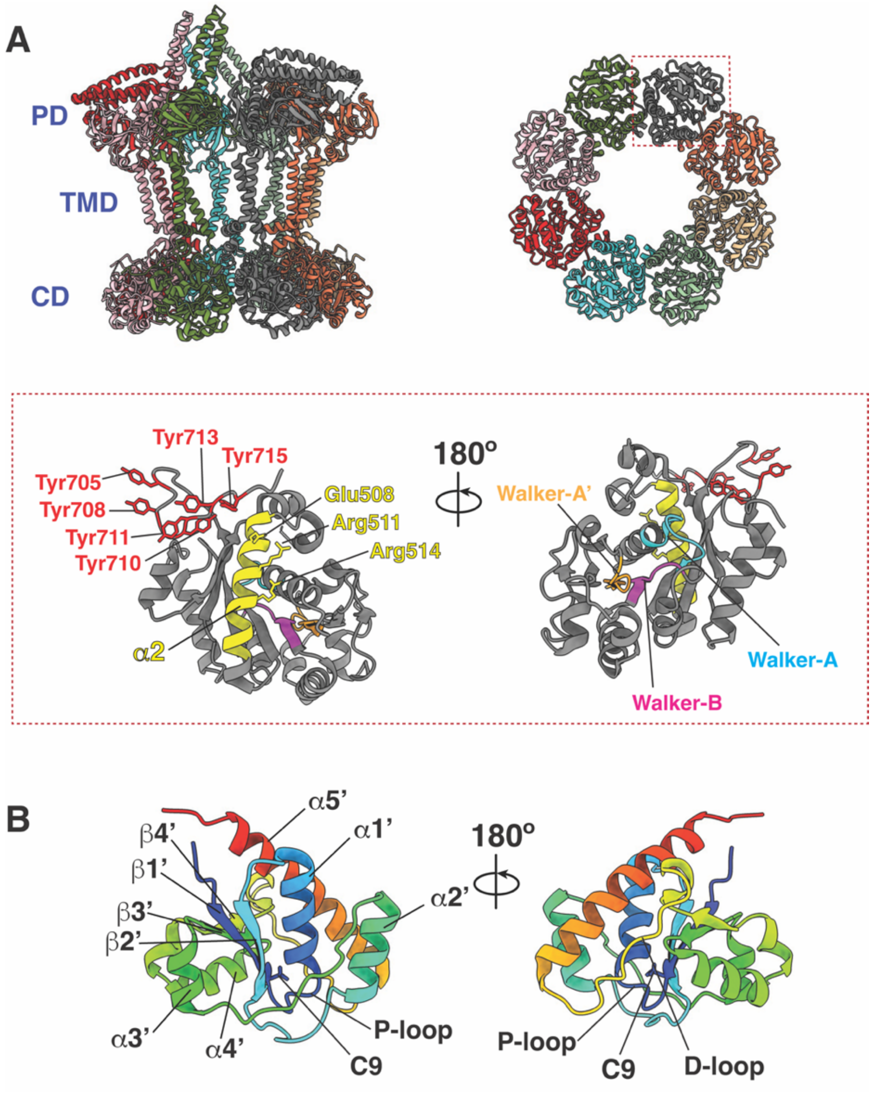


Fig. S1 (A) The structure (PDB: 7NII) (17) of full-length *E. coli* Wzc (K30) is shown on the top left panel; the periplasmic domain (PD), the two-pass trans-membrane domain (TMD) and the cytoplasmic catalytic domain (CD) are indicated. Each monomer of the octameric assembly is indicated by a different color. The top right panel depicts the CD viewed along the membrane normal from the cytoplasmic side; the TMD and PD have been removed to aid in visualization. A detailed view of a constituent monomer from the structure of the CD of *E. coli* (K-12) Wzc (PDB: 3LA6) (10) is shown in the bottom panel. The tyrosine residues, Tyr705, Tyr708, Tyr710, Tyr711, Tyr713, and Tyr715 that constitute the Y-cluster, are indicated (in red). The Walker A, A' and B motifs are colored cyan, orange and magenta, respectively. Also shown in yellow is the α2 helix with key conserved residues, Glu508, Arg511 and Arg514, indicated. **(B)** A representative structure (PDB: 2FEK) (9) from the NMR solution ensemble *E. coli* (K-12) Wzb indicating the regions of secondary structure (primed to distinguish them from Wzc_{CD}) as defined by Lescop *et al.* (9). The catalytic P-loop cysteine (C9) that is part of a conserved C-$[X]_5$-R-[S/T] PTP signature motif (18) is shown in stick representation and labeled. Also indicated is the D-loop that contains an invariant DPY sequence.

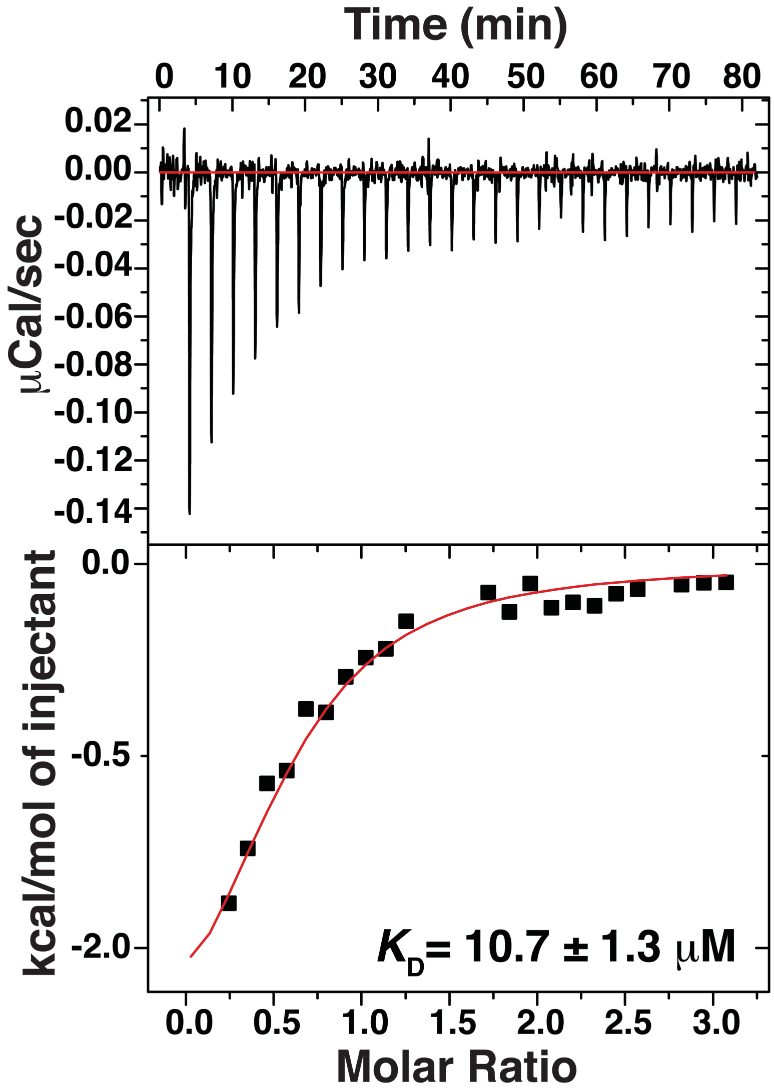


Fig. S2 A representative thermogram from isothermal titration calorimetry (ITC) measurements of the interaction of wild-type $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ with Wzb. The K_{D} value was obtained using a 1-site binding model. As shown here, the enthalpy change accompanying the interaction with Wzb is quite modest for wild-type $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$. This is especially true in case of the $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ variants generated for the attachment of spin-labels. Therefore, for precision and consistency, we relied on NMR methods to determine K_{D} values as described in the text.

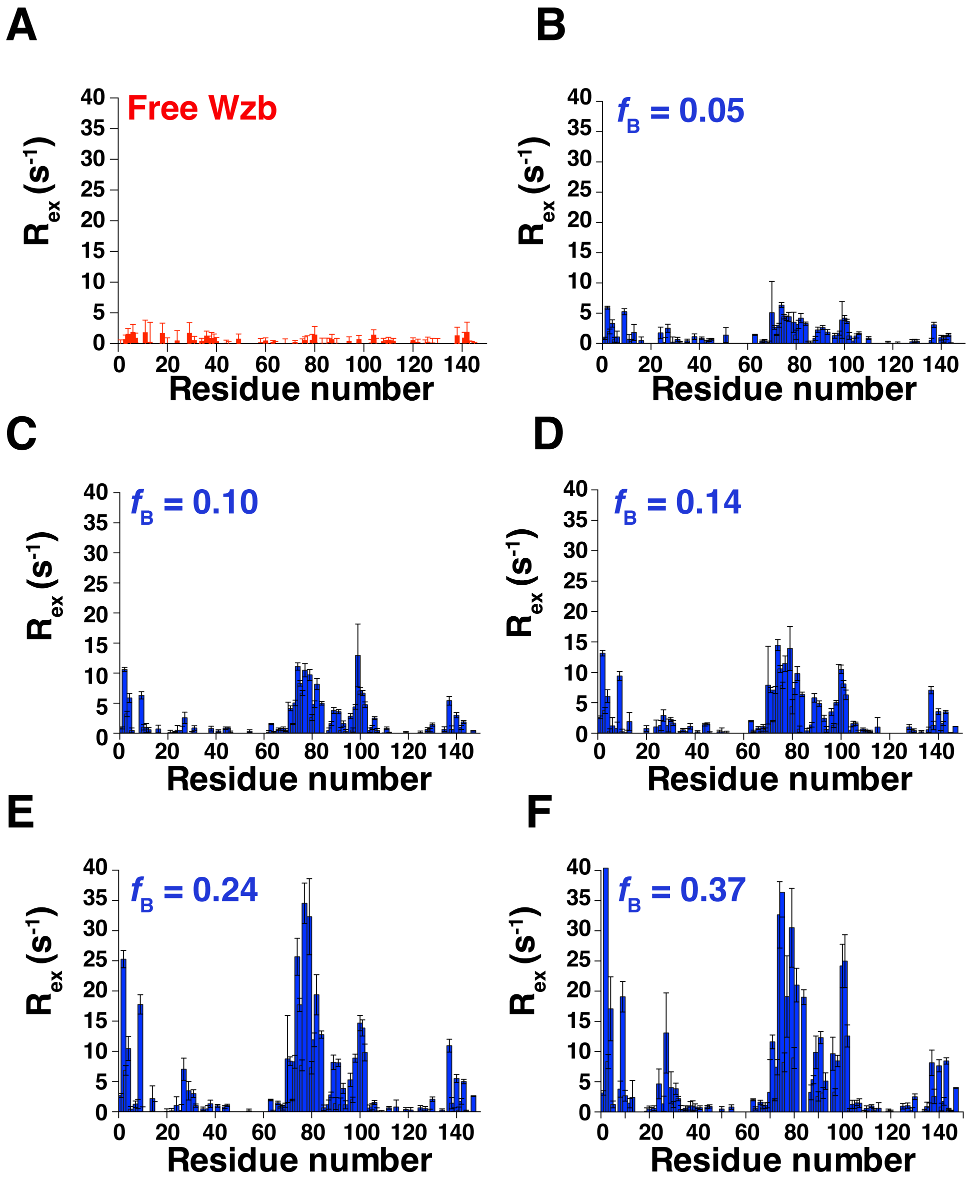


Fig. S3 Exchange contributions to Wzb amide ^{15}N relaxation in the presence of varying amounts of $Wzc_{CD\Delta C}$. **(A)** Free Wzb shows no residues with significant R_{ex} values. Increasing concentrations of $Wzc_{CD\Delta C}$ lead to increase in the number of residues that display substantial R_{ex} values and in the magnitudes of the values themselves. The R_{ex} values are plotted against Wzb residue number for several bound-state fractions (f_B): **(B)** 0.05, **(C)** 0.10, **(D)** 0.14, **(E)** 0.24, and **(F)** 0.37. Errors in the R_{ex} values were determined from the standard deviation of the noise for each spectrum using Equations S2 and S3.

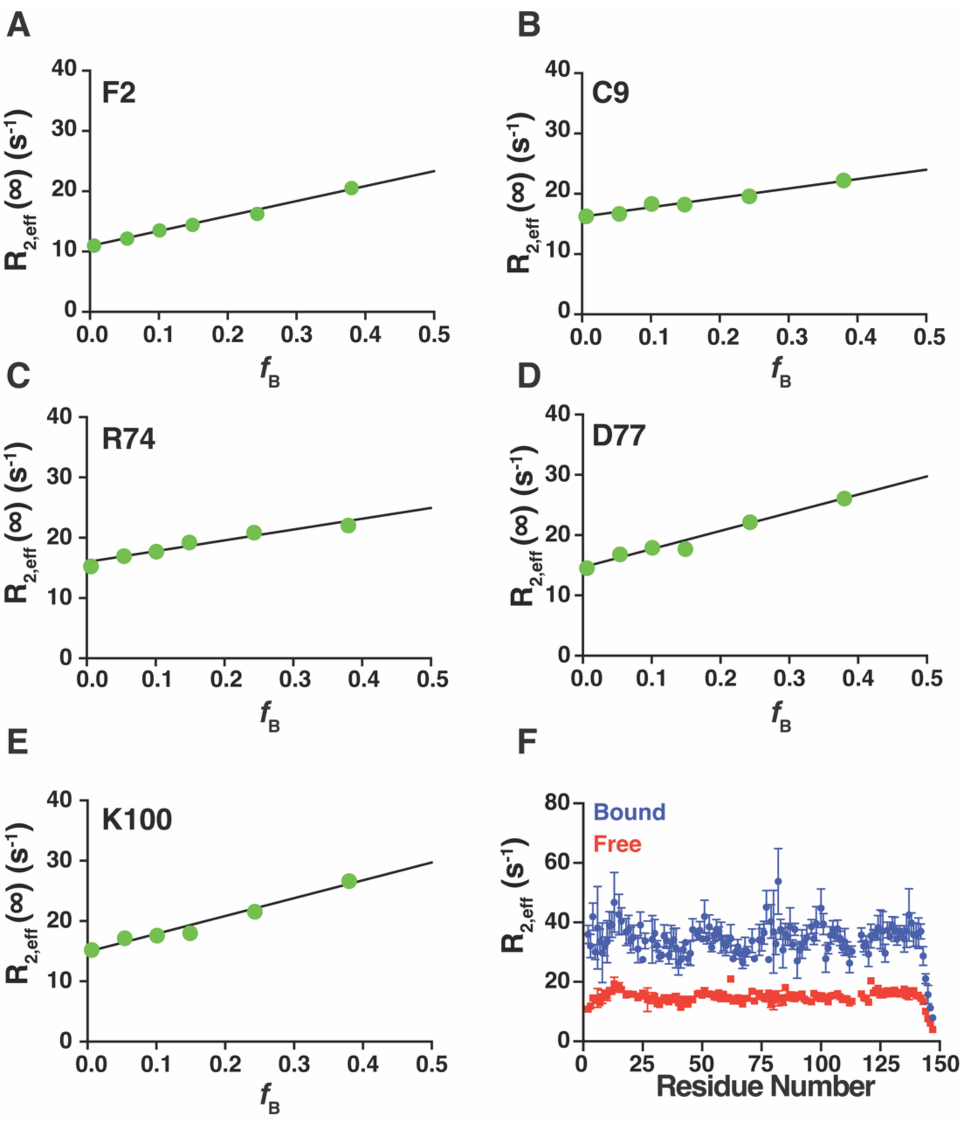


Fig. S4 (A-E) Representative regression plots demonstrating the linear dependence of the limiting effective relaxation rate [$R_{2,eff}$ (∞); using corresponding values at the highest pulsing frequency, ν_{CPMG} =1000 Hz, as proxy] against the fraction of Wzb in the Wzb•Wzc$_{CD\Delta C}$ complex (f_B). **(F)** Plots of extrapolated $R_{2,eff}$ values for the free (f_B=0, red) and fully bound states (f_B=1, blue) of Wzb. Data for G99 and D115 that have large errors have been excluded from the plot.

