Long-range dynamic correlations regulate the catalytic activity of the bacterial tyrosine kinase, Wzc

Fatlum Hajredini[1,2], 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.

*Correspondence: rghose@ccny.cuny.edu

Teaser: A mechanism for nucleotide exchange and activation is proposed for a family of protein tyrosine kinases unique to bacteria.

ABSTRACT

BY-kinases represent a highly conserved family of protein tyrosine kinases unique to bacteria without eukaryotic orthologs. BY-kinases are regulated by oligomerization-enabled trans-phosphorylation on a C-terminal tyrosine cluster through a process with sparse mechanistic detail. Using the catalytic domain (CD) of the archetypal BY-kinase, *Escherichia coli* Wzc, and employing enhanced-sampling MD simulations and supporting ITC and NMR measurements, we propose a mechanism for its activation and nucleotide exchange. We find that the monomeric Wzc CD preferentially populates states characterized by distortions at its oligomerization interfaces and by catalytic element conformations that allow high-affinity interactions with ADP but not with ATP•Mg^{2+}. We propose that oligomer formation stabilizes the inter-monomer interfaces and results in catalytic element conformations suitable for optimally engaging ATP•Mg^{2+}, facilitating exchange with bound ADP. This sequence of events, oligomerization i.e. substrate binding, prior to engaging ATP•Mg^{2+}, facilitates optimal auto-phosphorylation by preventing a futile cycle of ATP hydrolysis.

INTRODUCTION

BY-kinases (*1, 2*) comprise a unique family of protein tyrosine kinases (PTKs) that are highly conserved in both Gram-negative and Gram-positive bacteria (*3*). BY-kinase activity has been shown to be critical in many physiological processes, most notably in the synthesis and export of polysaccharides (*4*) responsible for biofilm and capsule formation. In Gram-negative bacteria, BY-kinases are localized to the inner membrane and comprise of a periplasmic domain (PD), a two-pass membrane domain with two transmembrane helices (TM1,2) and a cytoplasmic catalytic domain (CD), all encoded on a single polypeptide chain (fig. S1A). BY-kinases function as part of a multi-protein complex that spans both the inner and outer membranes (*5*). BY-kinases have no known eukaryotic orthologs and the structures of the CDs of the *Escherichia coli* BY-kinases, Etk (Etk_{CD}) (*6*) and Wzc (Wzc_{CD}) (*7*), together with that of *Staphylococcus aureus* CapB ($CapB_{CD}$) (*8*) illustrate their striking differences with eukaryotic PTKs (ePTKs) marked by the absence of conserved sequence motifs (*9*) and the classic two-lobed fold that are the signatures of the latter. BY-kinase CDs closely resemble P-loop ATPases, notably those of the MinD family (*10*), suggesting evolution from a common ancestor (*1*) in orthogonal fashion to ePTKs (*11*). Like the P-loop ATPases, the BY-kinase CDs contain NTP-binding Walker-A (A/GX_5GK[S/T], X is any residue) and Walker-B (ϕ_4DX_2P, ϕ is a hydrophobic residue) motifs (fig. S1B), together with an additional Walker-B-like motif, called Walker-A' (ϕ_4DxDxR) (*3*). Additionally, BY-kinases contain a loop that is rich in arginine and lysine residues (RK-cluster) that has been shown to be important for auto-phosphorylation (*7*). A defining feature of BY-kinases is a C-terminal cluster of 5-7 tyrosine residues (Y-cluster) whose intermolecular auto-phosphorylation (trans-phosphorylation) drives function (*12*). The number and precise location of individual tyrosine residues in the Y-cluster are poorly conserved and the removal of any individual tyrosine has no

significant effect on function. However, replacement of multiple tyrosines from the Y-cluster has deleterious functional consequences (*4, 13*), suggesting that the overall level of Y-cluster phosphorylation rather than the phosphorylation state of any particular tyrosine residue/s therein relates to function contrasting many ePTKs e.g. members of the Src-family (*14*).

As in the case of eukaryotic PTKs, a necessary condition for intermolecular auto-phosphorylation in BY-kinases is the formation of CD oligomers (fig. S1C). Trans-phosphorylation between subunits is then achieved by the insertion of a Y-cluster tyrosine of the substrate-acting (S-acting) CD into the catalytic site of an adjoining enzyme-acting (E-acting) CD (fig. S1B and fig. S1C). Indeed, $CapB_{CD}$ (*8*) and Wzc_{CD} (*7*) both crystallize as front-to-back octamers (fig. S1B). Biochemical studies suggest that two specific features are required for oligomer formation by the isolated BY-kinase CD – (1) a dephosphorylated Y-cluster (or that containing very low levels of phosphorylation) and (2) a conserved EX_2RX_2R oligomerization (fig. S1B) motif housed on a helical segment ($\alpha 2$) seen in the structures of the $CapB_{CD}$ (*8*) and Wzc_{CD} (*7*). While the high/low phosphorylation levels of the Y-cluster represent the primary drivers of the oligomerization state of the CD in solution, and likely in the membrane-bound full-length enzyme (fig. S2), the importance of the oligomerization motif is underscored by the fact that replacement of the conserved residues of this motif by alanine (E508A/R511A/R514A) in Wzc_{CD} prevents oligomer formation and leads to diminished levels of Y-cluster phosphorylation in the isolated CD, while almost completely abolishing it in the context of the full-length enzyme. *E. coli* K12 cells expressing the mutant protein are functionally compromised and produce altered levels of the exopolysaccharide colanic acid (*7*).

While crystallographic snapshots of BY-kinase CDs and supporting biochemical studies have provided significant insight into the properties of these unique enzymes (*1, 8*), the precise

mechanism/s of activation and regulation of BY-kinases are still poorly understood. A key open question is about the nature of the coupling between oligomerization and nucleotide exchange. The former, to bring the S-acting and E-acting CDs in close proximity to enable trans-phosphorylation (fig. S1C), and the latter, to engage ATP prior to chemistry and to release the ADP product after the chemical step to prime the catalytic site for the next phosphorylation event. Given their evolutionary origin, BY-kinases are, not surprisingly, highly efficient ATPases (*15*). Thus, the binding of ATP to a monomeric CD prior to oligomerization, i.e. substrate binding, is likely to result in a futile cycle of hydrolysis. Oligomerization, i.e. formation of the complex between the E-acting and S-acting CDs, would have to precede ATP binding to enable efficient auto-phosphorylation. One can then reasonably expect that this process is mediated by an allosteric pathway within the kinase core linking the oligomerization interfaces (involving the α2 helix and the α7/α9 helices on the complimentary surface; fig. S1B) to the catalytic site. In order to test this hypothesis, we utilized replica exchange molecular dynamics (REMD) validated by isothermal titration calorimetry (ITC) and solution NMR measurements using a construct ($Wzc_{CD\Delta C}$) (*16*) of *E. coli* (K12) Wzc_{CD} that lacks its highly dynamic C-terminal tail and represents its structural core. Our results illustrate the coupling between the catalytic elements and the oligomerization interfaces and help define a mechanism for oligomerization and ATP binding to enable efficient auto-phosphorylation in BY-kinases.

RESULTS

Monomeric apo-$Wzc_{CD\Delta C}$ assumes an open state with key catalytic elements in inactive conformations

We performed Replica Exchange with Solute Tempering (REST2) (*17*) simulations on the unliganded state of $Wzc_{CD\Delta C}$ (see Materials and Methods for details). An initial visual inspection

of the structures in the REST2-generated ensemble suggested the presence of two dominant global fluctuation modes. These include: (1) a downward displacement of α4, loop $L_{\alpha 5}$, together with additional elements connecting it to α4, with respect to α3, and (2) coupled outward rotations of α2 and α3 with respect to the protein core (indicated by the red arrows on the left panel of Fig. 1A). These global modes can be efficiently analyzed in a cylindrical coordinate system defined by a rise ($|h|$) and a polar angle (θ) (Fig. 1A, right panel; also see corresponding section in the Materials and Methods for additional details). $|h|$ defines the extent of the downwards displacement of the center of mass of α4. The polar angle, θ, measures the degree of outward rotation of α2 and α3 with respect to the protein core. The probability density function (PDF) of the sampled conformations projected onto this coordinate system reveals a single dominant state with average $|h|$ and θ values of ~4.5 Å, ~145°, respectively (Fig. 1B). This state is significantly more open (referred to as an open state, OS, from here on forward) compared to that represented by the crystal structure for which $|h|$ and θ take values of 3 Å and 131° (indicated by the green dot in Fig. 1B), respectively.

Analyses of representative structures drawn from the OS suggests significant evolution of the conformations of key catalytic elements of $Wzc_{CD\Delta C}$ with respect to the crystal structure (Fig. 1C), in addition to the differences in the global conformation mentioned earlier. In the crystal structure of Wzc_{CD} (*7*), the sidechains of T541 (Walker-A) and D642 (Walker-B) form a hydrogen-bond (Fig. 1D) generating an orientation that is critical for the proper coordination of Mg^{2+}. Further, the catalytic lysine on Walker-A (K540, a methionine in the crystal structure) adopts an orientation that allows it to contact the β- and γ-phosphates of ATP in a fashion similar to that seen in the structure of the homologous P-loop ATPase, MinD (*10*) (Fig. 1D). This conformation of the catalytic lysine represents a key structural feature in facilitating the bond breaking step during ATP

hydrolysis in typical ATPases (*18-21*). However, for the OS seen in our simulations, an outward rotation of α3 alters these sets of interactions (Fig. 1C) in a manner reminiscent of the structure of an inactive mutant (K15M) of shikimate kinase (SK) (*22*), a member of the P-loop kinase family (*23*) (Fig. 1E). Based on these observations, we suspect that the OS seen in our simulations represents an inactive state of $Wzc_{CD\Delta C}$ that is incompatible with Mg^{2+}, and consequently, with ATP coordination. Further, the rotation of α2 (that houses the conserved EX_2RX_2R motif) alters the orientation of the first oligomerization interface (I_1) resulting in a more extended geometry in the OS relative to that in the crystal structure making it incapable of forming the closed octameric ring seen *in crystallo* (*7*) (fig. S3).

The complex of monomeric $Wzc_{CD\Delta C}$ with ATP•Mg^{2+} is incompatible with progress toward efficient trans-phosphorylation

Next, we performed REST2 simulations on the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex (see Materials and Methods for details). The PDF of the resulting conformations projected onto the cylindrical coordinate frame described above produce three distinct ellipsoidal clusters that lie along the diagonal of the coordinate system (Fig. 2A) reflecting the tendency of the system to fluctuate (indicated by the magenta arrows in Fig. 2A) between an OS (~28%; similar to the one described above), a closed state (CS, ~45%; with similarities to the crystal structure), and a state that is significantly more closed compared to the crystal structure ($|h|$ and θ values of ~1.9 Å, ~128°, respectively) that we term a hyper-compact state (HS, ~11%) (Fig. 2A). Inspection of representative structures reveals that, as in our apo simulations, the OS displays the proximity between the sidechains K540 and D642 with the latter being disconnected from T541 (Fig. 2B). As a consequence, the adenosine moiety of ATP is displaced from its binding pocket; additional interactions critical for stabilizing the nucleotide-bound state e.g. those involving D480 of the RK-

cluster and Y569 of α4, are also broken (*7*). An inspection of 10 randomly selected structures from either the HS (Fig. 2C, left panel) or the OS (Fig. 2C, right panel) shows that the RK-cluster is highly ordered in the former and highly disordered in the latter. However, while the second oligomerization interface (I_2) is largely intact in the OS (though the extended structure and the resulting inability to form the octameric ring, discussed above, persist), helices α7 and α9, that comprise it, are partially unfolded in the HS.