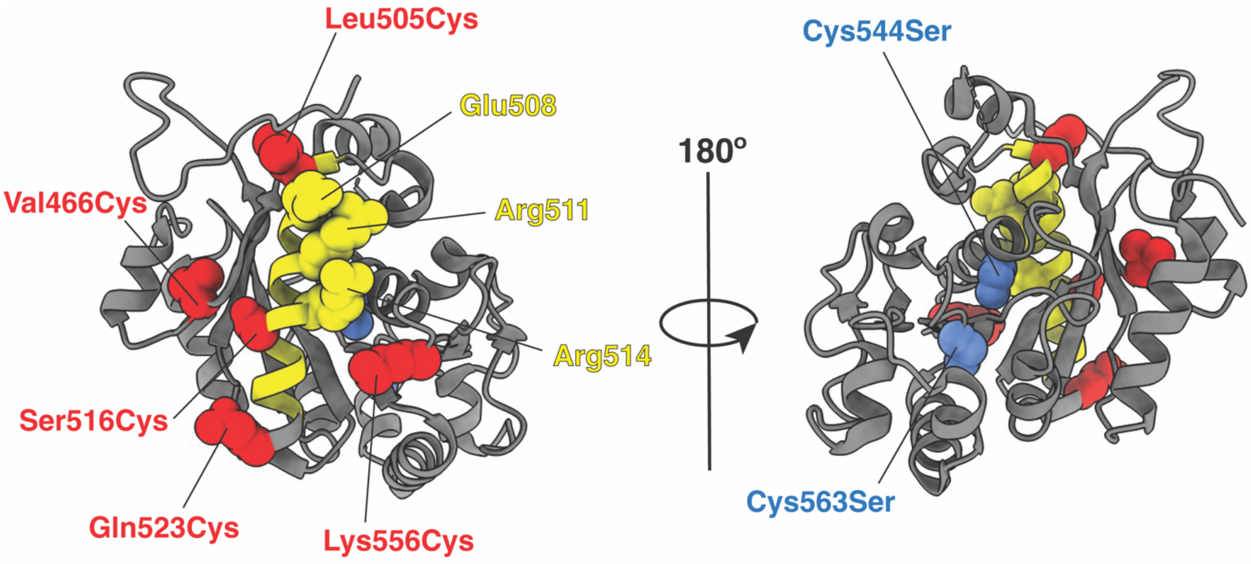


Fig. S5 $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ positions used for the attachment of the IPSL spin label. The two native cysteines (Cys563, Cys544; blue) were mutated to serine to generate a cysteine-less background ($\mathrm{Wzc}_{\mathrm{CD\Delta C,\Delta Cys}}$) into which non-native cysteines (Val466Cys, Leu505Cys, Ser516Cys, Gln523Cys and Lys556Cys; red) were introduced, one at a time, for the attachment of a single IPSL spin-label. Also shown (in yellow) are Glu508, Arg511 and Arg514 that are part of the conserved moiety implicated in recognizing Wzb.

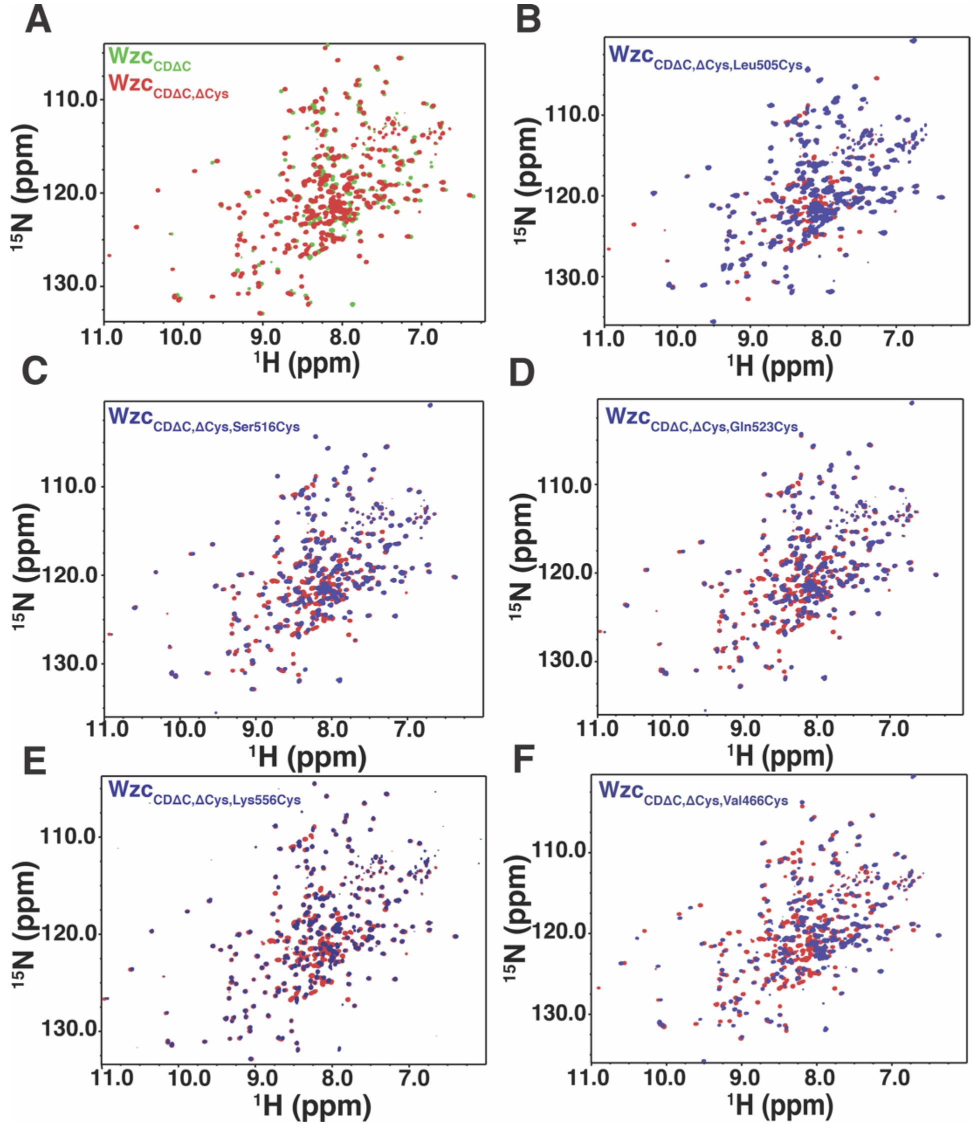


Fig. S6 (A) ^{15}N, ^{1}H TROSY spectra (600 MHz) of wild-type Wzc$_{CD\Delta C}$ (shown in green) compared to the corresponding cysteine-less mutant (Wzc$_{CD\Delta C,\Delta Cys}$: Cys544Ser/Cys563Ser, that is shown in red in all cases). This double mutation resulted in chemical shift perturbations (excluding residues within a 6 Å radius of each mutation site) of 0.04 ± 0.03 ppm (average ± standard deviation over all resonances) compared to Wzc$_{CD\Delta C}$. Corresponding spectra of the **(B)** Leu505Cys, **(C)** Ser516Cys, **(D)** Gln523Cys, (**E)** Lys556Cys, and **(F)** Val466Cys variants, all generated using the cysteine-less (Wzc$_{CD\Delta C,\Delta Cys}$) background, are shown. The spectra indicate folded species in all cases with modest chemical shift changes (relative to Wzc$_{CD\Delta C,\Delta Cys}$) of: Leu505Cys (0.04 ± 0.04 ppm), Ser516Cys (0.03 ± 0.03 ppm), Gln523Cys (0.03 ± 0.03 ppm) and Lys556Cys (0.02 ± 0.05 ppm). The perturbations were somewhat larger for Val466Cys (0.06 ± 0.07 ppm).

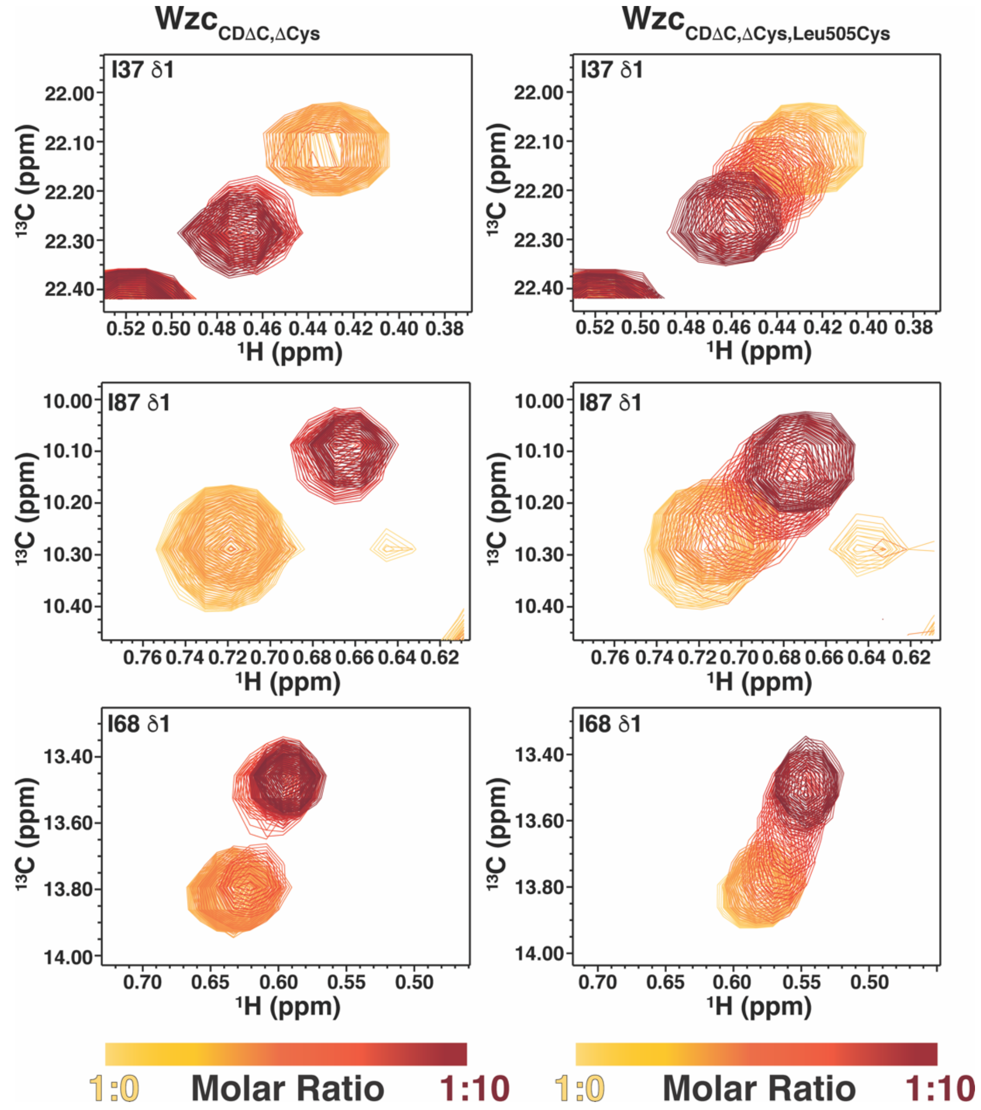


Fig. S7 Representative Ile-δ1 resonances in ^{13}C,^{1}H HMQC spectra of Wzb in the presence of increasing amounts of specific Wzc$_{\mathrm{CD\Delta C}}$ variants. The spectra illustrate the change in association dynamics comparing Wzc$_{\mathrm{CD\Delta C,\Delta Cys,Leu505Cys}}$ (fast exchange) with Wzc$_{\mathrm{CD\Delta C,\Delta Cys}}$ (slow exchange). The latter is reflective of the wild-type construct and all other variants.

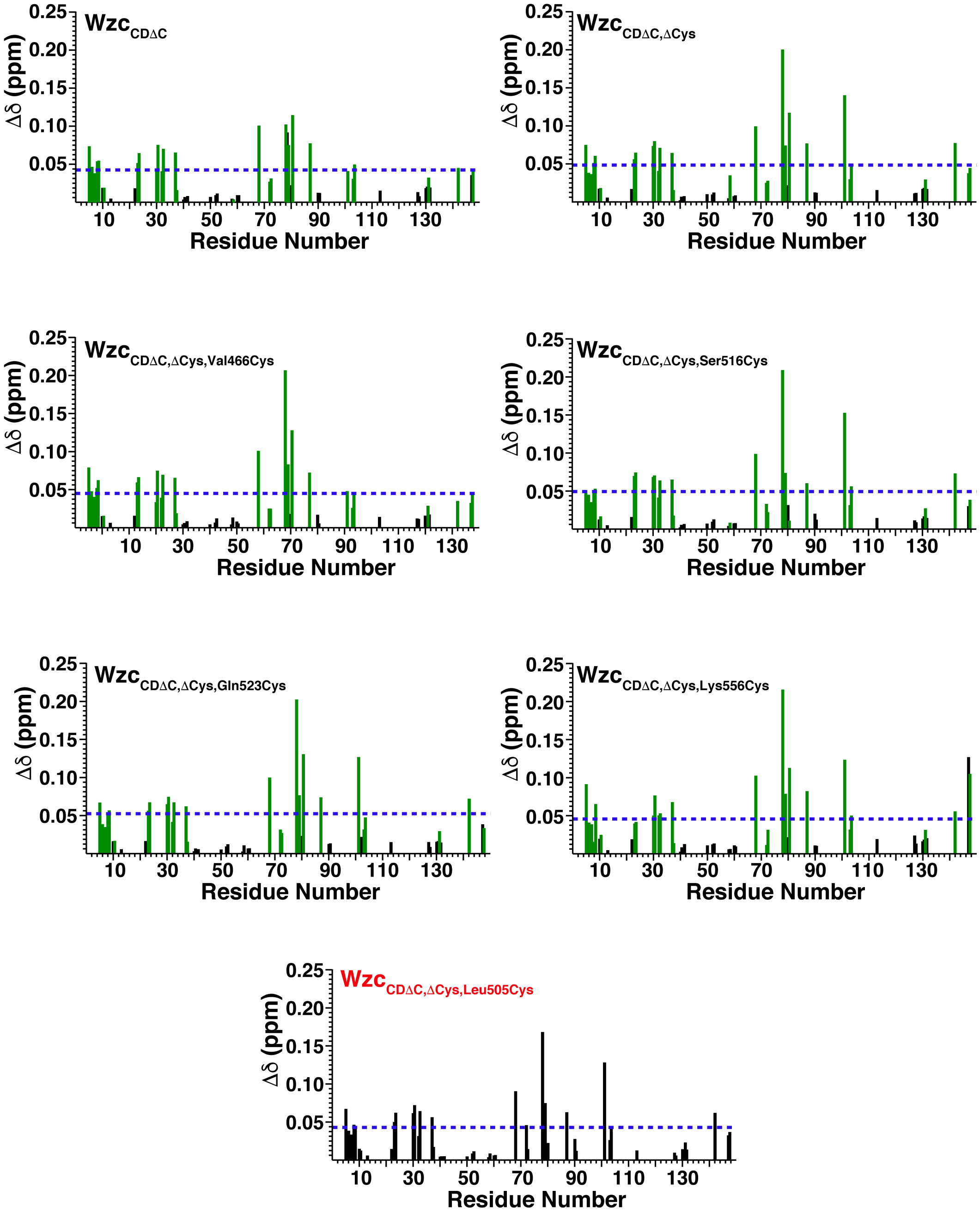


Fig. S8 Chemical shift changes of ^{13}C,^{1}H resonances of wild-type ILV-labeled Wzb in the presence of 10 molar equivalents of wild-type ^{2}H-labeled Wzc$_{CD\Delta C}$ (or variants). Green bars indicate chemical shift changes for residues in the slow exchange regime; black bars indicate residues for which the resonances are in fast exchange (or those that have small, unresolved differences between the chemical shifts of the free and bound states). Note that, unlike in all other cases, the association dynamics transitions into the fast exchange regime for the Wzc$_{CD\Delta C,\Delta Cys,Leu505Cys}$ variant (bottom panel, labeled in red) although the overall pattern of shifts remains similar to the other cases. The dashed lines indicate the standard deviation threshold above the mean in all cases.

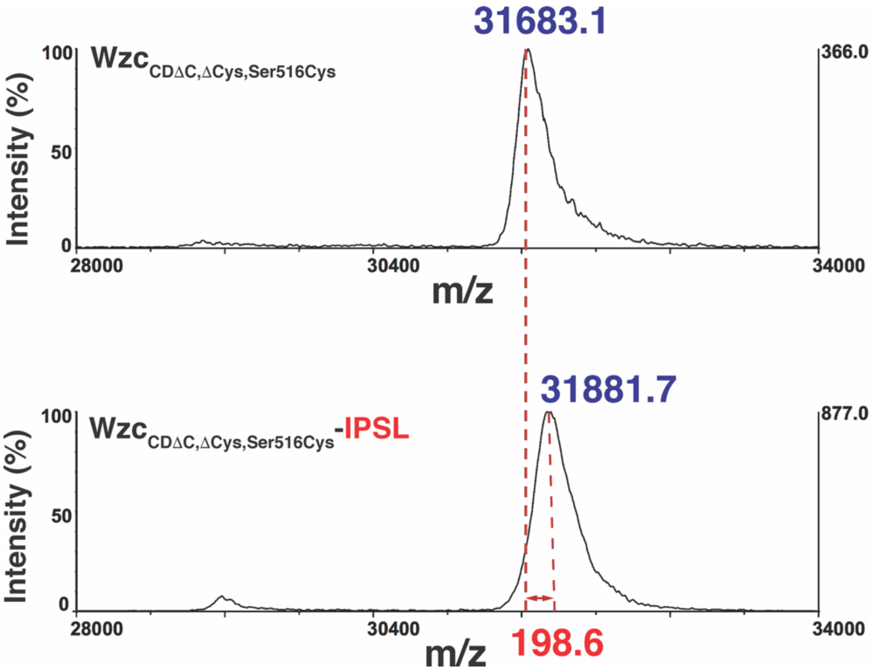


Fig. S9 Representative MALDI-TOF spectrum illustrating the covalent modification of a specific cysteine residue by IPSL. Data is shown for the $\text{Wzc}_{\text{CD}\Delta\text{C},\Delta\text{Cys,Ser516Cys}}$ variant before (top) and after attachment (bottom) of the spin-label. The bottom trace shows a mass increase of ~198.6 Da resulting from the conjugation of IPSL.

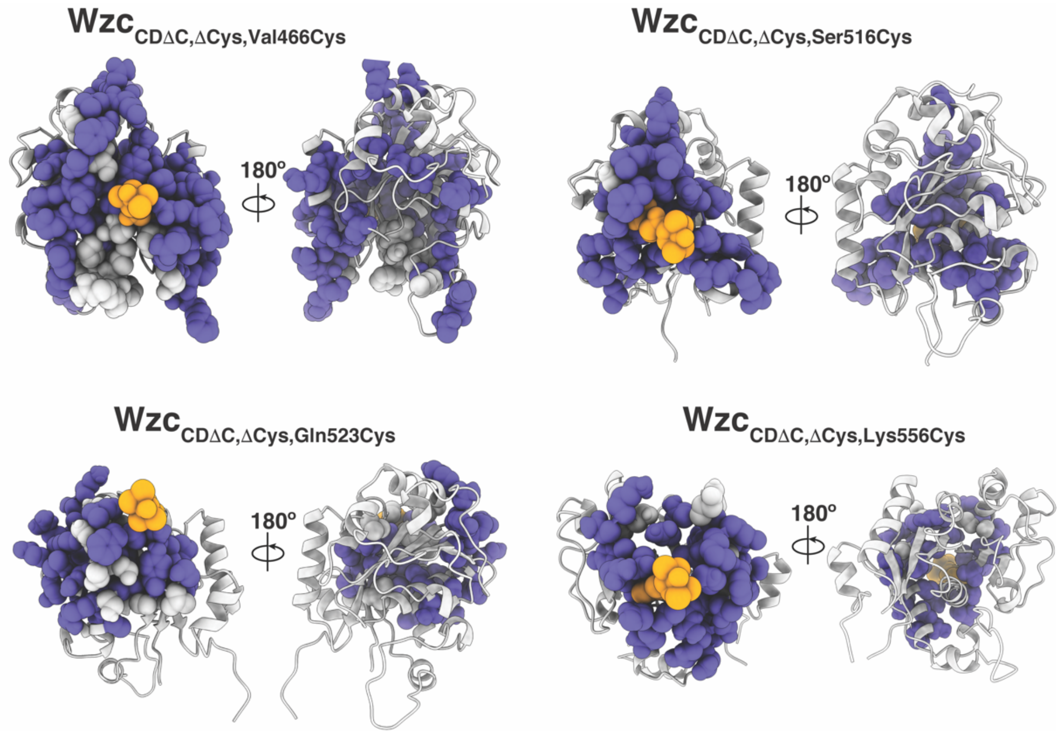


Fig. S10 Intramolecular paramagnetic relaxation effects seen upon conjugation of the IPSL spin-label for specific Wzc$_{CD\Delta C,\Delta Cys}$ variants. Residues that are affected by the presence of the spin-label (75% attenuation of the corresponding resonance in ^{15}N,^{1}H TROSY spectra at 600 MHz) are indicated by the blue spheres; the labeled cysteine is represented by yellow spheres. Residues that are spatially proximal to the spin-labeled cysteine but for which resonance assignments could not unambiguously transferred from those of wild-type Wzc$_{CD\Delta C}$ are represented by grey spheres. It is evident that the attenuations are limited to the face bearing the labeled cysteine suggesting the absence of non-specific interactions.

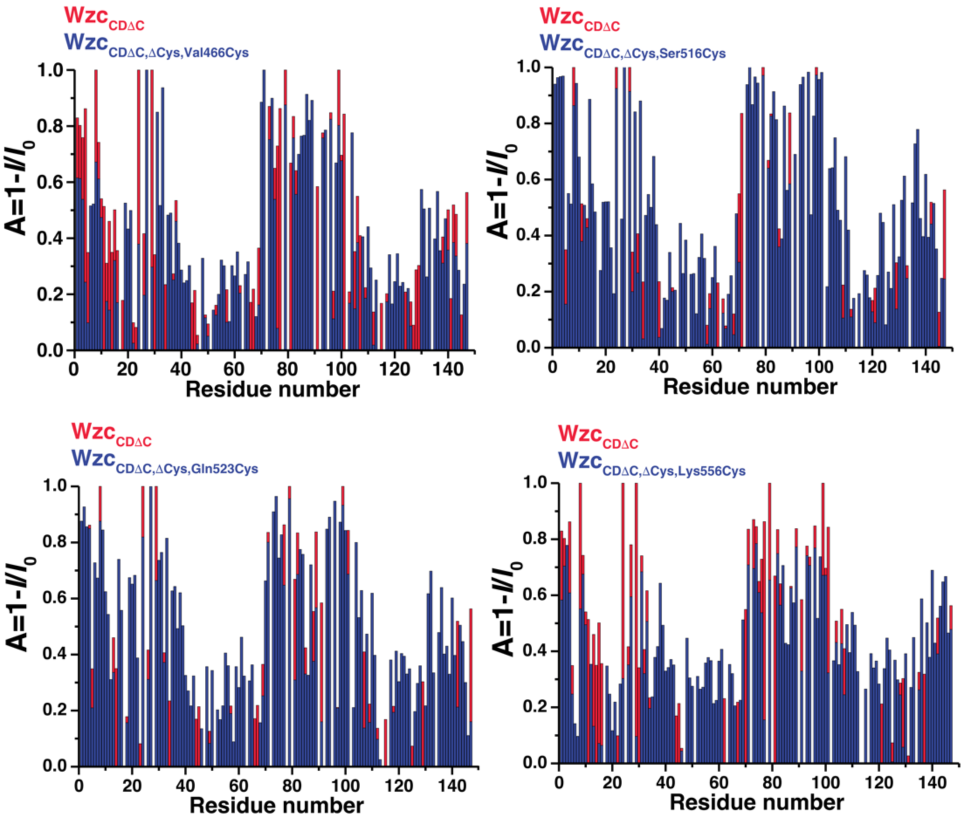


Fig. S11 Signal attenuation in $^1H,^{15}N$ TROSY spectra of uniformly $^2H,^{15}N$-labeled Wzb in the presence of the reduced forms the IPSL-modified, ^2H-labeled variants of $Wzc_{CD\Delta C}$ at 1:0.5 molar ratios are shown as blue bars. For comparison, the corresponding patterns of attenuation in the presence of an equivalent molar ratio of wild-type $Wzc_{CD\Delta C}$ (red bars) are also indicated. A=1 for resonances that are attenuated to below the noise level. The patterns of attenuations are qualitatively similar in all cases suggesting that biochemical manipulations of $Wzc_{CD\Delta C}$ including introduction of the mutations and subsequent the conjugation of the IPSL spin-label have minimal effect on the interactions with Wzb.

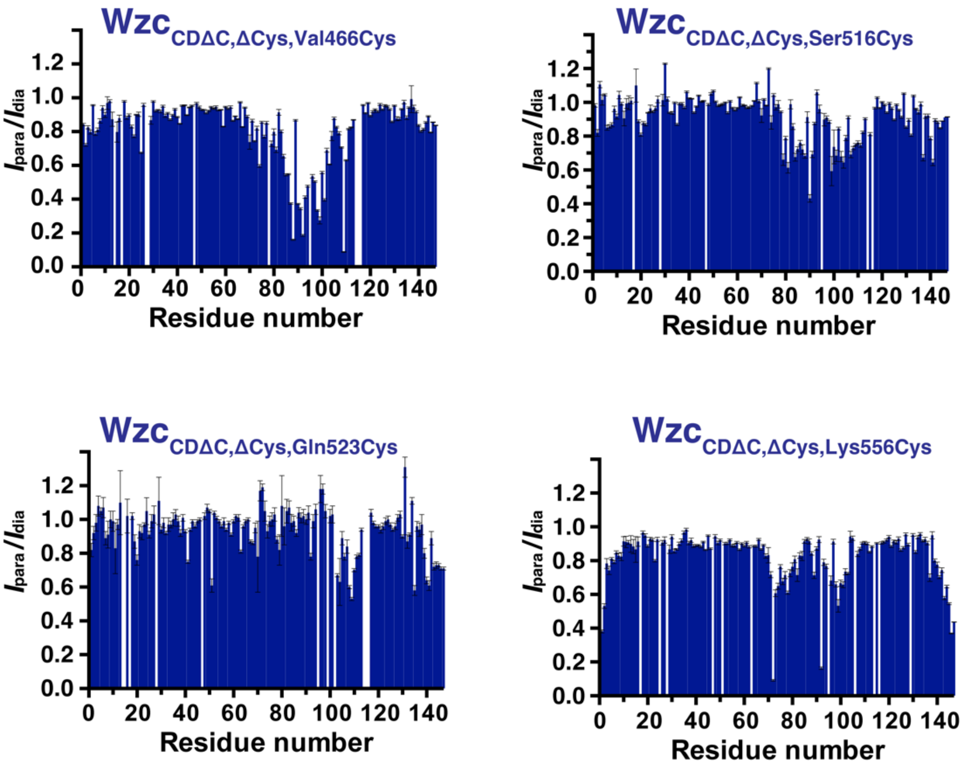


Fig. S12 Ratios of intensities of amide resonances in ^{15}N,^{1}H TROSY spectra of ^{2}H,^{15}N-labeled Wzb in the presence of IPSL-modified, ^{2}H-labeled variants of Wzc$_{\mathrm{CD\Delta C}}$ in their oxidized (I_{para}) and reduced forms (I_{dia}) plotted against Wzb residue number.