While it is relatively straightforward to analyze local features in the structures that comprise the OS and HS, conventional analysis by overlay and visual inspection proved difficult for the CS, that is the major state for the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex, due to a high degree of local structural heterogeneity. In order to facilitate the detailed structural analysis of the CS we developed a novel method we term Environmental Variability Analysis Coupled with Mean Shift (EVA-MS, described in detail in the Supplementary Materials). The EVA-MS approach identifies structural evolution in different parts of the protein and their correlations, if any. Briefly, the method determines the variation in the local environment of each residue (as defined by its local structural contacts) by defining a measure of its variance $V_i(t)$ relative to a reference structure (fig. S4A; eq. S3). By identifying a subset of the most highly varying residues across the structure and across all structures in the cluster (using a residue-specific normalized measure of variability across the ensemble, $V_{i,norm}$, see eq. S6 in the Supplementary Materials), a multi-dimensional PDF can be defined as a function of the $V_i(t)$ values of the most highly varying residues. Specific conformational modes within this PDF are then identified using the unsupervised Mean Shift algorithm (see fig. S4 for a simple illustration of the approach). For the CS, the most varying residues were found to be D480 (αB of the RK-cluster), S512 (α2), F519 (α2), S617 (α7) and G656 (α9) (fig. S5A). Application of EVA-MS to the 5-dimensional PDF comprising the $V_i(t)$

values for these residues yields several clusters of which four (CS1, CS2, CS4 and CS5) account for ~94% of the structures within the CS (see fig. S5B). Of these, CS2, that accounts for two-thirds of the CS, is characterized by the formation of a bend at the center of α2 (I_1) (Fig. 2D and fig. S5E). In the second most populated cluster (CS5), while α2 is largely intact, α7 and α9, that comprise the second oligomerization interface (I_2) are partially unfolded (Fig. 2D and fig. S5E). Taken together, these results suggest that the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} ensemble consists of structures that display features that are not compatible with efficient trans-phosphorylation (we refer to these as "distortions" for simplicity). The features include: (1) a significant bend in α2 thereby destabilizing I_1 (in CS2), (2) partial unfolding of α7 and α9 thereby destabilizing I_2 (in CS5) or (3) the displacement of ATP from its binding pocket (in the OS). These results suggest that in the monomeric state, $Wzc_{CD\Delta C}$ cannot simultaneously engage ATP•Mg^{2+} and present a stable oligomerization interface. The destabilization of I_1 and I_2 comes with the loss of several hydrogen bonds with a significant associated energetic cost. It is thus reasonable to expect that oligomerization, that provides a means to stabilize the I_1/I_2 interface, may be necessary for stable ATP•Mg^{2+} binding.

Characteristics of the hyper-compact state of the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex suggests a catalytic role for the RK-cluster

The HS (Fig. 2A), that is sampled the least of the three major states in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations, shows some remarkable structural features. Most notably, the RK-cluster, that is disordered in the crystal structure (*7*) with the exception of a small helical region (αB), shows a high degree of order and makes extensive contacts with ATP (Fig. 3A). The aliphatic segment of the K492 sidechain stacks against Y569 (that contacts the α-phosphate of ATP) and makes a salt bridge with E572 on α4 (Fig. 3A). The Walker-A K540 and the Walker-B D642 are

in their appropriate orientations for catalysis. Additionally, R490 is inserted into the active site where it contacts the γ-phosphate of ATP. Indeed, the coordination of ATP is very similar to the active, i.e. ATP hydrolysis capable states of MinD (*10*) (Fig. 3B) and F_1-ATPase (*24*) (Fig. 3C). This observation appears to suggest a critical role for the RK-cluster residue, R490. MinD has been shown to dimerize along the active site where its so-called deviant Walker-A lysine inserts into the active site of the neighboring protomer and contacts the γ-phosphate of ATP, thus activating hydrolysis (Fig. 3B) (*10*). The arginine-finger (Fig. 3C) plays a similar role in the context of the F_1-ATPase oligomer. It has now been established that the presence of a positively charged residue proximal to the γ-phosphate of ATP that serves to stabilize the negative charge build-up during nucleophilic attack by water (*18*) is a key initiator of hydrolysis in ATPases (*20, 21*). In the HS of $Wzc_{CD\Delta C}$, R490 appears to adopt a conformation that is similar to the arginine-finger in many GTPases and in AAA^+-family of ATPases (*25*) and the signature lysine of the deviant Walker-A in MinD (*26*). This suggests that R490 likely plays a similar role in activating ATP hydrolysis in Wzc. Indeed, a R490A variant of Wzc_{CD} shows drastically reduced levels of auto-phosphorylation without significant reduction in nucleotide affinity (*7*).

Monomeric $Wzc_{CD\Delta C}$ is hindered in its ability to efficiently co-ordinate Mg^{2+}

In the apo simulations discussed above, we noted that the coupled outward rotation of α2 and α3 to form the OS leads to a disengagement of T541 and D642 thereby disrupting the ability to coordinate Mg^{2+}. The displacement in α3 appears to represent a response to a dynamic coupling across the oligomerization interface involving α2 (I_1), and α7/α9 (I_2). The presence of Mg^{2+} in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations partially counteracts this effect by reinforcing the T541-D642 interaction, populating the CS that has alternative modes of destabilization e.g. through distortions in I_1 or I_2 (as in CS2 or CS5 illustrated in Fig. 2D). It is therefore reasonable to expect that this

effect would be modulated by the presence or the absence of Mg^{2+}. To test this hypothesis, we performed two additional sets of REST2 simulations on the $Wzc_{CD\Delta C}$•ADP and $Wzc_{CD\Delta C}$•ADP•Mg^{2+} complexes using similar protocols as before (see Materials and Methods for details).

For the $Wzc_{CD\Delta C}$•ADP•Mg^{2+} simulations, the system samples both the CS and OS with frequencies of ~60% and ~34% respectively (Fig. 4A, left panel). Application of the EVA-MS procedure on the CS identified the following residues with the highest degree of variation: Q475 and D480 (RK-cluster), S512 (α2), K567 ($L_{\alpha4}$), and S617 (α7) (fig. S6A). A majority of the resulting structures can be accounted for by six clusters (fig. S6B). Similar to the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations, the most populated clusters within the CS were found to be characterized by a bending of α2 (I_1) (CS1 in fig. S6D and fig. S6E), or an unfolding of α7 (I_2, CS2 in fig. S6D and fig. S6E). Additionally, the CS also samples structures where α2 is bent, α7 is unfolded, the RK-cluster is disordered, and the adenosine ring of ADP is displaced from its binding pocket (CS3 in fig. S6D and fig.S6E). These results suggest that the presence of Mg^{2+} leads to a CS in which the contact between Y569 of α4 and the α-phosphate of ADP persists (Fig. 4A, middle panel). However, there are significant distortions in I_1 and I_2 together with enhanced disorder within the RK-cluster and in some cases, breakage of key contacts between D480 and the ribose O3' and O2', resulting in the displacement of ADP from its binding pocket. Alternatively, both Y569 and D480 are displaced from ADP and the system assumes the OS (fig. S6E and the right panel in Fig. 4A). Taken together, the system evolves either by distortion of the interaction interfaces or by breaking key contacts with the nucleotide (ADP in this case) in similar fashion to the ATP•Mg^{2+} simulations described above.

In contrast to the $Wzc_{CD\Delta C}$•ADP•Mg^{2+} simulations, where discrete, well-defined states are seen, projection of the $Wzc_{CD\Delta C}$•ADP ensemble onto the cylindrical coordinate frame suggests the sampling of a relatively diffuse range of states (Fig. 4B, left panel). In the absence of well-defined boundaries between states, we performed EVA-MS analysis on the entire ensemble to define local features in the constituent structures (fig. S7A). The most highly varying residues in this simulation were found to be D480 (αB of the RK-cluster), S512 (α2), Y569 (α4), I592 ($L_{\alpha5}$), and E618 (α7) resulting in six major clusters within the ensemble (fig. S7B). The three most populated clusters, C1, C2 and C4, are all characterized by the evolution on S512 (fig. S7D) in the form of bending of α2 (fig. S7E and the middle panel of Fig. 4B), and differ slightly in the evolution of D480 (RK-cluster), Y569 (α4) and I592 ($L_{\alpha5}$) for each specific case. Given the very small differences observed in the projection of these clusters onto their constituent residues (fig. S7D), we wondered whether each cluster could represent different conformations of αB (RK-cluster), α4 and $L_{\alpha4}$, given that the latter two elements adopt distinct configurations in the CS and the OS described above. 500 structures were randomly selected from each of the C1, C2 and C4 clusters and projected onto the cylindrical coordinate system. These clusters sorted into distinct regions, populating states that, on average, varied from being somewhat closed (in black in fig. S7E; black part of the arrow in Fig. 4B) with occurrence of ~21%, to more open (blue in fig. S7E; blue part of the arrow in Fig. 4B) with an occurrence of ~23%, to a state that resembles the OS (red in fig. S7E; red part of the arrow in Fig. 4B) with an occurrence of ~18%. A comparison of representative structures from each cluster aligned on α3 shows an almost perfect alignment on α2 and α3 (Fig. 4B, middle panel), differing only in the degree of displacement of Y569 relative to ADP, with the difference being a little more pronounced between C1/C4 (black/blue) and C2 (red). Additionally, α4 and $L_{\alpha5}$ show different degrees of displacement with respect to α3 (illustrated in the right panel

of Fig. 4B). Interestingly this displacement appears to be coupled to larger, more global displacements across α7, α8, and α9 (Fig. 4B, right panel). Note that despite the relatively small variations in $V_i(t)$ values across the clusters (fig. S7D), the MS procedure is able to distinguish these clusters from each other. These clusters are characterized by multiple small local variations together with large global rearrangements that are reflected in the $|h|$-θ space. This shows that a subset of residues which show small variations in their local environment but together give rise to larger global rearrangements can be parsed using the EVA-MS approach. A close-up view of the active site in the partially closed configuration, C1 shows that, despite a higher degree of closure, the active site conformation resembles that in the OS (defined by the presence of a D642-K540 contact, Fig. 4C, left panel). This contrasts the CS seen for $Wzc_{CD\Delta C}$•ADP•Mg^{2+} in which T541 is proximal to D642 (Fig. 4C, right panel). Taken together, these data suggest that in the absence of Mg^{2+}, a stable interaction with ADP can occur despite the fact that the T541-D642 interaction is broken. This stable engagement of ADP is enabled by ability of the α4 and $L_{\alpha 5}$ elements to freely sample the open and partially closed states with similar probability without perturbing key ADP-coordinating residues. On the other hand, the presence of Mg^{2+} appears to force the formation of the CS thereby disturbing the orientation of the ADP-coordinating residues, (notably D480, also see CS3 in fig. S6D and fig. S6E) and occasionally displacing the nucleotide from its binding pocket (Fig. 4D). Overall, these findings support our model where formation of OS lies in direct opposition to Mg^{2+} binding, and presence of the divalent cation disrupts nucleotide coordination and destabilizes the oligomerization interface.

Experimental measurements validate *in silico* predictions

The simulations discussed above, suggest that the complex of monomeric $Wzc_{CD\Delta C}$ with ADP is stable in the absence of Mg^{2+} and the divalent cation appears to destabilize it. Additionally,

the ATP•Mg^{2+} complex of monomeric $Wzc_{CD\Delta C}$ appears to display various distortions, suggesting that binding of ATP•Mg^{2+} in the monomeric state carries with it an energetic cost. Given these observations and the fact that $Wzc_{CD\Delta C}$ is monomeric in solution (*16*), we would expect that its affinity for ATP•Mg^{2+} would be significantly lower than that for ADP, with the affinity of the latter also being reduced in the presence of Mg^{2+}. Indeed, the binding affinities measured using Isothermal Titration Calorimetry (ITC) indicate the following K_D values (averaged over 3 independent measurements); AMPPCP•Mg^{2+}: 163±23 μM, ADP•Mg^{2+}: 2.56±1.08 μM and ADP in the absence of Mg^{2+}: 0.70±0.38 μM (see Fig. 5A) in line with our predicted trends. This observation is relevant in the cellular context, since ATP is necessarily bound to Mg^{2+} while ADP may not be given their relative affinities for Mg^{2+} (*27*).

As shown above, our simulations suggest that apo-$Wzc_{CD\Delta C}$ exists in the OS (Fig. 1A) and ADP-bound $Wzc_{CD\Delta C}$ is capable of populating the OS where ADP is properly coordinated (Fig. 4B). However, in such cases, α2 (I_1) is bent, and the second interaction interface (I_2) together with α4 and $L_{\alpha 5}$ are free to undergo rigid-body oscillations occasionally disrupting the contact between Y569 and the α-phosphate of ADP. Given that ~80% of NMR assignments are available for the amide ^{15}N, ^{1}H resonances of $Wzc_{CD\Delta C}$ (*16*), we felt that these resonances were likely to sense the predicted ADP-induced changes and serve to validate the computational results. ^{15}N, ^{1}H TROSY spectra of ^{15}N,^{2}H-labeled $Wzc_{CD\Delta C}$ in the presence of sub-stoichiometric amounts of ADP show that the resonances of $Wzc_{CD\Delta C}$ in the ADP-bound and nucleotide free states are in the slow-exchange regime, a result that is not unexpected given the sub-μM affinity. Significant chemical shift differences between resonances corresponding to free and ADP-bound states are seen across the two interaction interfaces (I_1 and I_2) (top panels of Fig. 5B and Fig. 5C). S512, that forms the hinge in the bending of α2, shows a difference of 0.06 ppm between the free and ADP-bound

states. Strong perturbations are also seen on α7 (E618: 0.07 ppm, L619:0.05 ppm and L620: 0.07 ppm), α9 (A652: 0.12 ppm). Additional perturbations are also seen on α4, $L_{\alpha 5}$ and on α8 that contacts $L_{\alpha 5}$. This analysis suggests the presence of a long-range network, predicted in our simulations, that couples the catalytic site to remote regions, most notably the oligomerization interfaces I_1 and I_2.