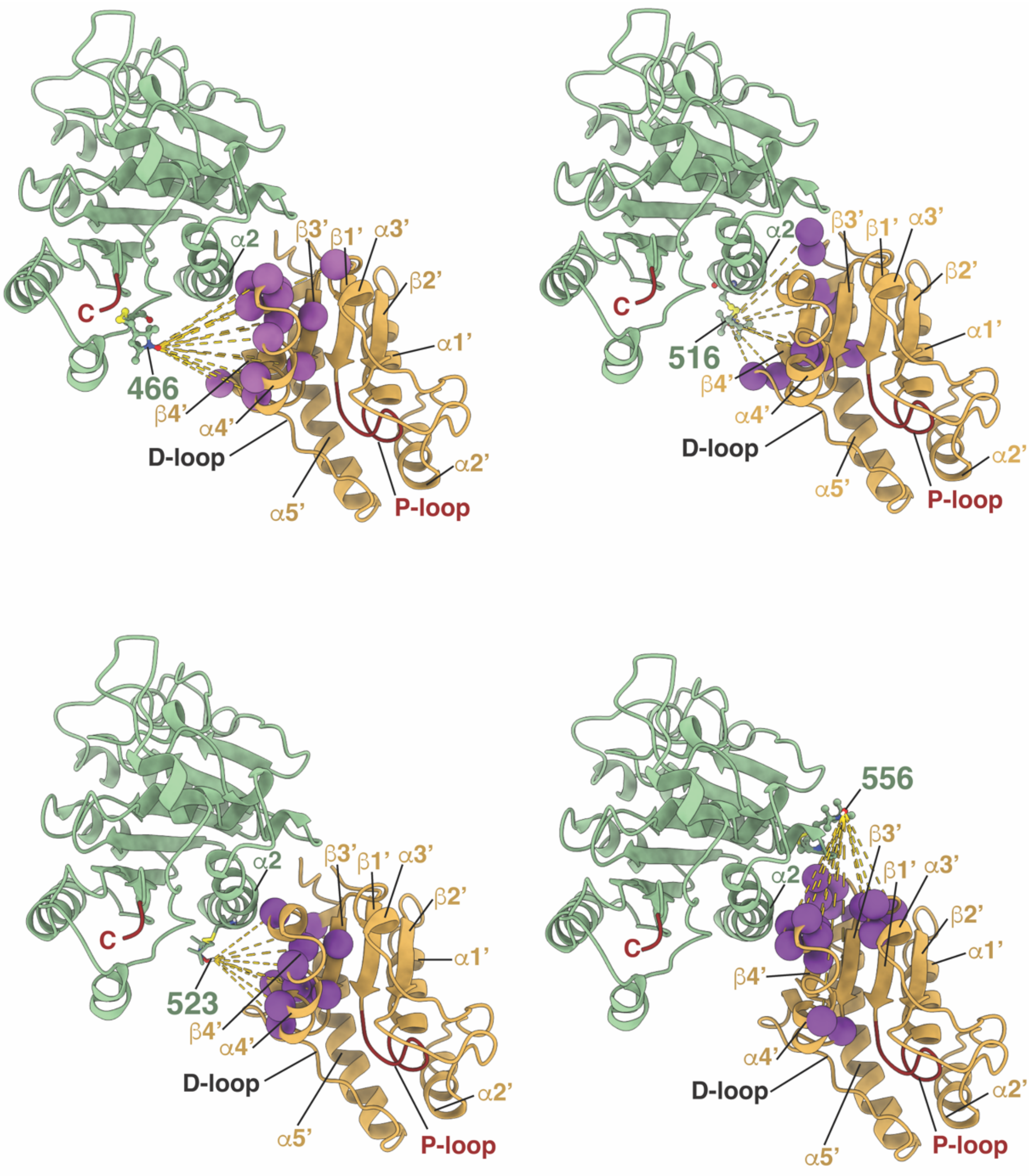


Fig. S13 Unambiguous PRE-based distance restraints are represented on the lowest energy structure obtained from the docking calculations using all i.e., unambiguous and ambiguous, restraints. The positions of the IPSL-labeled cysteines, Cys466, Cys516, Cys523, and Cys556 on $\text{Wzc}_{\text{CD}\Delta\text{C}}$ (in green) are displayed in ball-and-stick representation; the complementary N atoms of Wzb (in yellow) are shown as purple spheres. The distance restraints, listed in Table S3, used in the calculations are indicated by the dashed lines. The α2 helix of $\text{Wzc}_{\text{CD}\Delta\text{C}}$ and all elements of secondary structure on Wzb are indicated.

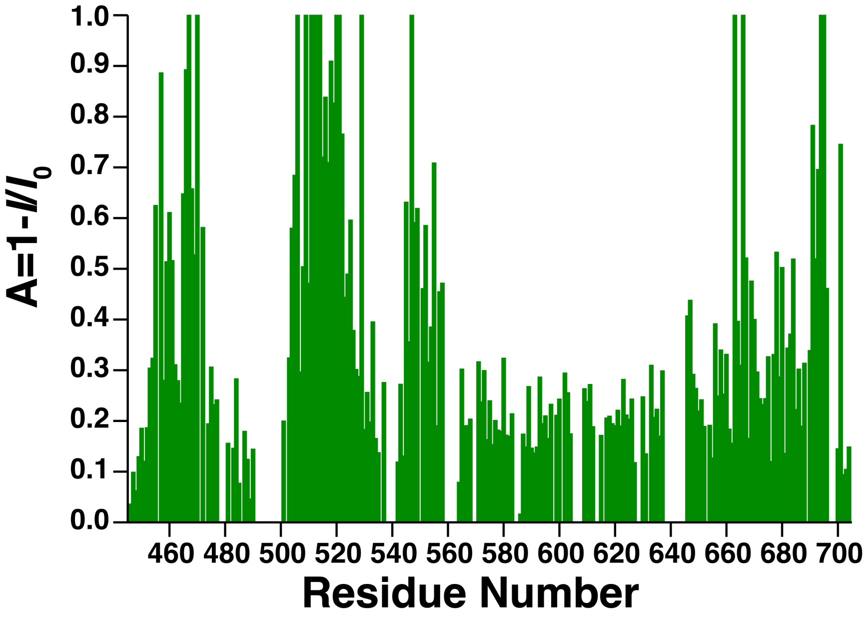


Fig. S14 Signal attenuation in ^{1}H,^{15}N TROSY spectra of uniformly ^{2}H, ^{15}N-labeled Wzc$_{\mathrm{CD\Delta C}}$ in the presence of ^{2}H-labeled Wzb at a 1:0.25 molar ratio plotted against Wzc$_{\mathrm{CD\Delta C}}$ residue number. A=1 for fully attenuated resonances, A=0 for resonances that are unaffected by the presence of Wzb.

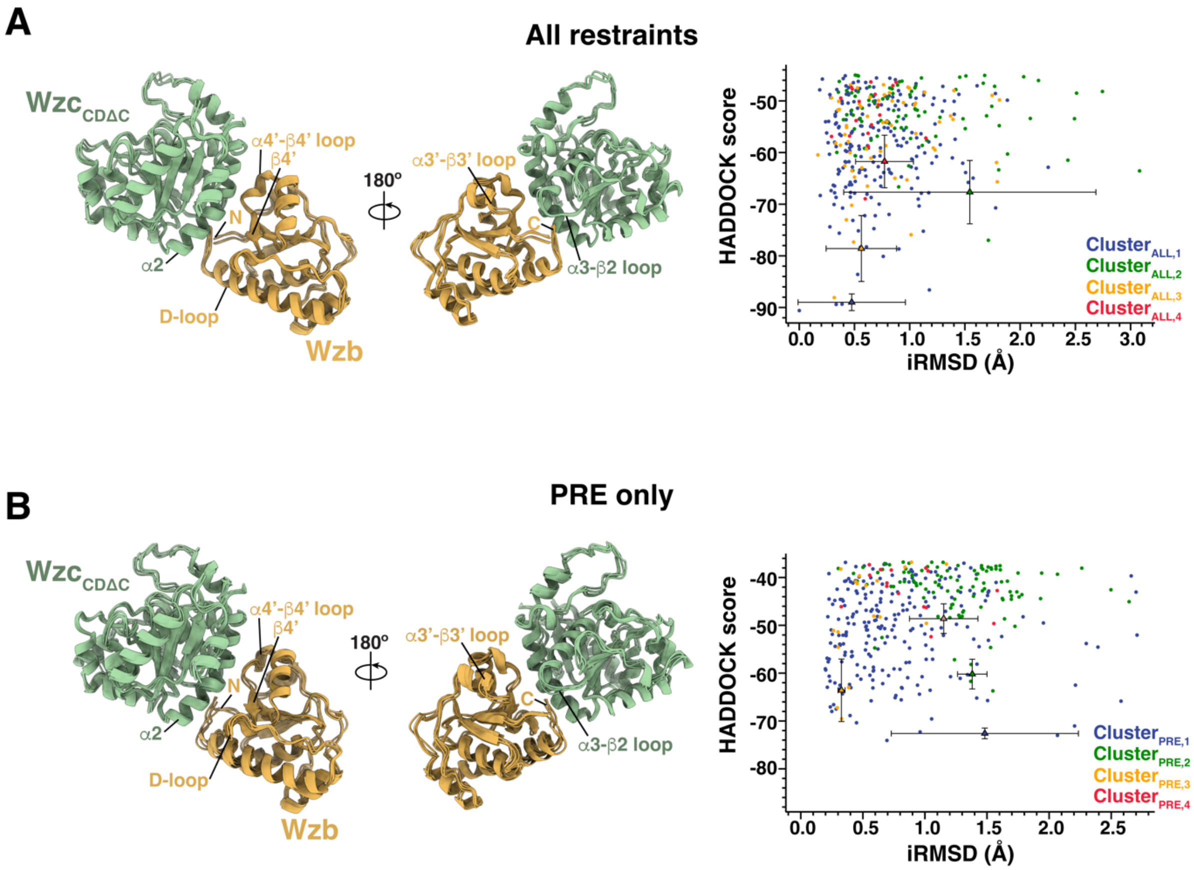


Fig. S15 Overlay of 4 lowest energy structures for the Wzb•Wzc$_{CD\Delta C}$ complex representing the results of docking calculations performed using either **(A)** all restraints (Cluster$_{ALL,i}$; i=1,2,3,4) or **(B)** only the PRE-based unambiguous distance restraints (Cluster$_{PRE,i}$; i=1,2,3,4) are shown on the left panels. The corresponding plots of HADDOCK scores versus interface RMSD (iRMSD) for all structures that comprise the relevant clusters are shown on the right panels. Cluster$_{ALL,1}$/Cluster$_{PRE,1}$ (calculated with/without AIRs) contain 212/237 structures; Cluster$_{ALL,2}$/Cluster$_{PRE,2}$: 85/101; Cluster$_{ALL,3}$/Cluster$_{PRE,3}$: 65/18; Cluster$_{ALL,4}$/Cluster$_{PRE,4}$: 19/15.

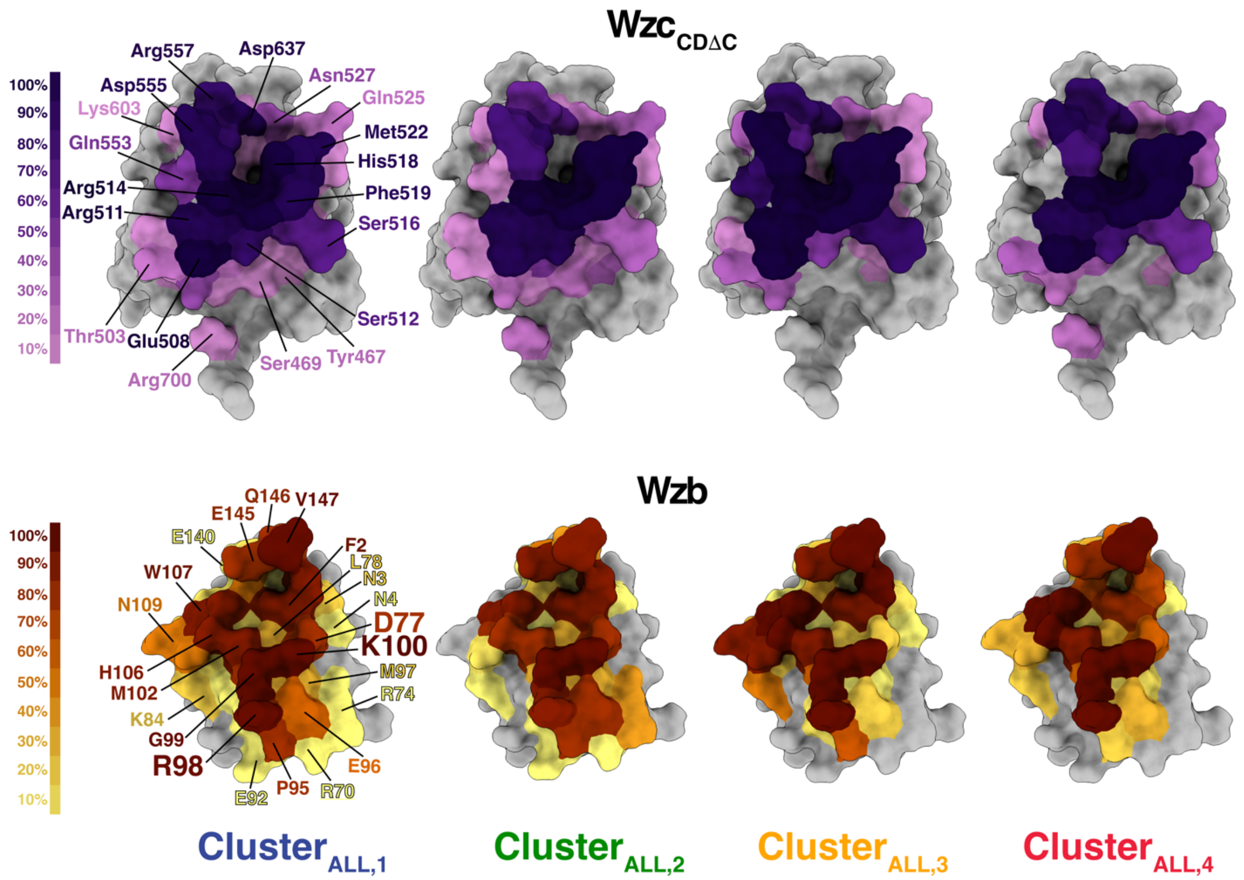
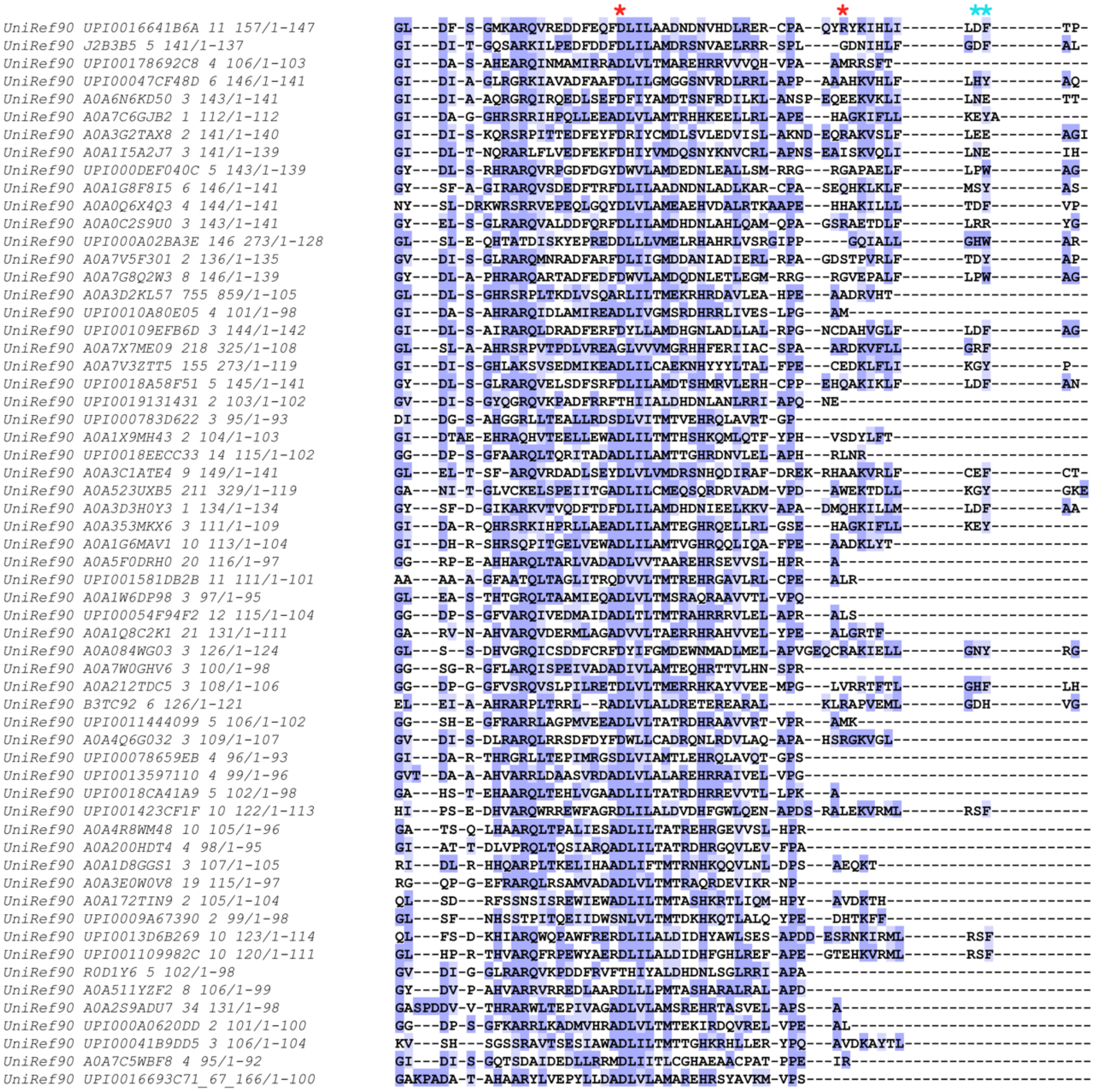


Fig. S16 Percent occurrence of Wzc$_{CD\Delta C}$ (top) and Wzb (bottom) residues at the protein-protein interface for the 4 major clusters calculated using all restraints (Cluster$_{ALL,i}$; i=1,2,3,4). Wzc$_{CD\Delta C}$ and Wzb are shown in surface representation with the occurrence of a specific residue at the interface colored using purple-to-pink or red-to-yellow gradients, respectively. An interfacial residue is defined as one that is located within 4 Å of any residue of its interaction partner. Selected residues are labeled. Residues on Wzb that have been mutated and the corresponding (R98A and D77K/K100D) mutants tested for their ability to dephosphorylate Wzc$_{CD}$ (see Fig. 4 in the main text) are indicated in larger font.

```
                                                                 *                        *              **
E. coli (K-12) Wzb seq/1-147            QL---SL-E-GHCARQISRRLCRNYDLILTMEKRHIERLCEM-APE---MRGKVMLF-------GHW--------DN-
UniRef90 A0A078LC37 1 147/1-147         NL---SL-E-GHCARQISGRMCRDYDLILAMEKRHIARLCEM-APE---MRGKVMLF-------GHW--------DS-
UniRef90 A0A5B0KH26 1 146/1-146         QL---SL-D-GHCARQISGAMCREYDLILAMERRHISKLHDI-APE---MRGKVMLF-------GHW--------DD-
UniRef90 UPI00137824A4 1 143/1-143      AL---SL-E-GHLAQQLTGTMCREYDLILVMEKGHIDAVCHF-APE---VRGKTMLY-------GHW--------LE-
UniRef90 UPI000F7E07EA 1 143/1-143      NL---SL-E-GHVAKQLTKEMCREYSLILVMEKGHIDAVCRL-APE---VRGKTMLF-------AHW--------LG-
UniRef90 A0A6P1Q0P6 3 145/1-143         SL---SL-E-GHRARKLTSAMCRENDLILVMEKQHINDISQI-APE---VRGKTMLF-------GHW--------CN-
UniRef90 A0A7U4I6Q9 1 142/1-142         GL---SL-D-GHQAQQLTAEMCRDFDLILVMEKRHIEQVNRI-DPA---ARGKTMLL-------GHW--------LN-
UniRef90 A0A7T7YS66 5 148/1-144         GL---SL-E-GHSGRQFTSALARQYDLILVMERSHIEQIGNI-APE---ARGKIMLF-------GHW--------LN-
UniRef90 A0A6I6IIB4 1 143/1-143         HL---SL-A-DHQGQQLTRQLAQQYDLILVMEKHHQEEVSAI-APE---VRGKTMLF-------AHW--------IN-
UniRef90 A0A085JA86 1 142/1-142         GL---SL-D-GHKGQQFTGKLGRQYDLILVMEKHHIEQVTKI-APE---ARGKTLLF-------GHW--------LG-
UniRef90 UPI00115AE06B 5 146/1-142      GL---SL-D-GHKGKQFTASLGRQYDLILVMERAHIEQIGNI-APE---ARGKVMLF-------GHW--------LD-
UniRef90 UPI001124D73D 5 146/1-142      GI---SL-D-NHKGRQFTASLGRKYELILAMEKYHIEQIGKI-APE---ARGKTMLF-------GQW--------IN-
UniRef90 A0A0P0YQS2 4 147/1-144         GI---SL-A-GHVGTQFTSALGRQFDLILVMEKSHIELIGRI-APE---ARGKVMLF-------GYW--------LE-
UniRef90 UPI0015DBD763 1 142/1-142      GL---SL-K-DHKGRQFTVRLGREYDLILVMEYTHLEEVTRI-SPE---IRGKTMLL-------GHW--------LN-
UniRef90 A0A368LFH4 1 146/1-146         GV---SL-E-AHQAQQLTNSLCRQYDLILVMEKGHLDAVHDI-APE---ARGKTMLF-------GQW--------IE-
UniRef90 A0A1C3T022 1 144/1-144         GL---DL-S-NHKGRQITSAMIRKYDLILVMEKGHIEKICTI-SPE---SRGKTMLL-------GHW--------ID-
UniRef90 A0A5P9CG24 1 145/1-145         GI---DL-S-NHRSQQVSSALCAEFDLILVMEKGHIEAITAI-APE---VRGKTMLF-------GHW--------TD-
UniRef90 A0A511QVN4 1 140/1-140         DC---DL-S-GHIARQITSAMIAESDLILVMEKGHIDSITKI-APA---ARGKTMLY-------GHW--------LS-
UniRef90 A0A377HMG9 6 150/1-145         GI---SL-E-GHASTQLTEAMAREYDLILVMEKGHIDAVCSI-APA---ARGKTMLF-------GQW--------IG-
UniRef90 Q84GX3 1 139/1-139             GL---SL-D-GHTARQLTSELCQQSSLILVMEKRHIEAVNRI-NLG---VRGKTMLF-------GHW--------LN-
UniRef90 A0A2N0W7W0 1 142/1-142         GL---SL-C-GHKGKQFNFTQMQKYDLILVMERLHIEYITKN-HPE---CRGKTMLL-------GHW--------MD-
UniRef90 UPI0018E32AC7 1 143/1-143      GI---DA-S-EHVARKLTRDMLVNADLILTMEPRHIKGINDL-APQ---VSGKTFLL-------GKW--------SN-
UniRef90 UPI000DEFCBB6 1 144/1-144      GV---DG-S-AHEAQQLNAAMLRDADLVLAMEREHVTQMMRM-APE---ASGKVMLL-------DRW--------GQ-
UniRef90 I4JGR3 1 142/1-142             GL---DF-T-AHRARQVDSHMLHDADLILAMEHSHIQHIRQI-APE---VHGKTFLI-------GKW--------QD-
UniRef90 UPI000BA8F6D8 1 147/1-147      EP---QANT-EHSAQQLSSELVRKADLILVMENGHLNVVNQK-YPQ---ASGKCMLL-------GKW--------DS-
UniRef90 UPI001606AA61 1 142/1-142      GL---EL-A-THRARQVKGDMLRQADLILAMERSHIQHISEI-APE---VHGKTFLI-------GKW--------QN-
UniRef90 A0A1H7YFB0 1 140/1-140         GL---DG-S-AHRARQLTPDLLAAADLILVMERHHHEQILRE-APQ---FSGKTFLA-------GKW--------RG-
UniRef90 B4WZ62 1 141/1-141             GL---DC-G-AHVARKLDSELCRDSDLILVMESRQKERLGRA-FPE---ASGKTFLL-------SHW--------NG-
UniRef90 A0A7X5QSM5 1 143/1-143         GV---DG-E-GHVARQLDAKLLRQSDLILAMERSHVDAIRRM-APE---ASGKVFLL-------DKW--------LG-
UniRef90 A0A078LP09 1 142/1-142         GL---EH-T-THRARQVDSEMLHQADLILTMEQSHIQHLRQI-APE---VHGKTFLM-------GKW--------LD-
UniRef90 A0A238IZY7 1 142/1-142         GL---SL-D-GHLARQFTAEIGAGAELILVMEPGHKREIARM-APE---LTGRMMLF-------DQW--------TG-
UniRef90 A0A1H1GM85 21 162/1-142        GY---AP-A-EHEATQLATQAISEADLILVMERKHVSGVLNI-APE---ARGKVLML-------GKW--------QN-
UniRef90 UPI0015537137 1 144/1-144      GI---NV-S-GHQARQLTPELLQKADLVLVMSDSQRRAVGEM-DPT---ALGKTLRL-------GRW--------LE-
UniRef90 A0A2W7QDH7 6 151/1-146         GL---SV-A-GHVARQFSHELGAAHALILVMEPGHKREIVKA-APD---LSGRVMLF-------DHW--------TG-
UniRef90 B2JS56 3 143/1-141             GH---SS-V-THRAKQVTPTMLRQSDLILTMSHRQRNHICSM-MPE---VRGKVFLL-------GKW--------DA-
UniRef90 A0A7Z0KZC6 18 159/1-142        GV---TL-D-GHVARQFTPEIGKAHDLILVMEPGHKREIARL-SPE---LSGRIMLF-------DHW--------TG-
UniRef90 Q5D1Y3 1 143/1-143             GL---DV-S-EHEARQLTPAMLQEADLVLVMTEGQRRAVADL-SPA---VLGKTMLL-------GRW--------LD-
UniRef90 A0A0Q2Z7E5 3 142/1-140         GV---AD-N-AHVARQFTREIGEQHDVILVLEGGHKREIERI-APE---LSGRVMRL-------SRW--------SG-
UniRef90 A0A432Z9J1 1 144/1-144         GH---SF-P-KHEARQLTREIVAANDLILVMDQEQRHYIGQR-YPE---ALAKTMLL-------GHW--------LG-
UniRef90 A0A2S7W626 1 147/1-147         GV---DL-T-HHKAQQLTKAMCLEADLILAMELKHIGLITDI-CPT---AHGKTMLL-------SQW--------VEM
UniRef90 UPI0018D26CCA 1 145/1-145      GL---DV-S-RHLARQLDARYLTQADLILVMSDEQRRLIGDR-WPQ---AVGKVMRL-------GHW--------LE-
UniRef90 A0A1G0DWF7 1 143/1-143         GV---DI-S-AHRARMLTDAIARDADIILVMDDLQKQQISMQ-YPY---ARGKVFKL-------GEA---------A-
UniRef90 A0A7W5J0Y5 3 142/1-140         GL---RL-D-HHVARQIDLRMLVEADLVLAMEQRQLRALEAI-APP---LRNRMHLL-------GRW--------KG-
UniRef90 Q4FU01 3 141/1-139             QI---DI-T-NHVAKQIDEQLAKKADLIFTMSDGQTKWIEER-WPF---CRGKTFKL-------GHW---------I-
UniRef90 A0A450TYR9_1_143/1-143         GI---DI-T-AHRARQLTLGILRGFDVVMVMEHQHRQTIERS-MPI---ARGRVFTL-------GHW---------N-
UniRef90 E6WXB6 1 142/1-142             GI---DP-D-THKARQATPSVLAAADLILVMEHAHQARIARE-VPQ---VSGKTFLL-------GKW--------RG-
UniRef90 A0A4R3Z348 1 100/1-100         GS---SF-E-GHVAKQLTSEMCRSFDMILVMENTHIKAVCRF-VPE---ARAR-------------------------
UniRef90 F8GNS2 1 142/1-142             GL---DI-T-GHRAQQLSRAMTTRADLILVMDNAQRQDIQHL-HPA---TTGRVFRI-------GEL---------A-
UniRef90 F9TBG0 1 144/1-144             GL---DI-S-KHRAKQVTKEILAQYDLILVMDEYQKELLSNY-SAG---IRGKTFLI-------GQW--------IN-
UniRef90 A0A2E2K5G4 1 142/1-142         GL---DL-S-AHRARQIEAEMLYDHDLVIVMEPSQRDWLIQR-FPQ---DRAKVVLL-------SHW--------SG-
UniRef90 UPI000A07348B 1 142/1-142      GV---DI-H-AHRAQQLNNVLASQADLILVMDNEQKVEVQRR-HPS---TTGKVFRL-------GEP---------G-
UniRef90 A0A1V8M8J3 1 143/1-143         GL---DI-S-DYRATQLNQDMIHKADLILVMESVHKEQIELK-QPS---AKGKVFRL-------GHW---------G-
UniRef90 A0A1E7RCL3 3 140/1-138         GI---DL-S-NHRAKQITAEDVKQVDVVLTMSTQQQKHLEQL-FPF---SKGKVFRL-------GHW---------Q-
UniRef90 A0A7X7SEG5 2 142/1-141         DL---DI-S-KHRAQQVSEDLLRWADLVLVMETGQKNELLSK-FPW---LDGKIFRY-------GDPV----------
UniRef90 A0A1I7FB54 1 143/1-143         GL---DV-S-AHEARQVTSEILQGADLVLVMSDGQRLAVGQL-DPA---ALGKTMLL-------GRW--------LD-
UniRef90 UPI000DD06D87 3 126/1-124      NI---DM-R-THVAKQITASLIKQADLILVMSNNQAQHIEKQ-WPF---AKGKVFRL-------GHW---------Q-
UniRef90 A0A063B6R1 1 142/1-142         GI---DL-S-AHRAQQLTAALIRGADIVLTMEASQKRAIVER-YPF---ALGRVFPL-------DKQ---------C-
UniRef90 A0A3A8EB79 4 141/1-138         GI---SM-D-NHVARKLNADMIKHADLILVMSQNQQKHLEQT-WPF---AKGKTFRL-------GHW---------Q-
UniRef90 A0A6H2H7A5 3 147/1-145         GI---DI-S-AHRGRQLASWMLQDADLVLVMETNHQKQLSQQ-YPL---SRGKIRCL-------GEFDAPGDFGAAG-
UniRef90 A0A329D2G4 1 143/1-143         GI---DL-R-AHVATAVDPRIIRAADLVLTANTAQKQSIERA-FPF---ARGRIFRL-------REH---------A-
UniRef90 UPI0011822F9D 1 139/1-139      GI---DL-R-GHSAMALNLQVVREAQLVLTMTQAQREYIETA-YPF---ARGRVYRL-------CD------------
UniRef90 UPI0018873099 3 131/1-129      NA---NM-N-KHIAKQISEELTVKADLILTMSTSQSKWIEDR-WPH---CRGKTFRI-------GYW---------I-
UniRef90 A0A2U3I1P0 1 142/1-142         GI---AL-G-AHVATQLDAPLARAADLILTMTKPQREWVEAM-FPS---TRGKVFRL-------CDT---------F-
UniRef90 A0A095PZV8 1 141/1-141         GV---DL-S-AHRAQQLGRAHCASADLILVMDRGQRRLVEER-YPL---ARGKVFRI-------GEH---------E-
UniRef90 A0A248VX54 3 140/1-138         GL---DL-A-CHRGQQFSYEAHRDSDLILVMDEAQKALLEKN-HTP---LRGRVFRL-------GHF---------S-
UniRef90 A0A0M4KUN8 1 142/1-142         GH---EL-R-GHRARVVDAQLVDWADLVLVMDLEQRDEVERR-FPD---ARGRVYRI-------GEFI----------
UniRef90 UPI00158A8527 6 151/1-146      GF---DI-A-AHRSRRLPPALGTHADLILTMDLAQKRWLEQC-VPA---LRGRVFRLGLPDPAGGRP---------T-
UniRef90 UPI001365E3D5 8 150/1-143      GY---QI-DPNKRSQRLAVRDVDRADVILAMEMNHLRHLGEQ-FPT---TIGRSFLI-------GHW--------SG-
UniRef90 UPI00142E22A3 10 153/1-144     GA---YF---PHVARQMTREDARRYDHILVMDRENLAEVQRR-FPE---ARGKVRLF-------LDY--------LS-
UniRef90 A0A7X7UFU8 9 143/1-135         GL---DI-S-YHRSTPLSTDLIKRADLIITMEQRHITYITTM-VPE---ATGKTFRL-------TEW--------AGE
UniRef90 UPI001A8ED085 3 135/1-133      EL---DL-S-VHRSRRVSREEVAAADWVLPMTKAHELRLKAL-FPE---YEHKIVRL-------GEW--------SDH