As noted earlier, our simulations suggest that Mg^{2+} destabilizes the ensemble by enforcing the T541 and D642 interaction to enable its coordination; the CS is then characterized by a bending (CS1) or distortion of the C-terminus of α2 (OS), the destabilization of I_2 (CS2) coupled with increased disorder in the RK cluster leading to the ultimate displacement of ADP (CS3, fig. S6E). Indeed, most of these predicted effects are consistent with the perturbations seen by the addition of an excess of Mg^{2+} to ADP-saturated $Wzc_{CD\Delta C}$ (bottom panels in Fig. 5B and Fig. 5C). While amide resonance assignments for T541 and D642 are not available, the state of this interaction is reported by C544 that is in spatial proximity to these residues and is therefore an excellent probe for their relative orientation (Fig. 4C). The perturbation seen for C544 (0.03 ppm) is among the largest spectral changes seen in the presence of Mg^{2+} (bottom panels in Fig. 5B and Fig. 5C). Additionally, perturbations exceeding 3 standard deviations above the mean are also seen on the RK-cluster, on α2 (I_1) and α7/ α9 (I_2) in line with the *in silico* predictions (bottom panels in Fig. 5B and Fig. 5C).

Stabilizing the closed state enables the appropriate co-ordination of ATP•Mg^{2+} by monomeric $Wzc_{CD\Delta C}$

As noted above, in the REST2 simulations on the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex, the dominant clusters of the CS comprise of structures with distorted oligomerization interfaces (I_1 or I_2; Fig. 2D). On the other hand, the OS, while displaying intact, i.e. unbroken, I_1 and I_2 (Fig. 2C)

is unable to co-ordinate ATP•Mg^{2+} appropriately (Fig. 2B). We asked the question whether it is possible to rationally engineer a mutation (or a set of mutations) that would stabilize the CS and prevent the displacement of ATP from its binding site. To design these mutants *in silico*, we focused on the α2 helix within I_1, since it carries the conserved EX_2RX_2R motif that has been confirmed to play a critical role in function (*7*). Further, while most of the α2 helix is largely ordered in the crystal structure (for which all the constituent monomers exist in the CS) of Wzc_{CD}, its C-terminus displays significantly enhanced B-factors values (*7*) (Fig. 6A), reminiscent of CS4 in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations (fig. S5E). Drawing further inspiration from the crystal structure of *S. aureus* $CapB_{CD}$ (*8*) in which the region corresponding to α2 in Wzc_{CD} is significantly shorter and terminates in a long flexible linker, we generated a $Wzc_{CD\Delta C}$ mutant where residues 523-526 (that includes the last two residues of α2) were replaced *in silico* by glycines. We performed REST2 simulations with this construct, $Wzc_{CD\Delta C}$-$(G)_4$, bound to ATP•Mg^{2+} using the same protocols described in the previous sections and analyzed the resultant PDF in the cylindrical co-ordinate system. Indeed, only a single major state, the CS, proximal to the crystal structure (Fig. 6B), is found. Inspection of randomly selected structures show that, as expected, the mutated region (C-term end of α2) forms a flexible loop, and the bending about S512 is never observed (Fig. 6C) nor is the unfolding of α7/α9. While the RK-cluster is still flexible, no other remarkable features are evident in the structures in which ATP and Mg^{2+} remain in their appropriate locations (Fig. 6D).

Formation of the octameric ring stabilizes the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex

As noted above, it is possible to stabilize the complex of ATP•Mg^{2+} with monomeric $Wzc_{CD\Delta C}$ by stabilizing the CS and preventing formation of the OS. It is notable that all of the monomers in the crystal structure of Wzc_{CD} are in the CS. This suggests that formation of the

octameric ring likely constitutes the mechanism by which the CS may be stabilized, enable the stable engagement of ATP•Mg^{2+} and allow trans-phosphorylation to proceed. As an initial test of this hypothesis, we performed three sets of classical MD simulations on a reduced system, a trimer. Coarse-grained simulations suggest that it is absolutely necessary to stabilize both interaction interfaces as in a ring configuration to prevent the dissociation of dangling monomers e.g. in a minimal open system such as a trimer. Therefore, for all-atom simulations, in lieu of the octameric ring, we utilized a trimer but imposed position restraints on all Cα atoms of the flanking monomers while the central monomer was unrestrained. This system, we call a "pseudo-ring" maintains the fixed arc-length of the octameric ring and is expected to reasonably approximate the effects of the ring on the global transitions between the CS and the OS in the central monomer. Further, since the pseudo-ring simulations are shorter and employ classical MD, quantitative comparisons with the REST2 simulations employing the $Wzc_{CD\Delta C}$ monomer would be inappropriate since the latter samples a significantly larger conformational landscape. In order to establish an appropriate reference for the pseudo-ring simulations, we preformed three additional sets of classical MD simulations of $Wzc_{CD\Delta C}$ in the monomeric state in complex with ATP•Mg^{2+}.

As expected, all three monomer trajectories populate the major states (though not all the constituent clusters are populated given the limited sampling of conformational space) seen in our REST2 simulations. In particular, one trajectory (T3) samples the OS (Fig. 7A), in which the interactions of ATP with Y569 and D480 are broken (Fig. 7B, right panel), presumably representing a state primed eject to ATP. Another trajectory (T1) samples a state similar to CS2 (Fig. 2D and fig. S5E) in the REST2 simulations where α2 is bent while the D480-ATP interaction remains intact (Fig. 7E, black traces). The third trajectory (T2) displays no bend in α2 though the D480-ATP interaction is broken (Fig 7E, red traces). Overall, these simulations reflect the

tendencies noted in the REST2 simulations for the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex reflecting the fact that the two conditions, a stable oligomeric interface and bound ATP•Mg^{2+}, cannot be simultaneously satisfied in the monomeric state. In the pseudo-ring simulations, on the other hand, the system only samples the CS (Fig. 7C); the D480-ATP contact is maintained and the bending of α2 is not observed (left panel of Fig. 7D, and the bottom panel of Fig. 7F). Interestingly, there is some disorder introduced at the C-terminal end of α2 (fig. S8). The disorder in this region (fig. S8) is similar to that observed in the crystal structure that shows elevated B-factor values (Fig. 6A) (*7*). This is also the condition exploited to stabilize the CS in the $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+} simulations described above (Fig. 6C). Interestingly, one of the pseudo-ring trajectories shows a re-introduction of order in this region and this effect is coupled with a small change in the orientation of D480 from its optimal position with respect to ATP (fig. S8). Indeed, these classical MD simulations provide further evidence of the intimate coupling of the oligomerization interfaces and the catalytic site of $Wzc_{CD\Delta C}$.

DISCUSSION

We used enhanced sampling methods to probe the conformational landscapes of a variety of complexes involving the catalytic domain of the *E. coli* BY-kinase, Wzc. Our results (shown schematically in Fig. 8) indicate the existence a complex dynamic network within the kinase core that links the conformations of the nucleotide-binding and catalytic elements to those at the oligomerization surfaces. Our studies suggest the presence of discrete global conformations represented by an open state (OS), a closed state (CS) that is similar to that seen *in crystallo* (*7*) and a sparsely populated hyper-compact state (HS) that only forms in the presence of ATP and Mg^{2+} (fig. S9). $Wzc_{CD\Delta C}$ bereft of any ligands assumes an open state (OS) that is characterized by the coupled rotations of helix α2, that harbors a conserved EX_2RX_2R motif required for

oligomerization, and helix α3 that contains the Walker-A motif. This rotation of α3 results in an interaction between the conserved Walker-A threonine (T541) and the conserved Walker-B aspartate (D642) being broken, with the catalytic Walker-A lysine (K540) establishing a salt bridge with the latter. This arrangement is incompatible with the coordination of ATP•Mg^{2+} or catalysis. The presence of ATP•Mg^{2+} opposes the tendency to assume the OS primarily due to the reinforcement of the T541-D642 interaction and leads to the partial occupancy of the CS that displays a destabilized oligomerization interface and variously involves a distorted α2 on the first oligomerization interface (I_1) or partial unfolding of α7 and α9 on the complimentary oligomerization interface (I_2). The presence of ATP•Mg^{2+}, however, leads to a low occupancy state, HS, that is far more compact than that seen in the crystal structure. The HS displays some interesting characteristics. In addition to the catalytic elements together with ATP•Mg^{2+} being in their appropriate positions, R490 on the RK-cluster contacts the γ-phosphate of ATP in much the same fashion as the arginine-finger in F_1-ATPase (*24*) or the signature lysine of the deviant Walker-A motif in MinD (*26*). We suspect that the HS represents the orientation of the catalytic elements in the fully active form of Wzc_{CD}. Though I_2 is partially destabilized in this state (see Fig. 2C), we expect that formation of the oligomer stabilizes I_2 and likely leads to an enhancement of the population of the HS and a resultant increase in bulk kinase activity.

As noted above, the D642-T541 interaction, required for the proper coordination of Mg^{2+} appears to be a major driver in introducing the CS. Indeed, the influence of Mg^{2+} is evident by comparing the ADP•Mg^{2+} and ADP-only simulations; a CS is still dominant in the former while a continuous distribution containing almost equal populations of CS-like and OS-like states is seen in the latter. The results of the ITC and NMR measurements are consistent with the predicted influence of Mg^{2+}. Indeed, it is possible to stabilize the CS, prevent formation of the OS, and allow

proper ATP•Mg^{2+} co-ordination by introduction of a poly-glycine segment at the C-terminal end of α2 on I_1 (Fig. 6).

Thus, our results taken together suggest a scenario in which ATP•Mg^{2+}, in contrast to ADP, would be unable to form a stable complex with Wzc_{CD} in its monomeric state since its engagement would lead to structural distortions presumably at an energetic cost through the loss of several hydrogen bonds at the oligomerization interfaces involving α2, α7 and α9. Indeed, ITC experiments confirm a significantly reduced affinity (by ~200-fold) for ATP•Mg^{2+} compared to ADP alone. We predict, that oligomerization and formation of the octameric ring stabilizes the two interfaces and thereby locks in the CS, enabling the T541-D642 interaction while decoupling the latter from the catalytic K540 and allowing the stable co-ordination of ATP•Mg^{2+} (Fig. 9). Indeed, our preliminary results through classical MD simulations are in line with this prediction. Clearly, a full set of REST2 simulations on the octameric state are needed to fully establish this model. Thus, given the enhanced affinity for ATP•Mg^{2+} in the oligomeric state compared to that of the monomer, the cellular ATP/ADP ratio would then ensure the exchange of ADP, that is bound to the monomer, for ATP•Mg^{2+}, upon oligomer formation. The system would then be able to sample the HS state that has an ordered RK-cluster, enabling the insertion R490 into the active site and initiating chemistry.

Our model for Wzc_{CD} activation predicts that oligomerization, i.e. substrate docking, must precede ATP•Mg^{2+} binding and the formation of an active conformation. It is notable that this scenario provides a distinct advantage to ensure efficient phosphorylation in that it avoids a futile cycle of ATP hydrolysis that would necessarily occur upon ATP•Mg^{2+} binding prior to substrate docking due to the high inherent ATPase activity of Wzc_{CD} (*15*). A similar mechanism to prevent a futile hydrolysis cycle has been proposed for ATP-binding cassette (ABC) transporters (*28, 29*).

ATPases such as F_1-ATPase have also developed mechanisms to prevent the wasteful expenditure of ATP (*30*). Indeed, this mechanism likely represents another example of the elaborate means evolved by bacteria to avoid these sorts of energy spilling reactions (*31*). Further, the formation of a closed ring in the case of Wzc_{CD} provides the possibility that all constituent monomers could appropriately engage ATP•Mg^{2+} and be optimally primed for catalysis without the non-productive consumption of ATP. It is of note that for several canonical i.e. eukaryotic protein kinases, there is growing evidence that a fully activated state is achieved only upon substrate docking. A prime example is the mitogen activated protein kinase (MAPK) p38α, for which it has been shown that docking of the substrate peptide, in addition to dual phosphorylation on its activation loop, is required to achieve a fully active conformation (*32*). Indeed, substrate docking leads to significant enhancement in the affinity of p38α for ATP•Mg^{2+} (*33*). Thus, many features of the proposed activation mechanism of Wzc, and by inference of other BY-kinases, appears to resemble that seen in several eukaryotic protein kinases despite the deployment of a unique catalytic domain. It is possible that the sequence of events suggested by our model, where the substrate is engaged (oligomerization) prior to binding ATP, allowed BY-kinases to select as their phospho-receiver, an -OH moiety housed on a tyrosine of the Y-cluster rather than on water, providing bacteria with the ability to phosphorylate on tyrosine deploying an ATPase-like fold, rather than a canonical dual-lobe kinase scaffold utilized by eukaryotic cells (*11*).