UniRef90 A0A5C7SFX3 5 143/1-139         GI---DI-S-GQRARRLAPADFGRFDLILAMDQDNLKNLRKA-APA--EVLGKLHLF-------NAL--------AL-
UniRef90 A0A1Q7DZI5 1 132/1-132         GL---DL-S-GHRARLLTRELVEEADLILTMARHHRARVDEL-GGE-----GRVFVL-------GEY--------GGR
UniRef90 A0A7X7SJZ6 13 151/1-139        GV---DL-E-GHRSQPLTEELVRWADLILVMSPSHLDAVREL-DPE---ADNRTALL-------SEF--------AD-
UniRef90 UPI0015F50EC5 3 142/1-140      GY---SF-K-GIRARKVKSKDFENFDMVLAADRANLADLLDI-CPS--EHSDKVSLF-------LSH--------SN-
UniRef90 A0A3A4F8S2 13 114/1-102        GV---SM-S-GFKAREVTAQLVREADLILTMTRDHRSSIVQL-VPA---ALRR-------------------------
UniRef90 A0A1Q9UI09 2 98/1-97           GG---ES-D-GFVARRISARIAEDADIVLTMTARHRDEVLAI-APR--------------------------------
UniRef90 A0A1E4Y959 5 143/1-139         GI---DI-S-DQRARRIASADFGKFDLILAMDQDNLKNLRKS-APA--DALGKLHLF-------NAL--------AL-
UniRef90 L0FAJ2 4 130/1-127             GI---DI-A-SHQARAVSRDILASADWVIPMTQTHERQLKDA-FPE---WSSKVKRL-------GAW--------HS-
UniRef90 A0A3D8VHD6 7 149/1-143         GV---DI-A-GLRARQLSASDYRRFDWLLCADRANLRDVRAR-APA--DAHARIALL-------LDW--------AG-
UniRef90 A0A2R9UAR9 4 100/1-97          GL---AG-E-SHRARQISEKLVSDAGLILALSREHRRAIVSL-VPS---A----------------------------
UniRef90 A0A0G3M6Y9 2 142/1-141         GI---DI-S-SQRSRPINRRDFEIFDKIYCMDIDVFEDVISK-AKNE-AERQKVSLF-------LEV--------LG-
UniRef90 A0A2W4BH16 5 146/1-142         GI---DI-A-AQRARRITAADFDRFDLILALDRDNLRVLEKM-ASG--ANAAKLHLF-------NRY--------AL-
UniRef90 A0A6L4B6K6 3 144/1-142         GY---DL-S-RLRARKVQPKDYAEFDLLLAMDREHFEYLRDR-APE--GCRHKVRMF-------MSY--------AR-
UniRef90 A0A5Q0C8N2 5 154/1-150         GV---DL-S-NQRARRVTPDDIESFDIILGMDRTHVASLRRL-APE----ARNVFLF-------GDR--------AL-
UniRef90 UPI0005CA9E7B 20 161/1-142     GV---DL-S-ALRARRVDAEDIQRYDLVLAMDRDHLDFLRRL-APP--ALRDRVQLF-------LDY--------AP-
UniRef90 A0A1G9NZY5 8 154/1-147         GY---DG-S-VHRARQFTRDWFDRYDLVLAMDRDNVATLRRL-TSE----EVEVELF-------RSFDP------AAP
UniRef90 UPI000E675679 3 145/1-143      GY---DL-T-PLRARQVSAEDFDQFDYILVMDKSNLENIQRI-QSK--NSRAQVALF-------LSY--------AD-
UniRef90 A0A7Y6M605 4 150/1-147         GY---DG-S-AHRARQFTHDWFDRYDLVLAMDRDNLAVLRGM-APS----GVEVSLF-------RSFDP------AAP
UniRef90 A0A4R3GIM8_6_143/1-138         GI---DI-A-GQRARKILPEDFDDFDLILAMDYSNVVELRRR-APH----ANNIHLF-------GDF--------AL-
```



```
                                                              *                        *              **
UniRef90 UPI0016641B6A 11 157/1-147  GL---DF-S-GMKARQVREDDFEQFDLILAADNDNVHDLRER-CPA--QYRYKIHLI-------LDF--------TP-
UniRef90 J2B3B5 5 141/1-137          GI---DI-T-GQSARKILPEDFDDFDLILAMDRSNVAELRRR-SPL----GDNIHLF-------GDF--------AL-
UniRef90 UPI00178692C8 4 106/1-103   GI---DA-S-AHEARQINMAMIRRADLVLTMAREHRRVVVQH-VPA---AMRRSFT----------------------
UniRef90 UPI00047CF48D 6 146/1-141   GI---DI-A-GLRGRKIAVADFAAFDLILGMGGSNVRDLRRL-APP--AAAHKVHLF-------LHY--------AQ-
UniRef90 A0A6N6KD50 3 143/1-141      GI---DI-A-AQRGRQIRQEDLSEFDFIYAMDTSNFRDILKL-ANSP-EQEEKVKLI-------LNE--------TT-
UniRef90 A0A7C6GJB2 1 112/1-112      GI---DA-G-GHRSRRIHPQLLEEADLVLAMTRHHKEELLRL-APE---HAGKIFLL-------KEYA----------
UniRef90 A0A3G2TAX8 2 141/1-140      GI---DI-S-KQRSRPITTEDFEYFDRIYCMDLSVLEDVISL-AKND-EQRAKVSLF-------LEE--------AGI
UniRef90 A0A1I5A2J7 3 141/1-139      GI---DL-T-NQRARLFLVEDFEKFDHIYVMDQSNYKNVCRL-APNS-EAISKVQLI-------LNE--------IH-
UniRef90 UPI000DEF040C 5 143/1-139   GY---DL-S-RHRARQVRPGDFDGYDWVLAMDEDNLEALLSM-RRG---RGAPAELF-------LPW--------AG-
UniRef90 A0A1G8F8I5 6 146/1-141      GY---SF-A-GIRARQVSDEDFTRFDLILAADNDNLADLKAR-CPA--SEQHKLKLF-------MSY--------AS-
UniRef90 A0A0Q6X4Q3 4 144/1-141      NY---SL-DRKWRSRRVEPEQLGQYDLVLAMEAEHVDALRTKAAPE---HHAKILLL-------TDF--------VP-
UniRef90 A0A0C2S9U0 3 143/1-141      GY---EL-S-GLRARQVALDDFQRFDLILAMDHDNLAHLQAM-QPA--GSRAETDLF-------LRR--------YG-
UniRef90 UPI000A02BA3E 146 273/1-128 GL---SL-E-QHTATDISKYEPREDDLLLVMELRHAHRLVSRGIPP-----GQIALL-------GHW--------AR-
UniRef90 A0A7V5F301 2 136/1-135      GV---DI-S-GLRARQMNRADFARFDLIIGMDDANIADIERL-RPA--GDSTPVRLF-------TDY--------AP-
UniRef90 A0A7G8Q2W3 8 146/1-139      GY---DL-A-PHRARQARTADFEDFDWVLAMDQDNLETLEGM-RRG---RGVEPALF-------LPW--------AG-
UniRef90 A0A3D2KL57 755 859/1-105    GL---DL-S-GHRSRPLTKDLVSQARLILTMEKRHRDAVLEA-HPE---AADRVHT----------------------
UniRef90 UPI0010A80E05 4 101/1-98    GI---DA-S-AHRARQIDLAMIREADLIVGMSRDHRRLIVES-LPG---AM---------------------------
UniRef90 UPI00109EFB6D 3 144/1-142   GI---DL-S-AIRARQLDRADFERFDYLLAMDHGNLADLLAL-RPG--NCDAHVGLF-------LDF--------AG-
UniRef90 A0A7X7ME09 218 325/1-108    GL---SL-A-AHRSRPVTPDLVREAGLVVVMGRHHFERIIAC-SPA---ARDKVFLL-------GRF-----------
UniRef90 A0A7V3ZTT5 155 273/1-119    GI---DI-S-GHLAKSVSEDMIKEADLILCAEKNHYYYLTAL-FPE---CEDKLFLI-------KGY--------P--
UniRef90 UPI0018A58F51 5 145/1-141   GY---DL-S-GLRARQVELSDFSRFDLILAMDTSHMRVLERH-CPP--EHQAKIKLF-------LDF--------AN-
UniRef90 UPI0019131431 2 103/1-102   GV---DI-S-GYQGRQVKPADFRRFTHIIALDHDNLANLRRI-APQ--NE----------------------------
UniRef90 UPI000783D622 3 95/1-93     DI---DG-S-AHGGRLLTEALLRDSDLVITMTVEHRQLAVRT-GP---------------------------------
UniRef90 A0A1X9MH43 2 104/1-103      GI---DTAE-EHRAQHVTEELLEWADLILTMTHSHKQMLQTF-YPH---VSDYLFT----------------------
UniRef90 UPI0018EECC33 14 115/1-102  GG---DP-S-GFAARQLTQRITADADLILAMTTGHRDNVLEL-APH---RLNR-------------------------
UniRef90 A0A3C1ATE4 9 149/1-141      GL---EL-T-SF-ARQVRDADLSEYDLVLVMDRSNHQDIRAF-DREK-RHAAKVRLF-------CEF--------CT-
UniRef90 A0A523UXB5 211 329/1-119    GA---NI-T-GLVCKELSPEIITGADLILCMEQSQRDRVADM-VPD---AWEKTDLL-------KGY--------GKE
UniRef90 A0A3D3H0Y3 1 134/1-134      GY---SF-D-GIKARKVTVQDFTDFDLILAMDHDNIEELKKV-APA--DMQHKILLM-------LDF--------AA-
UniRef90 A0A353MKX6 3 111/1-109      GI---DA-R-QHRSRKIHPRLLAEADLILAMTEGHRQELLRL-GSE---HAGKIFLL-------KEY-----------
UniRef90 A0A1G6MAV1 10 113/1-104     GI---DH-R-SHRSQPITGELVEWADLILAMTVGHRQQLIQA-FPE---AADKLYT----------------------
UniRef90 A0A5F0DRH0 20 116/1-97      GG---RP-E-AHHARQLTARLVADADLVVTAAREHRSEVVSL-HPR---A----------------------------
UniRef90 UPI001581DB2B 11 111/1-101  AA---AA-A-GFAATQLTAGLITRQDVVLTMTREHRGAVLRL-CPE---ALR--------------------------
UniRef90 A0A1W6DP98 3 97/1-95        GL---EA-S-THTGRQLTAAMIEQADLVLTMSRAQRAAVVTL-VPQ--------------------------------
UniRef90 UPI00054F94F2 12 115/1-104  GG---DP-S-GFVARQIVEDMAIDADLTLTMTRAHRRRVLEL-APR---ALS--------------------------
UniRef90 A0A1Q8C2K1 21 131/1-111     GA---RV-N-AHVARQVDERMLAGADVVLTAERRHRAHVVEL-YPE---ALGRTF-----------------------
UniRef90 A0A084WG03 3 126/1-124      GL---S--S-DHVGRQICSDDFCRFDYIFGMDEWNMADLMEL-APVGEQCRAKIELL-------GNY--------RG-
UniRef90 A0A7W0GHV6 3 100/1-98       GG---SG-R-GFLARQISPEIVADADIVLAMTEQHRTTVLHN-SPR--------------------------------
UniRef90 A0A212TDC5 3 108/1-106      GG---DP-G-GFVSRQVSLPILRETDLVLTMERRHKAYVVEE-MPG---LVRRTFTL-------GHF--------LH-
UniRef90 B3TC92 6 126/1-121          EL---EI-A-AHRARPLTRRL--RADLVLALDRETEREARAL------KLRAPVEML-------GDH--------VG-
UniRef90 UPI0011444099 5 106/1-102   GG---SH-E-GFRARRLAGPMVEEADLVLTATRDHRAAVVRT-VPR---AMK--------------------------
UniRef90 A0A4Q6G032 3 109/1-107      GV---DI-S-DLRARQLRRSDFDYFDWLLCADRQNLRDVLAQ-APA--HSRGKVGL----------------------
UniRef90 UPI00078659EB 4 96/1-93     GI---DA-R-THRGRLLTEPIMRGSDLVIAMTLEHRQLAVQT-GPS--------------------------------
UniRef90 UPI0013597110 4 99/1-96     GVT--DA-A-AHVARRLDAASVRDADLVLALAREHRRAIVEL-VPG--------------------------------
UniRef90 UPI0018CA41A9 5 102/1-98    GA---HS-T-EHAARQLTEHLVGAADLILTATRDHRREVVTL-LPK---A----------------------------
UniRef90 UPI001423CF1F 10 122/1-113  HI---PS-E-DHVARQWRREWFAGRDLILALDVDHFGWLQEN-APDS-RALEKVRML-------RSF-----------
UniRef90 A0A4R8WM48 10 105/1-96      GA---TS-Q-LHAARQLTPALIESADLILTATREHRGEVVSL-HPR--------------------------------
UniRef90 A0A200HDT4 4 98/1-95        GI---AT-T-DLVPRQLTQSIARQADLILTATRDHRGQVLEV-FPA--------------------------------
UniRef90 A0A1D8GGS1 3 107/1-105      RI---DL-R-HHQARPLTKELIHAADLIFTMTRNHKQQVLNL-DPS---AEQKT------------------------
UniRef90 A0A3E0W0V8 19 115/1-97      RG---QP-G-EFRARQLRSAMVADADLVLTMTRAQRDEVIKR-NP---------------------------------
UniRef90 A0A172TIN9 2 105/1-104      QL---SD---RFSSNSISREWIEWADLILTMTASHKRTLIQM-HPY---AVDKTH-----------------------
UniRef90 UPI0009A67390 2 99/1-98     GL---SF---NHSSTPITQEIIDWSNLVLTMTDKHKQTLALQ-YPE---DHTKFF-----------------------
UniRef90 UPI0013D6B269 10 123/1-114  QL---FS-D-KHIARQWQPAWFRERDLILALDIDHYAWLSES-APDD-ESRNKIRML-------RSF-----------
UniRef90 UPI001109982C 10 120/1-111  GL---HP-R-THVARQFRPEWYAERDLILALDIDHFGHLREF-APE--GTEHKVRML-------RSF-----------
UniRef90 R0D1Y6 5 102/1-98           GV---DI-G-GLRARQVKPDDFRVFTHIYALDHDNLSGLRRI-APA--------------------------------
UniRef90 A0A511YZF2 8 106/1-99       GY---DV-P-AHVARRVRREDLAARDLLLPMTASHARALRAL-APD--------------------------------
UniRef90 A0A2S9ADU7 34 131/1-98      GASPDDV-V-THRARWLTEPIVAGADLVLAMSREHRTASVEL-APS---A----------------------------
UniRef90 UPI000A0620DD 2 101/1-100   GG---DP-S-GFKARRLKADMVHRADLVLTMTEKIRDQVREL-VPE---AL---------------------------
UniRef90 UPI00041B9DD5 3 106/1-104   KV---SH---SGSSRAVTSESIAWADLILTMTTGHKRHLLER-YPQ---AVDKAYTL---------------------
UniRef90 A0A7C5WBF8 4 95/1-92        GI---DI-S-GQTSDAIDEDLLRRMDLIITLCGHAEAACPAT-PPE---IR---------------------------
UniRef90 UPI0016693C71_67_166/1-100  GAKPADA-T-AHAARYLVEPYLLDADLVLAMAREHRSYAVKM-VPS--------------------------------
```