MATERIALS AND METHODS

Computational Details

Structure Preparation

Chain A (average RMSD over the all chains based on Cα atoms: 0.12±0.04 Å) was randomly selected from the crystal structure (PDB: 3LA6) of the catalytic domain of the K540M

mutant of Wzc (Wzc_{CD}) (*7*). The missing segment of the RK-cluster was built using the Rosetta KIC protocol (*34*) using a single round of modeling without subsequent energy minimization (we refer to this as the crystal structure from here on). In all cases, the bound Ca^{2+} ion present in the crystal structure (that already contains bound ADP) was replaced by a Mg^{2+}. M540 on the Walker-A motif was mutated back to the native lysine using the mutagenesis tool in PyMOL. Additionally, all residues spanning the dynamic Y-cluster at the C-terminal tail (705-720) were removed to generate the $Wzc_{CD\Delta C}$ kinase core construct that was used in all simulations. A series of classical molecular dynamics (MD) simulations (using the setup described below) were performed using the structure generated above containing an intact RK-cluster with ADP•Mg^{2+} bound to the active-site to generate starting structures for all subsequent simulations. The structure was selected based on the presence of all ADP contacts that are observed in the crystal structure after equilibration of the conformation of the RK-cluster in the MD runs. In case of the $Wzc_{CD\Delta C}$•ADP•Mg^{2+} simulations, this structure was used as the starting structure, without further processing. For the $Wzc_{CD\Delta C}$•ADP simulations, the Mg^{2+} ion was removed from the starting structure in $Wzc_{CD\Delta C}$•ADP•Mg^{2+} to initiate the run. For the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations, the ATP•Mg^{2+} configuration was obtained from the crystal structure of the Wzc homolog MinD (PDB: 3Q9L) (*10*) and used to replace ADP•Mg^{2+} moiety. A short classical MD simulation using the setup described below was performed using $Wzc_{CD\Delta C}$•ATP•Mg^{2+} at 400 K to achieve proper solvation of the Mg^{2+} ion. The appropriately solvated $Wzc_{CD\Delta C}$•ATP•Mg^{2+} configuration was then used as the starting structure for further simulations. The starting $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+} structure used in the REST2 simulations was generated using the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} structure by replacing residues 523-526 by glycines using the mutagenesis tool in PyMOL. Following this, the system was prepared in the same manner as in other cases.

All simulations were performed using the GROMACS 2019 software suite (*35, 36*); the CHARMM36m (*37*) force-field was used to parameterize the system. The system was solvated using TIP3P waters in a cubic box with 1.5 nm edge distance from the solute. In all cases, charge neutralizing ions were added to the system. Energy minimization with gradient descent was performed with a 0.1 fs time step until energy convergence was obtained. Prior to the production runs the system was equilibrated, first in the NVT, and then in the NPT ensemble, for 100 ps each. NVT and NPT equilibration were conducted with position restraints on all heavy atoms using a 2 fs integration step. The production run was performed in the NPT ensemble using the Berendsen (*38*) and the Parrinello-Rahman (*39*) thermostat and barostat, respectively. Classical MD simulations were carried out at 300 K with a 2 fs integration step. The particle-mesh Ewald (PME) summation procedure with a cutoff of 1.0 nm was used to compute long-range electrostatic interactions and periodic boundary conditions were used to handle edge effects. In all cases, coordinates were stored every 10 ps.

REST2 Simulations

Structures were prepared as described above and used after NVT and NPT equilibration to initiate the REST2 simulations (*17*). Simulations were performed using the PLUMED 2 (*40*) plugin patched GROMACS 2019 suite. The "hot" region in our simulation encompassed all solute atoms including the nucleotide and Mg^{2+} (where present). A total of 14 replicas were used, where the lowest temperature (T_{min}) was set to 300 K. A series of short 2 ns REST2 simulations were performed by varying the temperature of the highest temperature replica (T_{max}) until the average replica exchange probability fell in the acceptable range. The desired exchange probability was achieved with a T_{max} of 400 K using 14 replicas with exchange probabilities of 23.6±0.8% for apo, 23.0±1.0% for $Wzc_{CD\Delta C}$•ATP•Mg^{2+}, 23.2±0.9% for $Wzc_{CD\Delta C}$•ADP•Mg^{2+}, 23.4±0.5% for

$Wzc_{CD\Delta C}$•ADP and 21.0±6.0% for $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+}. In all cases the temperature difference between replicas was obtained using the following expression:

$$T_n = T_{min} * e^{n*\ln\frac{\left(\frac{T_{max}}{T_{min}}\right)}{n-1}} \quad (1)$$

Where n is the n^{th} replica, T_n is its corresponding temperature; T_{min} and T_{max} are the minimum and maximum temperatures, 300 K and 400 K, respectively. The factor λ, that describes that scaling of the solute intramolecular potentials, was defined by the T_{min}/T_n ratio which then was used to re-parameterize the system for each replica n. Production runs in all cases were carried out for 200 ns per replica, for an effective 28000 ns sampling time; exchange attempts were made every 20 ps. Only the final 160 ns was used in the analyses in each case. The final temperatures and scaling factors (λ) for each replica were 1: 300.0, 1.000; 2: 306.7, 0.978; 3: 313.6, 0.957; 4: 320.6, 0.936; 5: 327.8, 0.915; 6: 335.1, 0.895; 7: 342.6, 0.875; 8: 350.3, 0.856; 9: 358.1, 0.838; 10: 366.1, 0.819; 11: 374.3, 0.801; 12: 382.7, 0.784; 13: 391.3, 0.767 and 14: 400.0, 0.750.

Definition of the Cylindrical Frame of Reference

A cylindrical co-ordinate system with an angle, θ and a rise, $|h|$, was defined to characterize the conformations of the structures generated in the REST2 simulations (see Fig. 1A). In order to define this frame, four reference points (labeled P1, P2, P3 and P4 in Fig. 1A) were selected - the center of mass (COM) of the Cα atoms of a segment comprising residues 541-547 (the first 2 turns of the N-terminal end of helix α3; the COM is indicated by a yellow sphere and labeled P1 in the left panel of Fig. 1A), the COM of the Cα atoms of residues 548-552 (the penultimate turn from the C-terminal end of helix α3, indicated by the magenta sphere and labeled P2 in the left panel of Fig. 1A), the COM of the Cα atoms of residues 569-573 (α4, indicated by the red sphere and labeled P3 in the left panel of Fig. 1A), and the COM of the Cα atoms of residues 505-509 (the

first turn of the N-terminal end of helix α2, indicated by the green sphere and labeled P4 in the left panel of Fig. 1A). Alignment of structures was then performed as follows - all structures were aligned such that points P1 and P2 were oriented along the z-axis. Next, the structures were rotated about the z-axis such that P3 was placed on the x-axis in the positive direction. The rise, $|h|$ (these, as defined were negative in all cases hence we use the absolute value for simplicity), was then the z-coordinate of P1, and the angle θ was defined as the angle between the x-axis and the xy-projection of the vector pointing from P1 to P4 (see Fig. 1A). These manipulations were performed using appropriate rotation matrices utilizing in-house code that used the C++ programing language. The PTEROS (*41*) library was used for trajectory file handling.

Classical MD Simulations on the Monomeric and Trimeric ("Pseudo-ring") States of $Wzc_{CD\Delta C}$

A single structure from cluster CS1 (that is almost in the same as the conformation seen in the crystal structure, see fig. S5D) of our $Wzc_{CD\Delta C}$•ATP•Mg^{2+} REST2 simulations was randomly selected and in the case of the monomer simulations served as the starting structure. In the case of the trimer simulations, chains A, B, and C were selected from the Wzc_{CD} crystal structure and the CS1 structure was aligned to chain B via Cα atoms. Chain B was subsequently removed from the configuration. Additionally, the C-terminal tails (705-720) were also removed from chains A and C, while the ADP•Mg^{2+} was kept in place. The simulation set-up was the same as described above. In the case of the trimer (pseudo-ring) simulations during the production runs, harmonic position restraints were imposed on all Cα atoms of chains A and C with a force constant of 1000 kJ mol^{-1} nm^2 acting along the x, y, and z directions. 100 ns simulations were performed in triplicate and in each case the initial velocities were obtained from a Maxwell distribution at 300 K using a random

seed. In the case of the trimer, the structural evolution was only analyzed for the central (unrestrained) monomer.

Expression and Purification of $Wzc_{CD\Delta C}$ and Wzc_{CD}

A pET15b vector containing either His_6-tagged $Wzc_{CD\Delta C}$ (447-704) or Wzc_{CD} (447-720), was used to transform BL21 DE3 (ThermoFisher) cells for protein expression. A single colony was used to inoculate 20 mL of Luria-Bertani (LB) media in the presence of 0.1 mg/mL ampicillin, grown at 37 °C overnight and used to inoculate 500 mL LB medium in the presence of 0.1 mg/mL ampicillin. The culture was grown at 37 °C until an OD_{600} of 1 was reached at which time expression was induced by the addition of 400 μM isopropyl β-D thiogalactopyranoside (IPTG). Expression was conducted at 37 °C for 4 hours, after which the cells were harvested via centrifugation at 4000 rpm for 30 minutes using a Fiberlite F12-6 x500 LEX rotor (ThermoFisher Scientific). All subsequent steps were carried out at 4 °C. The cells were lysed via sonication (3 minutes, pulse on for 0.3 seconds, pulse off for 0.7 seconds with a power amplitude of 30%) using a Sonic Dismembrator Model 500 (ThermoFisher Scientific) in lysis buffer containing 25 mM PIPES pH 6.5, 300 mM NaCl, 30% glycerol, 5 mM β–mercaptoethanol (BME), 1:500 lysozyme, and supplemented with half a Roche protease inhibitor tablet. The cell lysate was centrifugated at 10,000 rpm for 40 minutes to remove cell debris using a Fiberlite F21-8x50y rotor. The soluble lysate was then added to 5 mL of Ni-NTA affinity beads (GoldBio) and incubated for 1 hour with gentle rotation. Following this, the slurry was poured over a chromatography column and the flow through was discarded. The bound protein was washed with 50 mL of wash buffer (lysis buffer containing 20 mM imidazole), and then eluted with 50 mL elution buffer (lysis buffer containing 250 mM imidazole). The expressed protein in all cases was concentrated and further purified through gel-filtration using a Superdex 200 10/300 (GE HealthCare Biosciences) column.

$Wzc_{CD\Delta C}$ and Wzc_{CD} samples were dialyzed in dialysis buffer containing 25 mM Tris-HCl pH 7.5, 300 mM NaCl, 30% glycerol, and 5 mM BME. A portion of the Wzc_{CD} sample was also treated with ~0.25 nmol of λ-phosphatase (prepared in house) in the presence of 5 mM $MnCl_2$ in dialysis buffer.

Samples purified in the fashion described above contained bound ADP. In order to generate nucleotide-free protein, samples containing the His_6-tagged proteins were bound to a 1 mL HiTrap HP Ni-NTA (GE HealthCare Biosciences) affinity column and washed repeatedly with dialysis buffer for a total wash volume of 200 mL. Subsequently the protein was eluted using 10 mL of buffer containing 25 mM Pipes pH 6.5, 300 mM NaCl, 30% glycerol, 5 mM BME, and 1 M imidazole.

Isothermal Titration Calorimetry Measurements

After removal of bound nucleotide, $Wzc_{CD\Delta C}$ samples were dialyzed against Mg^{2+}-free buffer (25 mM Pipes pH 6.5, 300 mM NaCl, 5 mM BME, 10% glycerol) or Mg^{2+}-containing buffer (25 mM Pipes pH 6.5, 300 mM NaCl, 5 mM BME, 10% glycerol, and 5 mM $MgCl_2$) at 4 °C. Following this, protein samples were concentrated to 40 μM via spin columns. ADP (sodium salt, Sigma) was dissolved in the same buffer and the pH was subsequently adjusted to 6.5 using sodium hydroxide (NaOH). AMPPCP (disodium salt, Sigma) was dissolved in the Mg^{2+}-containing buffer and the pH was adjusted to 6.5 using sodium hydroxide (NaOH). Isothermal titration calorimetry (ITC) measurements were performed at 25 °C using a MicroCal iTC200 isothermal titration calorimeter (Malvern). All measurements were done in triplicate. The protein/titrant concentrations used were 40 μM/400 μM and 40 μM/1.6 mM for the ADP•Mg^{2+} and AMPPCP•Mg^{2+} titrations, respectively. Separate affinity measurements for ADP in the absence of Mg^{2+} where done using 40 μM/400 μM, 20 μM/300 μM and 15 μM/180 μM protein/titrant

concentrations. All titrations were performed with a rotation speed of 500 rpm. A total of 16 injections were made in each case where the first one consisted of 0.4 μL with a duration 0.8 seconds, and the rest consisted of 2.4 μL with a duration of 4.8 seconds and filter periods of 5 seconds. The spacing between injections used was 240 seconds. In all cases the same experiment was performed by titrating the titrant into buffer alone and used to normalize the data. The normalized data were fitted to the one-site binding model using Origin (OriginLab).

Analysis of the Oligomerization State of Wzc_{CD}

Nucleotide-free Wzc_{CD} or nucleotide-free, λ-phosphatase-treated Wzc_{CD} were each concentrated to 200 μM and then passed through a Superdex 200 10/300 (GE HealthCare Biosciences) gel-filtration column pre-equilibrated with buffer containing 25 mM PIPES pH 6.5, 300 mM NaCl, 30% glycerol, 5 mM β–mercaptoethanol (BME) to assess their respective oligomerization states. Fractions suspected to contain Wzc_{CD} were analyzed by SDS-PAGE gels.

Chemical Shift Perturbation Analysis

Uniformly ^{2}H,^{15}N-labeled $Wzc_{CD\Delta C}$ was expressed and purified using previously described protocols (*16*) with the following minor modifications. After elution from the Ni^{2+}-affinity column, the protein was incubated overnight with an excess of ATP (~20-fold) and Mg^{2+} (~200-fold). After adding an excess of EDTA, the sample was concentrated and injected on a size exclusion column (Superdex 75, GE Healthcare) pre-equilibrated with NMR buffer containing 50 mM phosphate, pH 6.0, 50 mM NaCl, 25 mM DTT, 200 mM EDTA. Unlike the previously published protocol, no further nucleotides were added at this point. To eliminate bound ADP, the sample was buffer exchanged (1/10000 dilution factor) against the NMR buffer using spin columns (10 kDa cutoff). The NMR samples typically consisted of ~150 μL (in 4 mm Shigemi tubes) of ~100 μM U-^{2}H,^{15}N-labeled $Wzc_{CD\Delta C}$. Ligands (Mg^{2+}, 200 molar equivalents; ADP, up to 20 molar equivalents) were

added directly to the NMR tubes from stock solutions prepared in the same buffer. ^{15}N, ^{1}H TROSY experiments (64 scans, 1.5 s recycle delay) were collected using 2048 and 256 complex points, with sweep-widths of 15 and 32 ppm in the direct and indirect dimensions, respectively. All experiments were acquired at 25 °C on a 700 MHz Avance III Bruker spectrometer equipped with cryogenic probes capable of applying pulse-field gradients along the z-axis. The spectra were processed using nmrPipe (*42*) and analyzed using nmrViewJ (*43*). Amide chemical shift perturbations were calculated using the formula:

$$\Delta\delta = \sqrt{\Delta\delta_{HN}^2 + (0.154\Delta\delta_N)^2} \quad (2)$$

$\Delta\delta_{HN}$ and $\Delta\delta_N$ represent the perturbations for amide ^{1}H and ^{15}N chemical shifts, respectively. The $\Delta\delta$ values represent the differences between well-resolved, free and ADP-bound resonances of $Wzc_{CD\Delta C}$ under partial ADP saturation or the perturbations induced by Mg^{2+} on ADP-saturated $Wzc_{CD\Delta C}$.

REFERENCES

1. C. Grangeasse, S. Nessler, I. Mijakovic, Bacterial tyrosine kinases: evolution, biological function and structural insights. *Philos. T. Roy. Soc.* **367**, 2640-2655 (2012).
2. C. Grangeasse, R. Terreux, S. Nessler, Bacterial tyrosine-kinases: structure-function analysis and therapeutic potential. *Biochim. Biophys. Acta* **1804**, 628-634 (2010).
3. F. Jadeau, E. Bechet, A. J. Cozzone, G. Deléage, C. Grangeasse, C. Combet, Identification of the idiosyncratic bacterial protein tyrosine kinase (BY-kinase) family signature. *Bioinformatics* **24**, 2427-2430 (2008).
4. B. Obadia, S. Lacour, P. Doublet, H. Baubichon-Cortay, A. Cozzone, C. Grangeasse, Influence of tyrosine-kinase Wzc activity on colanic acid production in *Escherichia coli* K12 cells. *J. Mol. Biol.* **367**, 42-53 (2007).

5. R. F. Collins, K. Beis, C. Dong, C. H. Botting, C. McDonnell, R. C. Ford, B. R. Clarke, C. Whitfield, J. H. Naismith, 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).

6. D. Lee, J. Zheng, Y. She, Z. Jia, Structure of *Escherichia coli* tyrosine kinase Etk reveals a novel activation mechanism. *EMBO J.* **27**, 2030 (2008).

7. E. Bechet, J. Gruszczyk, R. Terreux, V. Gueguen-Chaignon, A. Vigouroux, B. Obadia, A. J. Cozzone, S. Nessler, C. Grangeasse, Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. *Mol. Microbiol.* **77**, 1315-1325 (2010).

8. V. Olivares-Illana, P. Meyer, E. Bechet, V. Gueguen-Chaignon, D. Soulat, S. Lazereg-Riquier, I. Mijakovic, J. Deutscher, A. J. Cozzone, O. Laprévote, S. Morera, C. Grangeasse, S. Nessler, Structural basis for the regulation mechanism of the tyrosine kinase CapB from *Staphylococcus aureus*. *PLoS Biol.* **6**, e143 (2008).

9. 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).

10. W. Wu, K. T. Park, T. Holyoak, J. Lutkenhaus, Determination of the structure of the MinD-ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC. *Mol. Microbiol.* **79**, 1515-1528 (2011).

11. T. Hunter, The genesis of tyrosine phosphorylation. *CSH Perspect. Biol.* **6**, a020644 (2014).

12. 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).

13. 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).

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

15. D. Soulat, J. M. Jault, C. Geourjon, P. Gouet, A. J. Cozzone, C. Grangeasse, Tyrosine-kinase Wzc from *Escherichia coli* possesses an ATPase activity regulated by autophosphorylation. *FEMS Microbiol. Lett.* **274**, 252-259 (2007).

16. D. B. Temel, K. Dutta, S. Alphonse, J. Nourikyan, C. Grangeasse, R. Ghose, Regulatory interactions between a bacterial tyrosine kinase and its cognate phosphatase. *J. Biol. Chem.* **288**, 15212-15228 (2013).

17. L. Wang, R. A. Friesner, B. J. Berne, Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **115**, 9431-9438 (2011).

18. M. Dittrich, S. Hayashi, K. Schulten, On the mechanism of ATP hydrolysis in F_1-ATPase. *Biophys. J.* **85**, 2253-2266 (2003).

19. S. C. Kamerlin, P. K. Sharma, R. B. Prasad, A. Warshel, Why nature really chose phosphate. *Q. Rev. Biophys.* **46**, 1-132 (2013).

20. P. I. Hanson, S. W. Whiteheart, AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519-529 (2005).

21. S. N. Gates, A. Martin, Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Sci.* **29**, 407-419 (2020).

22. T. Krell, J. Maclean, D. J. Boam, A. Cooper, M. Resmini, K. Brocklehurst, S. M. Kelly, N. C. Price, A. J. Lapthorn, J. R. Coggins, Biochemical and X-ray crystallographic studies on shikimate kinase: the important structural role of the P-loop lysine. *Protein Sci.* **10**, 1137-1149 (2001).

23. D. D. Leipe, E. V. Koonin, L. Aravind, Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781-815 (2003).

24. J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. *Nature* **370**, 621-628 (1994).

25. G. N. Nagy, R. Suardiaz, A. Lopata, O. Ozohanics, K. Vekey, B. R. Brooks, I. Leveles, J. Toth, B. G. Vertessy, E. Rosta, Structural characterization of arginine fingers: identification of an arginine finger for the pyrophosphatase dUTPases. *J. Am. Chem. Soc.* **138**, 15035-15045 (2016).

26. J. Lutkenhaus, M. Sundaramoorthy, MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. *Mol. Microbiol.* **48**, 295-303 (2003).

27. E. Gout, F. Rebeille, R. Douce, R. Bligny, Interplay of Mg^{2+}, ADP, and ATP in the cytosol and mitochondria: unravelling the role of Mg^{2+} in cell respiration. *Proc. Natl. Acad. Sci. USA* **111**, E4560-4567 (2014).

28. J. Chen, S. Sharma, F. A. Quiocho, A. L. Davidson, Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport. *Proc. Natl. Acad. Sci. USA* **98**, 1525-1530 (2001).

29. K. Hollenstein, D. C. Frei, K. P. Locher, Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213-216 (2007).

30. J. E. Walker, The ATP synthase: the understood, the uncertain and the unknown. *Biochem. Soc. Trans.* **41**, 1-16 (2013).

31. J. B. Russell, G. M. Cook, Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol. Rev.* **59**, 48-62 (1995).

32. A. Kuzmanic, L. Sutto, G. Saladino, A. R. Nebreda, F. L. Gervasio, M. Orozco, Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *eLife* **6**, e22175 (2017).

33. Y. Tokunaga, K. Takeuchi, H. Takahashi, I. Shimada, Allosteric enhancement of MAP kinase p38α's activity and substrate selectivity by docking interactions. *Nature Struct. Mol. Biol.* **21**, 704-711 (2014).

34. A. Stein, T. Kortemme, Improvements to Robotics-Inspired Conformational Sampling in Rosetta. *Plos One* **8**, e63090 (2013).

35. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701-1718 (2005).

36. C. Kutzner, S. Pall, M. Fechner, A. Esztermann, B. L. de Groot, H. Grubmuller, More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J. Comput. Chem.* **40**, 2418-2431 (2019).

37. J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller, A. D. MacKerell, Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Meth.* **14**, 71-73 (2017).

38. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).

39. M. Parrinello, R. A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182 (1981).

40. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604-613 (2014).

41. S. O. Yesylevskyy, Pteros 2.0: Evolution of the fast parallel molecular analysis library for C++ and python. *J. Comput. Chem.* **36**, 1480-1488 (2015).

42. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).

43. B. A. Johnson, From raw data to protein backbone chemical shifts using NMRFx processing and NMRViewJ analysis. *Meth. Mol. Biol.* **1688**, 257-310 (2018).

44. M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Riateri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, M. Parrinello, PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **180**, 1961-1972 (2009).

45. M. Bonomi, D. Branduardi, F. L. Gervasio, M. Parrinello, The unfolded ensemble and folding mechanism of the C-terminal GB1 beta-hairpin. *J. Am. Chem. Soc.* **130**, 13938-13944 (2008).

46. M. Levitt, A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**, 59-107 (1976).

47. Y. Cheng, Mean shift, mode seeking, and clustering. *IEEE T. Pattern Anal.* **17**, 790-799 (1995).

48. D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis. *IEEE T. Pattern Anal.* **24**, 603-619 (2002).

49. W. Jia, H. Zhang, X. He, M. Piccardi, *Mean shift for accurate license plate localization*. Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005 (Vienna, 2005), pp. 566-571.

50. H. Zhou, Y. Yuan, C. Shi, Object tracking using SIFT features and mean shift. *Comput. Vis. Image Und.* **113**, 345-352 (2009).

51. J. Einbeck, Bandwidth selection for nonparametric unsupervised learning techniques – a unified approach via self-coverage. *J. Pattern Recogn. Res.* **6**, 175-192 (2011).

52. J. Einbeck, G. Tutz, L. Evers, Local principal curves. *Stat. Comput.* **15**, 301-313 (2005).

Acknowledgements

NMR experiments were carried out in the CUNY ASRC Biomolecular NMR facility. The authors thank Dr. Christophe Grangeasse (MMSB, CNRS/Université Lyon) and Dr. Paul Robustelli (Dartmouth) for critical comments on this manuscript.

Funding

This work is supported by NSF grants MCB1937937 and PHY1811770. FH acknowledges support from the United States Department of Education GAANN award P200A150068.