Fig. S17. Alignment of sequences related to *E. coli* (K-12) Wzb using the ConSurf server and default parameters (19). Residues are colored using BLOSUM62 scores. The red '*' indicate positions corresponding to D77 and R98; the cyan '*' indicate positions corresponding to H106 and W107 in *E. coli* (K-12) Wzb. Only the relevant segment of the alignment is shown. A total of 150 sequences were obtained from the non-redundant UniRef90 database (20) by use of the HMMER algorithm (E-value cutoff 0.0001) (21). The sequences were aligned using MAFFT-NS-i (22) with maximum and minimum cutoffs of 95% and 35%, respectively, in sequence identity.

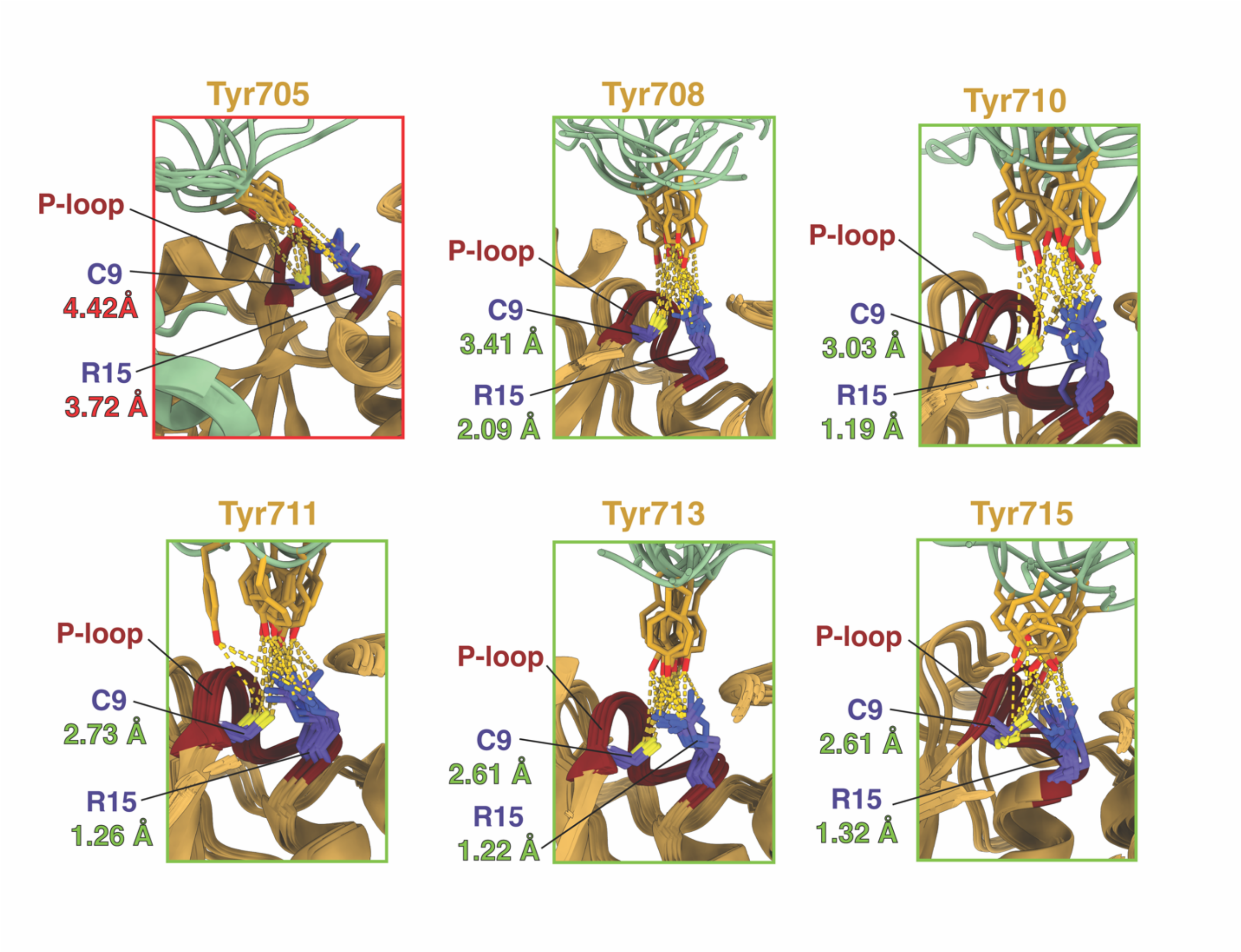


Fig. S18 The protein-protein interface in the Wzb•Wzc$_{\text{CD}\Delta\text{C}}$ complex is compatible with dephosphorylation of the Y-cluster. The distances of the -OH moieties for each of Tyr708, Tyr710, Tyr711, Tyr713 and Tyr715 with the thiol group of the catalytic C9 and the guanidino group of the conserved R15 of the phosphatase signature motif are within (or close to) their optimal distances of 3.3 Å and 2.6 Å, respectively. These distances are greatly increased for Tyr705. Thus, all tyrosine positions, except Tyr705, appear to be optimally accommodated at the active site without the introduction of strain or distortion at the binding interface suggesting that their dephosphorylation is compatible with the generated structural models.

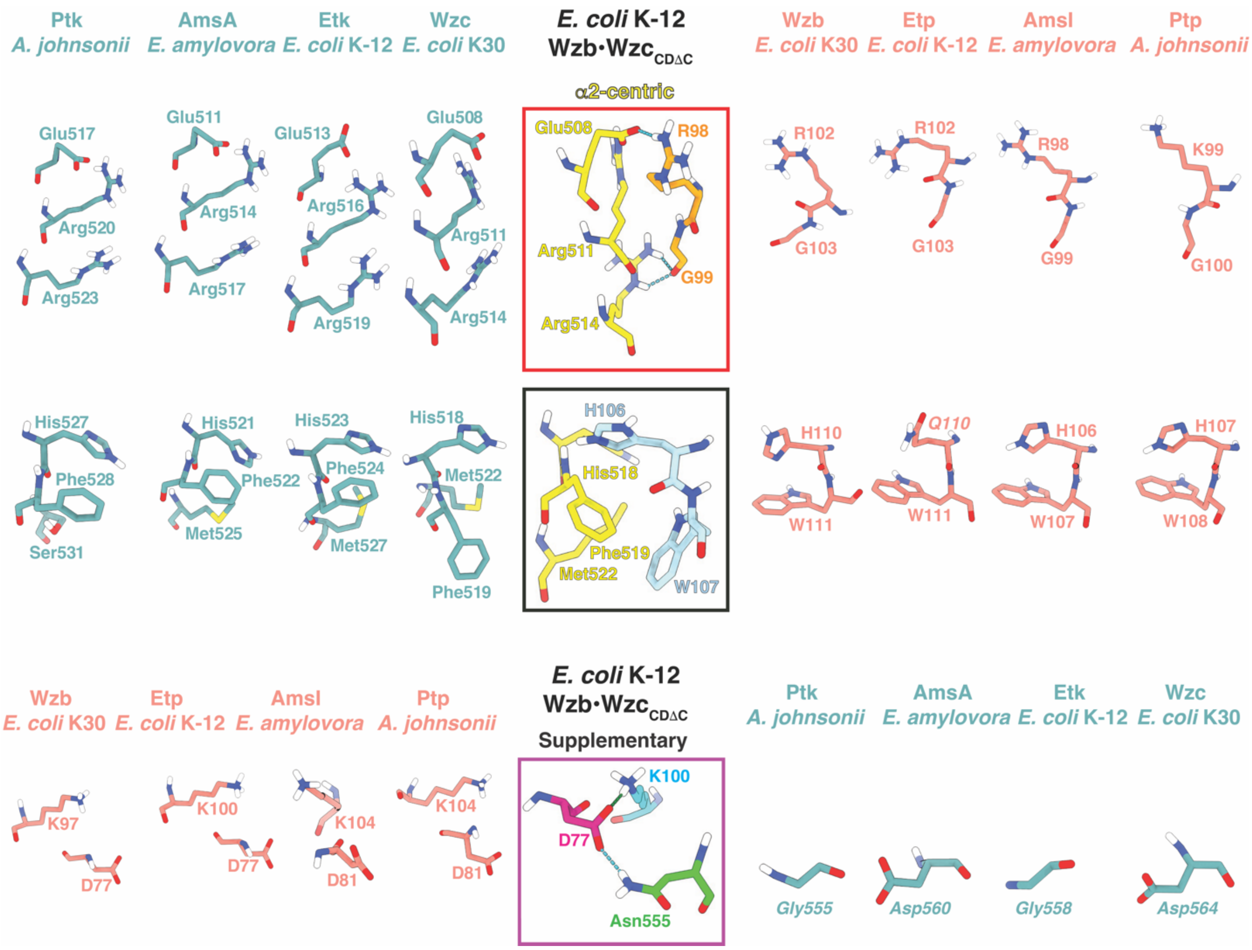


Fig. S19 Key interactions at the protein-protein interface for the Wzb•Wzc$_{CD\Delta C}$ complex compared with structurally similar cognate BY-kinase/LMW-PTP pairs from Gram-negative bacteria. Individual structures of the BY-kinase CDs and the LMW-PTPs from *E. coli* K30 [Wzc$_{K30}$ kinase PDB: 7NHS (17), Wzb$_{K30}$ phosphatase PDB: 2WJA (23)], *E. coli* K-12 [Etk kinase PDB: 3CIO (24), Etp phosphatase], *Erwinia amylovora* [AmsA kinase, AmsI phosphatase PDB:4D74 (25)] and *Acinetobacter johnsonii* [Ptk kinase, Ptp phosphatase] were overlaid on their counterparts within the Wzb•Wzc$_{CD\Delta C}$ complex without further optimization and residues at the putative interfaces were compared with the corresponding residues deemed to be key in stabilizing the Wzb•Wzc$_{CD\Delta C}$ complex (see Fig. 3). It is apparent that interactions centered at the kinase α2 helix are largely conserved (top two panels) while the supplementary interactions (bottom panels) appear to deviate. In the absence of experimental structures, models determined using AlphaFold (26) were used for Etp, AmsA and both *A. johnsonii* enzymes. Residues that differ from their *E. coli* K-12 counterparts are italicized.

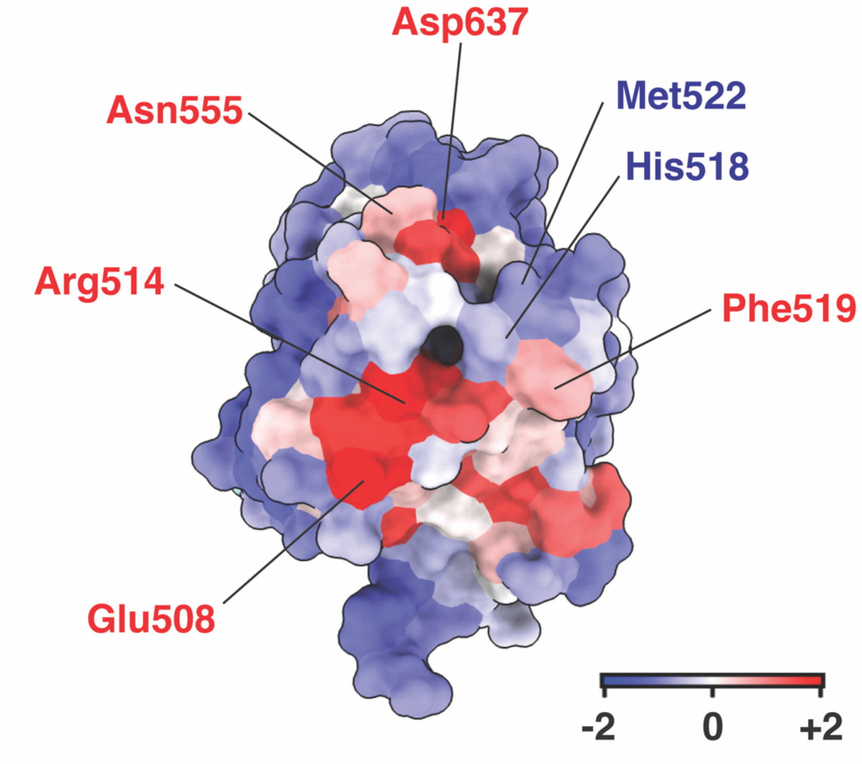


Fig. S20 Positional conservation mapped onto the structure of *E. coli* (K-12) $\mathrm{Wzc}_{\mathrm{CD\Delta C}}$ displaying the face that interacts with Wzb. Sequence conservation was obtained using the ConSurf server with default parameters (19). A total of 150 sequences were obtained from the non-redundant UniRef90 database (20) by use of the HMMER algorithm (E-value cutoff 0.0001) (21). The sequences were aligned using MAFFT-NS-i (22) with maximum and minimum cutoffs of 95% and 35%, respectively, in sequence identity. The surface is colored using AL2CO scores (27). The least and most conserved residues are shaded blue and red, respectively. Residues whose sidechains are identified as interacting with Wzb (see Fig. 3) are indicated, and the highly and poorly conserved residues are labeled in red and blue, respectively. The α2 helix residues, Glu508 and Arg514, that enable oligomerization (10), and recognize Wzb, are invariant. His518 and Met522 that take part in hydrophobic interactions with Wzb (see Fig. 3) appear poorly conserved. However, the position corresponding to His518 is a histidine in > 50% of the cases, and that corresponding to Met522 is almost always a hydrophobic residue (methionine or leucine in most cases).