Author contributions

R. G. conceived the project; F.H. performed and analyzed all simulations and experiments except the NMR experiments that were conducted by A.P. F. H. prepared a first draft of the paper and figures that were refined by R. G. with input from A. P. and F. H.

Competing interests

The authors declare that they have no competing interests.

Data and materials availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or in the Supplementary Material. Additional data related to this paper may be requested from the authors.

FIGURE LEGENDS

Fig. 1. REST2 simulations on unliganded $Wzc_{CD\Delta C}$. (A) The cylindrical coordinate frame (defined by θ and $|h|$) to represent the global conformations observed in the simulations. **(B)** Density of states in θ-$|h|$ space plotted using kernel density estimation; P(θ,$|h|$) represents the probability density. The green dot indicates the reference values for the crystal structure. Only a single major state (an open state, OS) is sampled. **(C)** Close-up of the active-site in the OS showing the orientation of α3 and the proximity between the Walker-A K540 and the Walker-B D642. **(D)** In the crystal structure (representing a closed conformation, CS, cyan) D642 forms a hydrogen-bond with the Walker-A T541 (dotted oval). K540 (methionine in the crystal structure) is in a downward orientation similar to that in the homologous MinD (yellow). **(E)** In wild-type shikimate kinase (SK_{WT}, magenta), the catalytic elements are appropriately oriented for chemistry; the Walker-A threonine (T16) is hydrogen-bonded to the Walker-B aspartate (D32) (green dotted circle). In the inactive K15M mutant (SK_{K15M}, blue), T16 and D32 have drifted apart due to the motion of the helix (green arrow) and display orientations similar to the corresponding sidechains in the $Wzc_{CD\Delta C}$ OS.

Fig. 2. REST2 simulations on the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} complex. (A) Projection onto θ-$|h|$ space shows three distinct states: OS, CS (similar to the crystal structure; green dot), and a hyper-compact state (HS). **(B)** Close-up view of the active site in the OS. The catalytic elements, key ATP contacting residues, and ATP are indicated. The OS is similar to that in the apo simulations; K540 forms a salt-bridge with D642; key ATP-contacting residues (D480 and Y569) have moved away from ATP (green stick representation). Additionally, the adenosine moiety of ATP is displaced from its binding pocket relative to that in the HS (tan). **(C)** The HS (left panel) shows an ordered RK-cluster, however helices α7 and α9 of the second interaction interface (I_2) are partially

unfolded (green ellipses). In contrast, the OS (right panel) is characterized by an intact I_2 but a disordered RK-cluster. **(D)** The two largest clusters (CS2 and CS5) obtained through an EVA-MS decomposition of the CS (also see fig. S5) are characterized by mutually exclusive distortions at the oligomerization interface: bending of α2 in CS2 (I_1, red arrows) or partial unfolding of helices α7 and α9 in CS5 (I_2, green ellipses).

Fig. 3. Active site conformation for the HS in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations. (**A**) All catalytic and key ATP-interacting residues are shown in stick representation for an ensemble of 10 structures (left panel) and a single representative structure (right panel). Key catalytic residues on the Walker-A (dark red, K540 and T541) and Walker-B (blue, D642) motifs are indicated. ATP-coordinating residues (D480 in magenta, R490 in cyan, and Y569 in yellow) and additional ones involved in stabilizing the conformation (K492 in cyan and E572 in green) are also indicated. R490 and K492 of the RK-cluster are ordered in this state. Close-up views of the catalytic sites in the active conformations of MinD **(B)** and the F_1-ATPase **(C)** are shown in approximately similar orientations as the right panel of **A**. Key catalytic elements are indicated and colored similarly as their Wzc counterparts in **A**.

Fig. 4. REST2 simulations of $Wzc_{CD\Delta C}$•ADP• Mg^{2+} and $Wzc_{CD\Delta C}$•ADP complexes. (A) θ-$|h|$ space projection of the $Wzc_{CD\Delta C}$•ADP•Mg^{2+} simulations. The CS and OS are highlighted by green ellipses (left); the red dot indicates crystal structure values. Close-up (middle panel) of the active-sites of major EVA-MS decomposed CS clusters (CS1, CS2, CS3; also see fig. S6). ADP (cyan ellipse) and key nucleotide-coordinating residues, D480 and Y569 (green ellipse), are colored according to their constituent clusters. The right panel illustrates a representative OS structure. **(B)** States sampled in the $Wzc_{CD\Delta C}$•ADP simulations; the tricolor arrow indicates regions in θ-$|h|$ space. Conformations of α2 (bending indicated by red arrows), α3, and αB of the major EVA-

MS-decomposed clusters (C1, C2, C4; also see fig. S7) are shown in the middle panel. Y569, whose position varies between the three clusters, is also indicated. Collective displacements of secondary structural elements are indicated by tricolor arrows representing specific regions in θ-$|h|$ space. (**C**) Active-site configurations for C1 of $Wzc_{CD\Delta C}$•ADP (left) and CS1 of $Wzc_{CD\Delta C}$•ADP•Mg^{2+} (right) are compared. **(D)** Normalized distribution of D480,Cγ–ADP,O3' distances (3.6±0.3 Å over 16 chains *in crystallo*) in the $Wzc_{CD\Delta C}$•ADP•Mg^{2+} (blue) and $Wzc_{CD\Delta C}$•ADP (red) simulations.

Fig. 5. **Experimental verification of *in silico* predictions. (A)** Representative ITC thermograms of the interaction of $Wzc_{CD\Delta C}$ with ADP (left), ADP•Mg^{2+} (middle), and AMPPCP•Mg^{2+} (right). **(B)** (Top panel) Amide chemical shift perturbations (Δδ) induced by ADP on $Wzc_{CD\Delta C}$ in the absence of Mg^{2+}. Δδ values are shown only for well-resolved resonances corresponding to the apo and ADP-bound species under partial ADP saturation for the spectra that are in slow exchange. (Bottom panel) Amide chemical shift perturbations induced by 200 molar equivalents of Mg^{2+} on the ADP-saturated spectra of $Wzc_{CD\Delta C}$. Red and cyan dashed lines indicate two and three standard deviation (σ) threshold beyond the mean (<Δδ>), respectively; green bars indicate exchange broadened resonances. **(C)** (Top panel) ADP-induced perturbations from the top panel of **B** depicted as red spheres on the structure of $Wzc_{CD\Delta C}$. (Bottom panel) Mg^{2+}-induced perturbations from the bottom panel of **B** mapped onto the structure of $Wzc_{CD\Delta C}$. Residues for which resonances show Δδ values larger than <Δδ>+3σ or are exchange-broadened, are indicated by blue and green spheres, respectively. C544, a key probe of the active-site conformation and shows a large perturbation is indicated.

Fig. 6. Stabilization of the CS *in silico.* (A) B-factor values from the crystal structure of Wzc_{CD} reveal a high degree disorder at the C-terminal end of α2 (dotted circle). **(B)** Projection of the

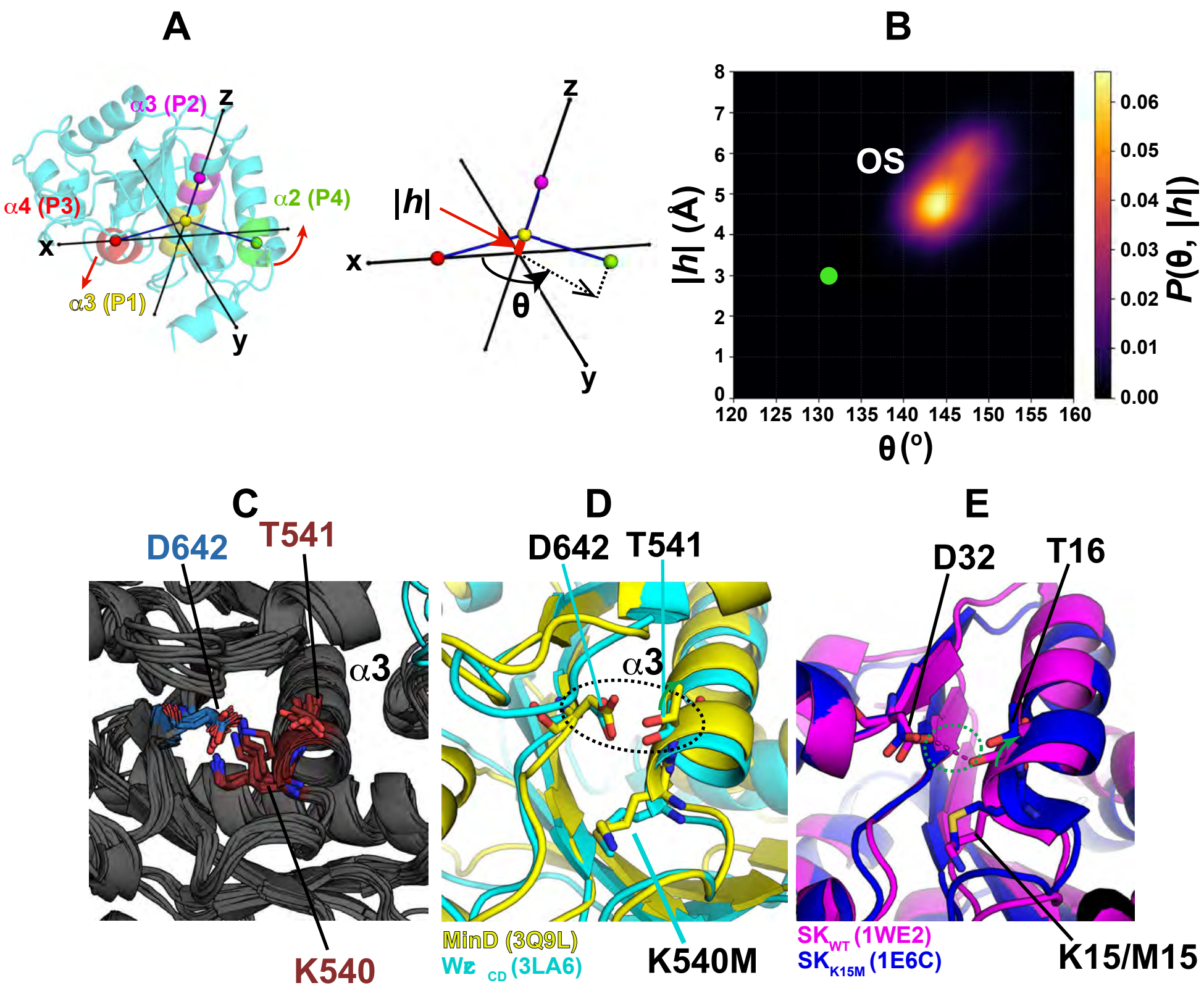
REST2-generated ensemble of $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+} onto θ-|*h*| space reveals the sampling of only a CS (the Wzc_{CD} crystal conformation is indicated by the green dot); the OS and HS seen in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations (see Fig. 2A) are no longer seen. **(C)** 10 randomly selected structures superimposed on Cα atoms are shown. The α2 configuration resembles that seen in the crystal structure (i.e. no bending is observed) and the introduction of the $(G)_4$ introduces flexibility at the C-terminal end of α2. The RK-cluster (cyan and magenta) is dynamic. However, no other structural evolution (e.g. the unfolding of the α7 or α9 helices i.e. disruption of I_2) is seen. **(D)** The key ATP coordinating elements at the catalytic site are ordered and ATP remains stably bound.

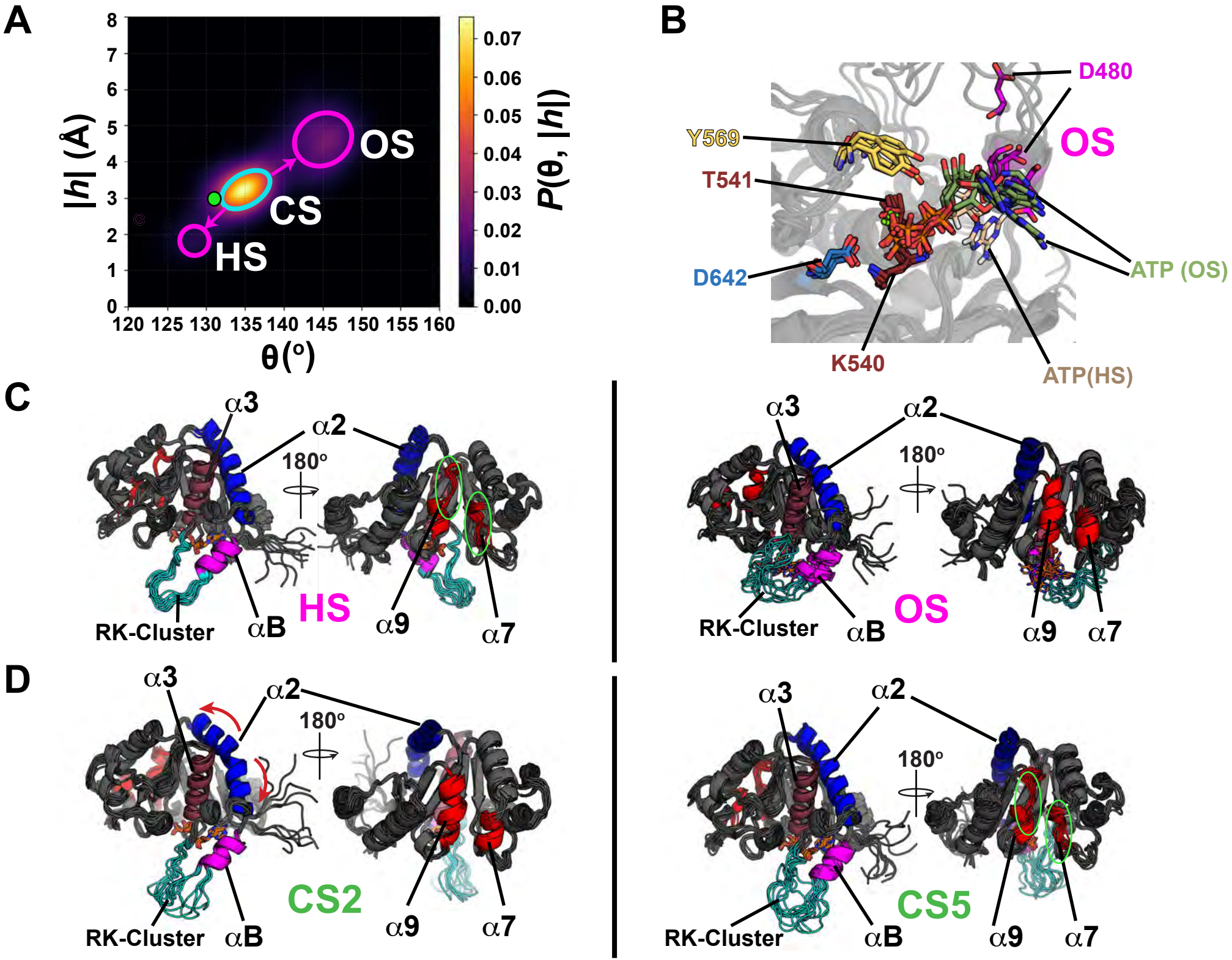
Fig. 7. Classical MD simulations of Wzc_{CD}•ATP•Mg^{2+} in the monomeric and pseudo-ring states. (A) θ-|*h*| space representation of a monomer trajectory (T3) with the points color-coded according to their timestamps (in ns, the black dot indicates crystal structure values); OS-like and CS-like states are observed. **(B)** Comparison of structures in the early (left panel, bending of α2 shown by red arrows) and late (right panel, displacement of Y569 and D480 from ATP indicated by blue arrows) time-points from **A**. **(C)** Conformations of the central monomer from a representative pseudo-ring trajectory (T1) plotted in θ-|*h*| space shows only a CS-like state. **(D)** The last frame from T1 shows that α2 is intact and ATP continues to be engaged by Y569 and D480. For the monomer (**E**) and the central monomer in the pseudo-ring **(F)** simulations, the evolution of the S512,Cα–T515,Cα (proxy for α2 bending, top panel) and D480,Cγ–ATP,O3' (proxy for ATP-contact, lower panel) distances are shown. The left panels indicate time-courses in the three independent trajectories (T1, T2, and T3), and right panels indicate frequencies of the corresponding distances. The green/black arrows highlight the inverse correlation between the distances for the monomer. Also see fig. S8.

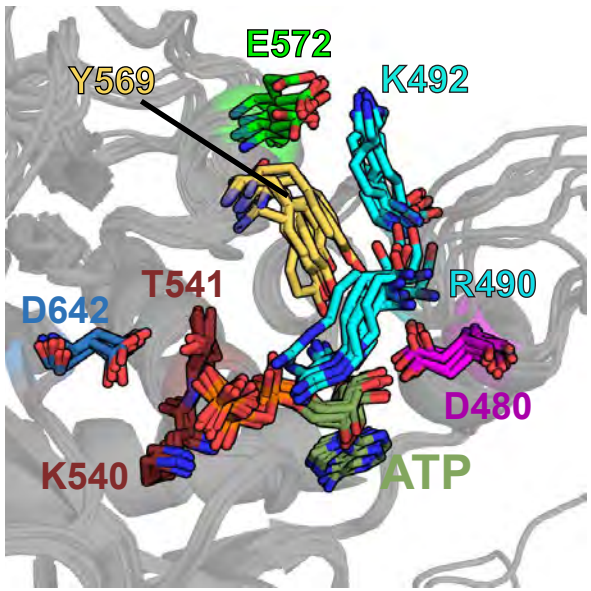
Fig. 8. **Schematic representation of key conformational features identified in the REST2 simulations.** The RK-cluster (cyan, with αB in magenta), α2 (I_1, dark-blue), α7 and α9 (I_2, red), α3 (black) and α4 (dark-red) are indicated. Monomeric, unliganded $Wzc_{CD\Delta C}$ exists in the OS; binding of ATP•Mg^{2+} forces formation of the CS (red arrows). This state can be qualitatively conceptualized as a strained network (black arcs) of connected deformed springs. The CS comprises of several exchanging sub-states (clusters) with distortions that are incompatible with trans-phosphorylation. The most populated of these clusters show different sorts of distortions: CS5 - partial unfolding of α7 and α9 (I_2); CS4 - enhanced disorder in the RK-cluster (dashed line) and at the N-terminal end of α2 (red arcs); CS2 - bending of α2 (I_1; green arrows). Alternatively, ATP is displaced from the binding site (purple arrow) restoring the OS. These various distortions may be considered to be independent mechanisms by which this mechanical network "releases its internal strain". A HS (proposed to represent a reactive state) that is also populated, albeit sparsely, when bound to ATP•Mg^{2+}, is not shown for simplicity.

Fig. 9. Model for Wzc_{CD} activation and nucleotide exchange. In the monomeric state, Wzc_{CD} is stably bound to ADP in the absence of Mg^{2+}. Though α2 is bent (green arrows), the absence of Mg^{2+} allows a coupled rigid body motion (red arrows; no unfolding occurs) of α7 and α9 (I_2) together with α4 allowing for the proper coordination of ADP. Oligomerization and formation of the octameric ring (only three constituent monomers are represented), stabilizes I_1 and I_2 (light green arrows) and allows formation of the CS. This state also carries some "strain" that is "released" by introducing disorder at the C-terminus of α2 (red arcs), a segment that does not participate in oligomerization. In this oligomeric state, ATP•Mg^{2+} can be stably engaged with an affinity that is similar to (or possibly exceeds that) of ADP. Consequently, the cellular ATP/ADP ratio allows nucleotide exchange and the formation of an active conformation with ATP•Mg^{2+}

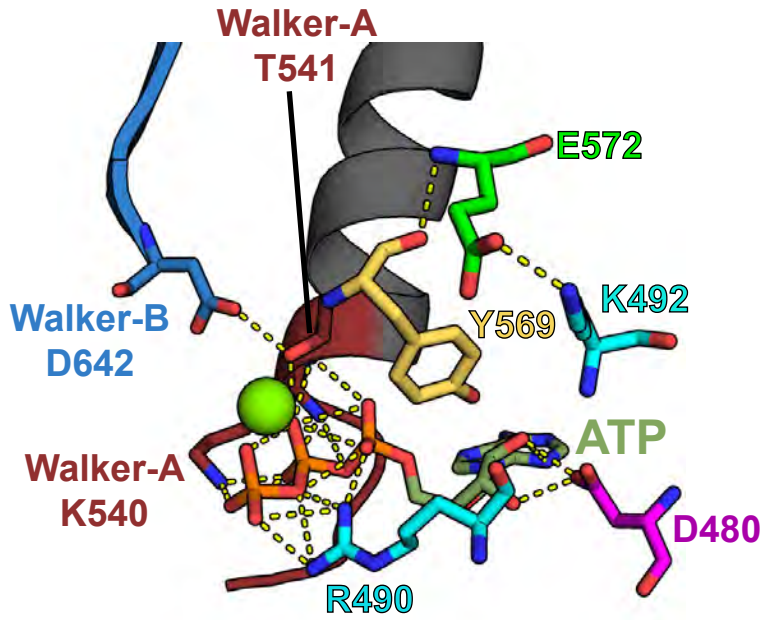
correctly coordinated, the catalytic elements properly oriented and the RK-cluster well-ordered (thick lines) and appropriately engaged to enable catalysis.


A
α3 (P2)
z
α4 (P3)
α2 (P4)
x
α3 (P1)
y
|h|
z
x
θ
y
B
OS
|h| (Å)
θ (°)
P(θ, |h|)
C
D642
T541
α3
K540
D
D642
T541
α3
MinD (3Q9L)
WzCD (3LA6)
K540M
E
D32
T16
SKWT (1WE2)
SKK15M (1E6C)
K15/M15

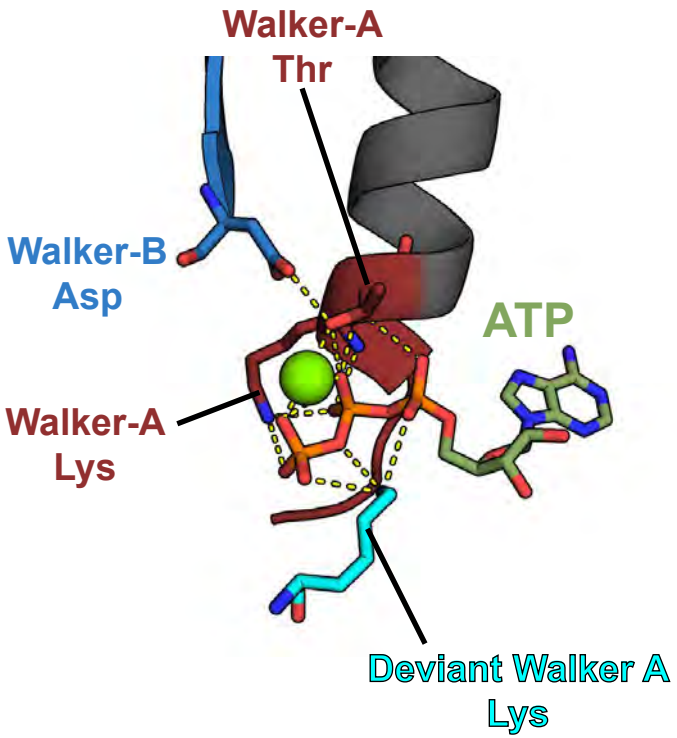

A
|h| (Å)
θ(°)
P(θ, |h|)
OS
CS
HS
B
D480
Y569
T541
D642
K540
ATP (OS)
ATP(HS)
OS
C
α3
α2
180°
RK-Cluster
αB
HS
α9
α7
α3
α2
180°
RK-Cluster
αB
OS
α9
α7
D
α3
α2
180°
RK-Cluster
αB
CS2
α9
α7
α3
α2
180°
RK-Cluster
αB
CS5
α9
α7