Table S1. Templates and primers used in this study

Construct	Template; Primers	Reference
Wzc$_{\mathrm{CD}}$	pET15b; 5’-agctacatatgtcgttgtttaagctacatatgtcgttgttta-3’; 5’-cctagtggatccttatttcgcatccgacttat-3’	Temel et al. (6)
Wzc$_{\mathrm{CD\Delta C}}$	pET15b; 5’- actggccatatgtcgttgtttaatcgcgctggcgcgg-3’; 5’- attataggatccttacgcgctggcgcgg-3’	Temel et al. (6)
Wzb	pET15b; 5’- gggagggcatatgtttaacaacatcttagttgtctgtg-3; 5’-aggaaatggatccttatacct gctctgcgttcaat-3’	Temel et al. (6)
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser}}$	Wzc$_{\mathrm{CD\Delta C}}$; 5’-gtcaattggtaaaacctttgtctccgccaacctggc-3’; 5’-acgcgtgttgttgatcgactccgatatgcgca-3’	Temel et al. (6)
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser/Val466Cys}}$	Wzc$_{\mathrm{CD\Delta C,C544S/C563S}}$; 5’-agcgggatgctggcatagcagctaataccgtgttcttc-3’; 5’-gaagaacacggtattagctgctatgccagcatcccgct-3'	This study
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser/Leu505Cys}}$	Wzc$_{\mathrm{CD\Delta C,C544S/C563S}}$; 5’-gatggcttcaatcgcgcaatcggttggattccccaccg-3’; 5’-cggtggggaatccaaccgattgcgcgattgaagccatc-3'	This study
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser/Ser516Cys}}$	Wzc$_{\mathrm{CD\Delta C,C544S/C563S}}$; 5’-cgtagtctgcgtacctgtttgcacttcgcga-3’; 5’-tcgcgaagtgcaaacaggtacgcagactacg-3’	This study
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser/Gln523Cys}}$	Wzc$_{\mathrm{CD\Delta C,C544S/C563S}}$; 5’-agtttgcacttcgcgatgatgtgcgcacagaacaatgtgttgatg-3’; 5’-catcaacacattgttctgtgcgcacatcatcgcgaagtgcaaact-3'	This study
Wzc$_{\mathrm{CD\Delta C,Cys544Ser/Cys563Ser/Lys556Cys}}$	Wzc$_{\mathrm{CD\Delta C,C544S/C563S}}$; 5’-gatcaacaacacgcggcaattggtctggctgatcaccgcc-3’; 5’-ggcggtgatcagccagaccaattgccgcgtgttgttgatc-3'	This study
Wzb$_{\mathrm{D77A}}$	Wzb; 5’-ccatggtcaaaatcagggcgtagttgcgacacaga-3’; 5’-tctgtgtcgcaactacgccctgattttgaccatgg-3’	This study
Wzb$_{\mathrm{D77K/K100D}}$	Wzb; 5’-tttccatggtcaaaatcagcttgtagttgcgacacagacgg-3’; 5’-accaaacagcatcacatcgccgcgcatctccgg-3’	This study
Wzb$_{\mathrm{R70A/E96A}}$	Wzb; 5’-tgcgacacagacgggcgctgatttgacggg-3’; 5’-tttgccgcgcatcgccggtgccatctc-3’	This study
Wzb$_{\mathrm{R98A}}$	Wzb; 5’- tggcaccggagatggccggcaaagtgatgc -3’; 5’-gcatcactttgccggccatctccggtgcca-3’	This study

Table S2. Ambiguous interaction restraints used in the structure calculations

Protein	Active residues	Passive residues
Wzc$_{CD\Delta C}$	467, 511,512,514,518	460, 465, 466, 468, 469, 502, 503, 508, 509, 515, 516, 519, 522, 523, 552, 553, 554, 691
Wzb	2,4,74,75,77,84,100, 101, 137	1, 3, 30, 31, 32, 69, 70, 71, 72, 85, 87, 88, 91, 96, 98, 99, 102, 106, 107, 108, 109, 110, 111, 112, 132, 133, 136, 140, 141, 145.

Table S3. Unambiguous distance restraints from PRE measurements*

Residue	Sγ atom	Distance (Å)
R74	Val466Cys	18.3
I79	Val466Cys	19.7
T81	Val466Cys	19.0
K84	Val466Cys	18.6
R85	Val466Cys	17.4
H86	Val466Cys	17.9
I87	Val466Cys	15.8
E88	Val466Cys	14.1
R89	Val466Cys	14.7
C91	Val466Cys	15.5
E92	Val466Cys	14.5
M93	Val466Cys	16.1
A94	Val466Cys	16.7
E96	Val466Cys	17.1
M97	Val466Cys	19.8
R98	Val466Cys	17.2
G99	Val466Cys	15.0
K100	Val466Cys	17.4
V101	Val466Cys	16.2
M102	Val466Cys	21.1
L103	Val466Cys	19.2
D108	Val466Cys	19.5
N109	Val466Cys	13.2
E110	Val466Cys	18.7
I79	Ser516Cys	17.6
T81	Ser516Cys	16.8
K84	Ser516Cys	17.4
E88	Ser516Cys	17.9
C91	Ser516Cys	17.6
G99	Ser516Cys	16.3
V101	Ser516Cys	17.5
L103	Ser516Cys	17.3
F104	Ser516Cys	17.2

W107	Ser516Cys	17.5
A141	Ser516Cys	17.3
M103	Gln523Cys	19.3
F104	Gln523Cys	18.0
D108	Gln523Cys	18.1
N109	Gln523Cys	17.5
A135	Gln523Cys	17.9
Q140	Gln523Cys	18.8
A141	Gln523Cys	18.1
Q146	Gln523Cys	19.0
V147	Gln523Cys	19.9
N4	Lys556Cys	18.3
C73	Lys556Cys	16.7
R74	Lys556Cys	18.1
Y76	Lys556Cys	17.6
D77	Lys556Cys	18.6
M82	Lys556Cys	18.1
R89	Lys556Cys	18.1
E96	Lys556Cys	17.6
M97	Lys556Cys	16.9
R98	Lys556Cys	19.0
G99	Lys556Cys	16.2
K100	Lys556Cys	17.7
V101	Lys556Cys	17.9
N143	Lys556Cys	17.2
A144	Lys556Cys	17.7
E145	Lys556Cys	17.1
Q146	Lys556Cys	15.6
V147	Lys556Cys	16.2

*Each distance was modified by ± 4 Å to determine the upper and lower bounds.

Table S4. Statistics for the Wzb•Wzc$_{CD\Delta C}$ complex calculated using *all* restraints**

	Cluster$_{ALL,1}$	**Cluster$_{ALL,2}$**	**Cluster$_{ALL,3}$**	**Cluster$_{ALL,4}$**
HADDOCK score	-89.0 ± 1.6	-67.7 ± 5.5	-78.6 ± 5.7	-61.7 ± 4.5
Cluster size	212	85	65	19
RMSD from the lowest-energy structure (Å)	1.6 ± 1.0	2.5 ± 0.6	2.5 ± 0.3	1.6 ± 0.2
Van der Waals energy (kcal/mol)	-54.6 ± 1.1	-33.2 ± 7.2	-51.1 ± 6.1	-39.0 ± 6.1
Electrostatic energy (kcal/mol)	-135.6 ± 29.5	-153.5 ±17.7	-132.8 ± -17.3	-93.1 ± 46.6
Desolvation energy (kcal/mol)	-20.6 ± 7.1	-12.8 ± 4.1	-17.3 ± 6.1	-19.2 ±8.5
Restraint violation energy (AIRs) (kcal/mol)	133.2 ± 24.8	90.9 ± 53.7	164.0 ± 27.6	150.6 ± 29
Buried Surface Area (Å^2)	1664.8 ± 84.8	1264.0 ± 87.6	1716.3 ± 104.9	1325.2 ± 149

Table S5. Statistics for the Wzb•Wzc$_{CD\Delta C}$ complex calculated using *only* PREs**

	Cluster$_{PRE,1}$	**Cluster$_{PRE,2}$**	**Cluster$_{PRE,3}$**	**Cluster$_{PRE,4}$**
HADDOCK score	-72.6 ± 1.1	-60.2 ± 2.6	-63.6 ± 5.7	-48.6 ± 2.6
Cluster size	237	101	18	15
RMSD from the lowest-energy structure (Å)	2.0 ± 0.2	2.4 ± 0.2	1.7 ± 0.1	2.8 ± 0.2
Van der Waals energy (kcal/mol)	-38.6 ± 1.6	-23.4 ± 3.6	-29.5 ± 4.1	-16.0 ± 2.0
Electrostatic energy (kcal/mol)	-126.5 ± 14.8	-88.0 ± 35.3	-113.2 ± 5.8	-106.3 ± 10.9
Desolvation energy (kcal/mol)	-20.5 ± 1.9	-19.4 ± 5.9	-20.3 ± 1.4	-14.6 ± 3.2
Restraint violation energy (AIRs) (kcal/mol)	117.9 ± 29.4	2.6 ± 1.7	89.0 ± 23.8	31.9 ± 17.0
Buried Surface Area (Å^2)	1354.8 ± 114.8	1045.7 ± 56.3	1237.0 ± 87.7	838.8 ± 80.0

Table S6. Statistics for the Wzb•Wzc$_{CD\Delta C}$ complex calculated using *only AIRs**

	Cluster$_{AIR,1}$	**Cluster$_{AIR,2}$**	**Cluster$_{AIR,3}$**	**Cluster$_{AIR,4}$**
HADDOCK score	-119.2 ± 3.9	-120.6 ± 5.2	-123.9 ± 3.7	-128.4 ± 1.3
Cluster size	59	54	38	23
RMSD from the lowest-energy structure (Å)	2.0 ± 1.3	1.8 ± 1.1	1.5 ± 0.9	2.7 ± 1.6
Van der Waals energy (kcal/mol)	-56.1 ± 1.7	-57.6 ± 5.6	-61.2 ± 6.4	-62.4 ± 8.2
Electrostatic energy (kcal/mol)	-212.5 ± 28.22	214.9 ± 32.3	-240.32 ± 14,2	-273.3 ± 24.4
Desolvation energy (kcal/mol)	-22.8 ± 4.5	-25.4 ± 3.3	-20.2 ± 3.1	-17.6 ± 2.4
Restraint violation energy (AIRs) (kcal/mol)	21.7 ± 19.22	54.1 ±37.7	55.5 ± 32.1	63.2 ± 43.3
Buried Surface Area (Å^2)	1952.4 ± 120.9	1840.2 ± 79.2	2110.1 ± 80.7	2211.3 ± 79.3

**Statistics for the 4 lowest energy structures in each significant cluster are shown for all cases.

REFERENCES

1. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).
2. B. A. Johnson, From raw data to protein backbone chemical shifts using NMRFx processing and NMRViewJ analysis. *Meth. Mol. Biol.* **1688**, 257-310 (2018).
3. M. Gal, P. Schanda, B. Brutscher, L. Frydman, UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *J. Am. Chem. Soc.* **129**, 1372-1377 (2007).
4. C. A. Waudby, A. Ramos, L. D. Cabrita, J. Christodoulou, Two-Dimensional NMR lineshape analysis. *Sci. Rep.* **6**, 24826 (2016).
5. J. P. Loria, M. Rance, A. G. Palmer, A Relaxation-compensated Carr-Purcell-Meiboom-Gill Sequence for characterizing chemical exchange by NMR spectroscopy. *J. Am. Chem. Soc.* **121**, 2331-2332 (1999).
6. D. B. Temel *et al.*, Regulatory interactions between a bacterial tyrosine kinase and its cognate phosphatase. *J. Biol. Chem.* **288**, 15212-15228 (2013).
7. J. Garcia de la Torre, M. L. Huertas, B. Carrasco, HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.* **147**, 138-146 (2000).
8. J. L. Battiste, G. Wagner, Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39**, 5355-5365 (2000).
9. E. Lescop *et al.*, The solution structure of *Escherichia coli* Wzb reveals a novel substrate recognition mechanism of prokaryotic low molecular weight protein-tyrosine phosphatases. *J. Biol. Chem.* **281**, 19570-19577 (2006).
10. E. Bechet *et al.*, Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. *Mol. Microbiol.* **77**, 1315-1325 (2010).
11. F. Hajredini, A. Piserchio, R. Ghose, Long-range dynamic correlations regulate the catalytic activity of the bacterial tyrosine kinase Wzc. *Sci. Adv.* **6**, eabd3718 (2020).
12. C. Kutzner *et al.*, Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.* **36**, 1990-2008 (2015).
13. C. Dominguez, R. Boelens, A. M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737 (2003).
14. N. Eswar, D. Eramian, B. Webb, M. Y. Shen, A. Sali, Protein structure modeling with MODELLER. *Meth. Mol. Biol.* **426**, 145-159 (2008).
15. S. Wang *et al.*, Crystal structures of a low-molecular weight protein tyrosine phosphatase from *Saccharomyces cerevisiae* and its complex with the substrate p-nitrophenyl phosphate. *Biochemistry* **39**, 1903-1914 (2000).
16. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).
17. Y. Yang *et al.*, The molecular basis of regulation of bacterial capsule assembly by Wzc. *Nat. Commun.* **12**, 4349 (2021).
18. D. Barford, Z. Jia, N. K. Tonks, Protein tyrosine phosphatases take off. *Nat. Struct. Biol.* **2**, 1043-1053 (1995).

19. H. Ashkenazy *et al.*, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350 (2016).
20. B. E. Suzek *et al.*, UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926-932 (2015).
21. M. Wistrand, E. L. Sonnhammer, Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. *BMC Bioinformatics* **6**, 99 (2005).
22. K. Katoh, D. M. Standley, MAFFT: iterative refinement and additional methods. *Meth. Mol. Biol.* **1079**, 131-146 (2014).
23. G. Hagelueken, H. Huang, I. L. Mainprize, C. Whitfield, J. H. Naismith, Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly. *J. Mol. Biol.* **392**, 678-688 (2009).
24. D. C. Lee, J. Zheng, Y. M. She, Z. Jia, Structure of *Escherichia coli* tyrosine kinase Etk reveals a novel activation mechanism. *EMBO J.* **27**, 1758-1766 (2008).
25. S. Benini, L. Caputi, M. Cianci, Cloning, purification, crystallization and 1.57 Å resolution X-ray data analysis of AmsI, the tyrosine phosphatase controlling amylovoran biosynthesis in the plant pathogen *Erwinia amylovora*. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 1693-1696 (2014).
26. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
27. J. Pei, N. V. Grishin, AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700-712 (2001).