A

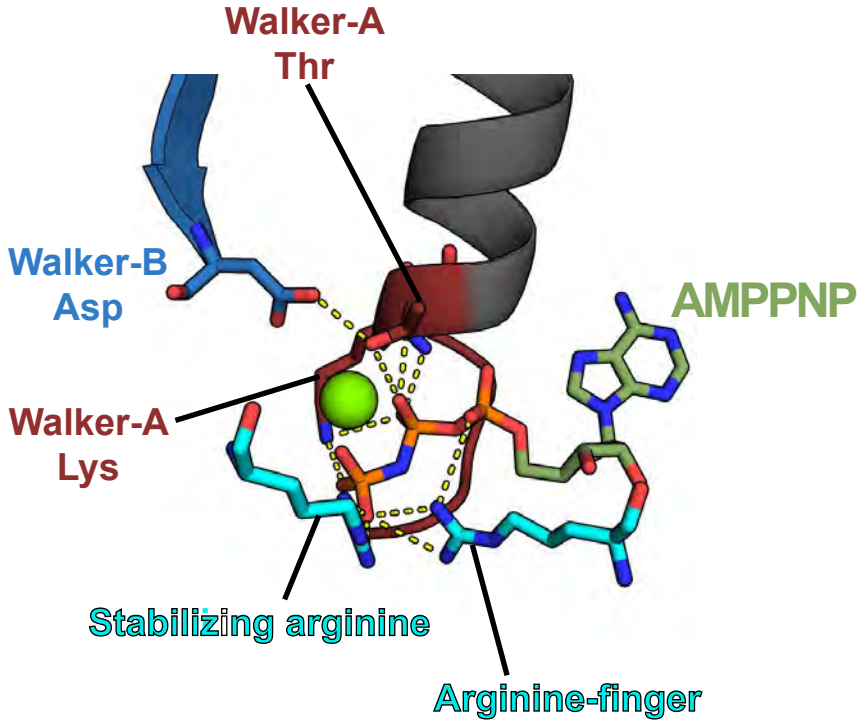


B

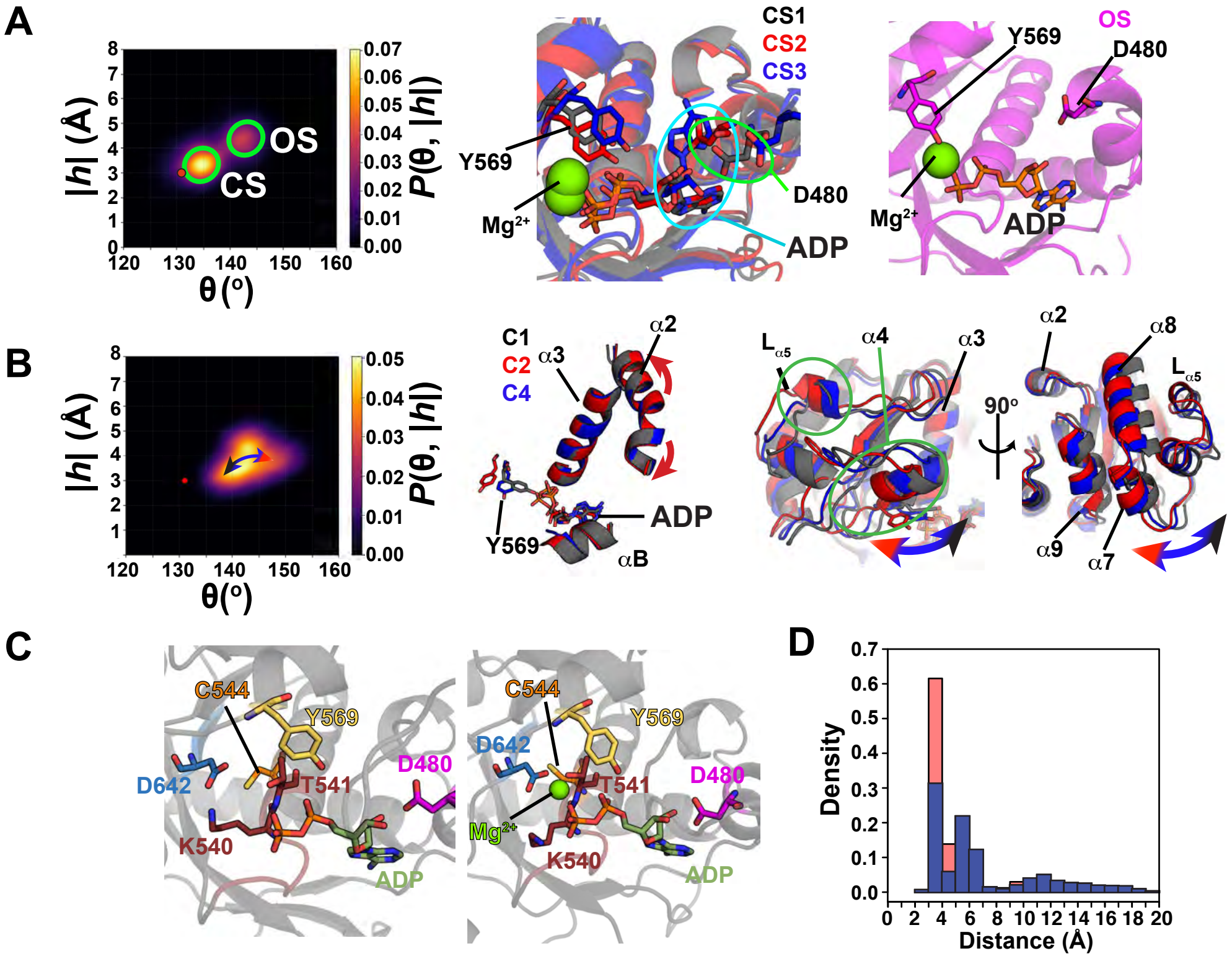


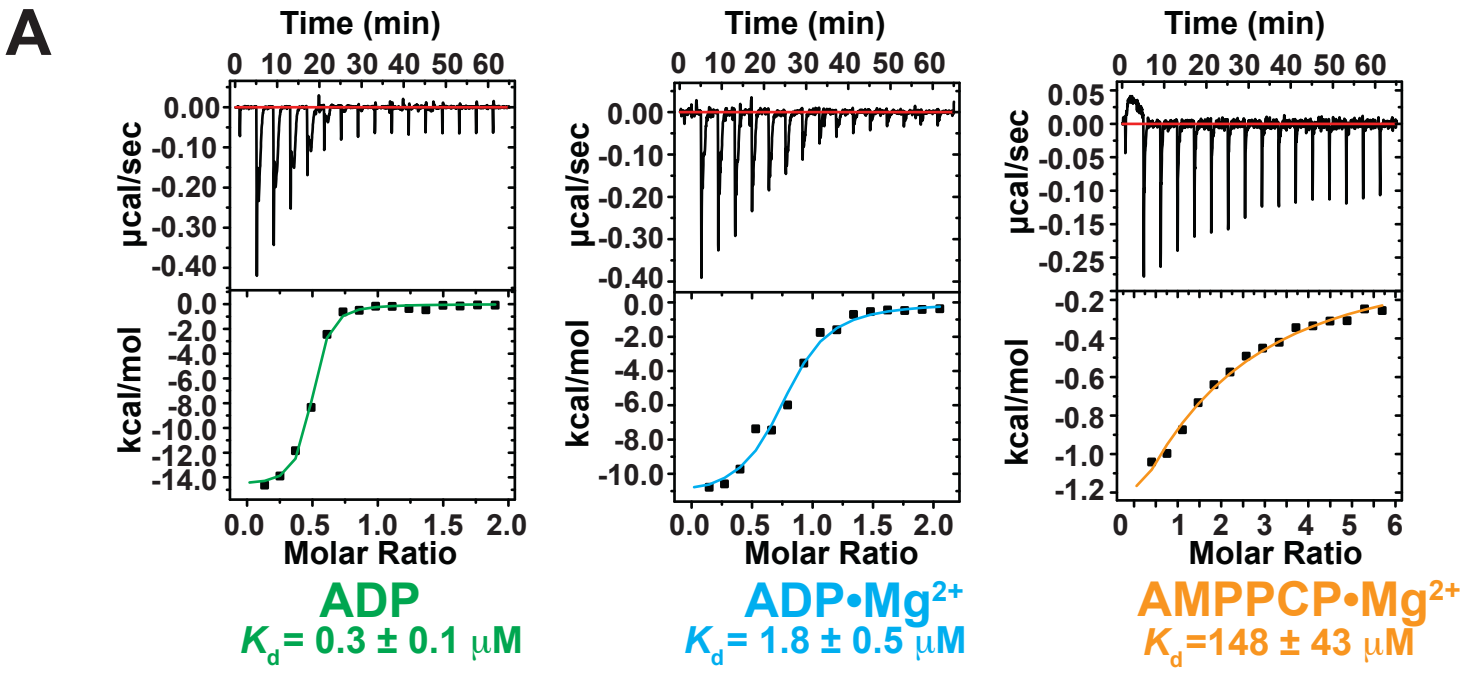
MinD (3Q9L)

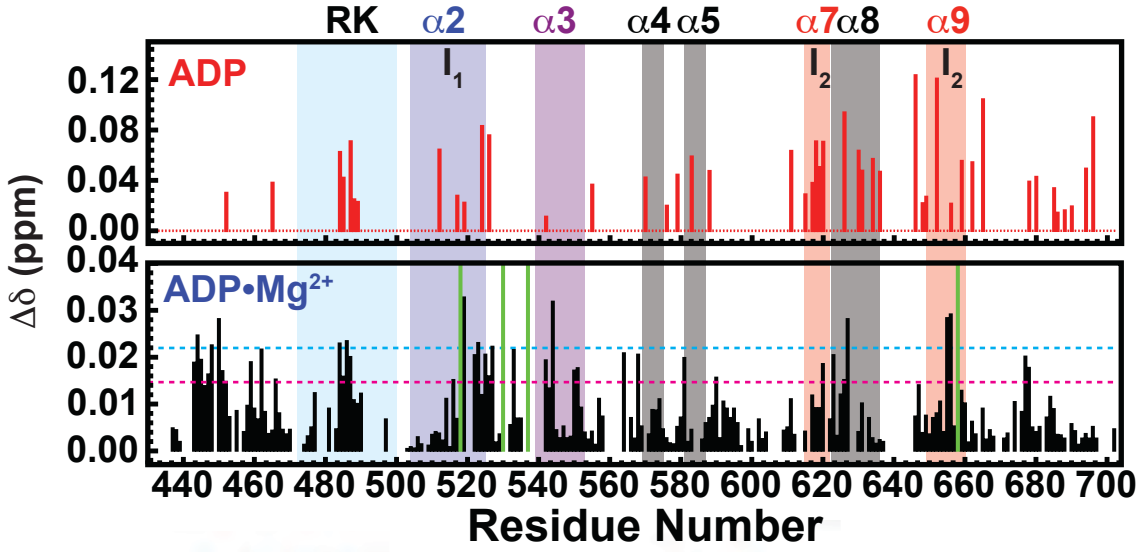
C

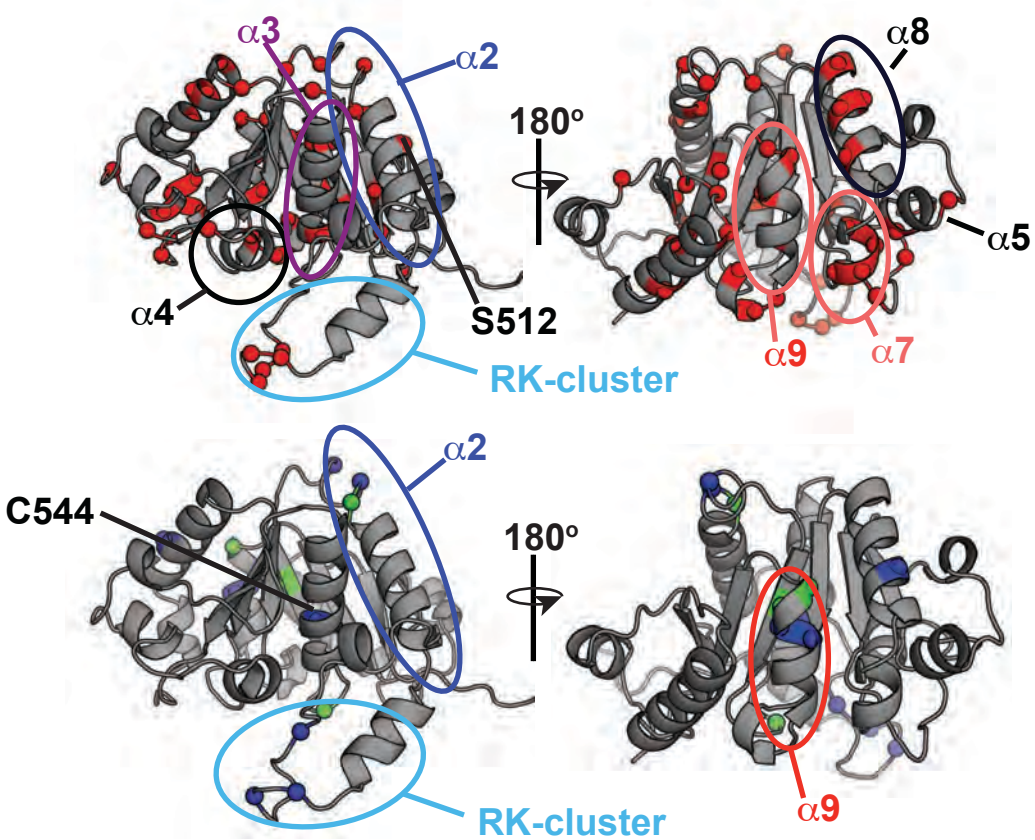


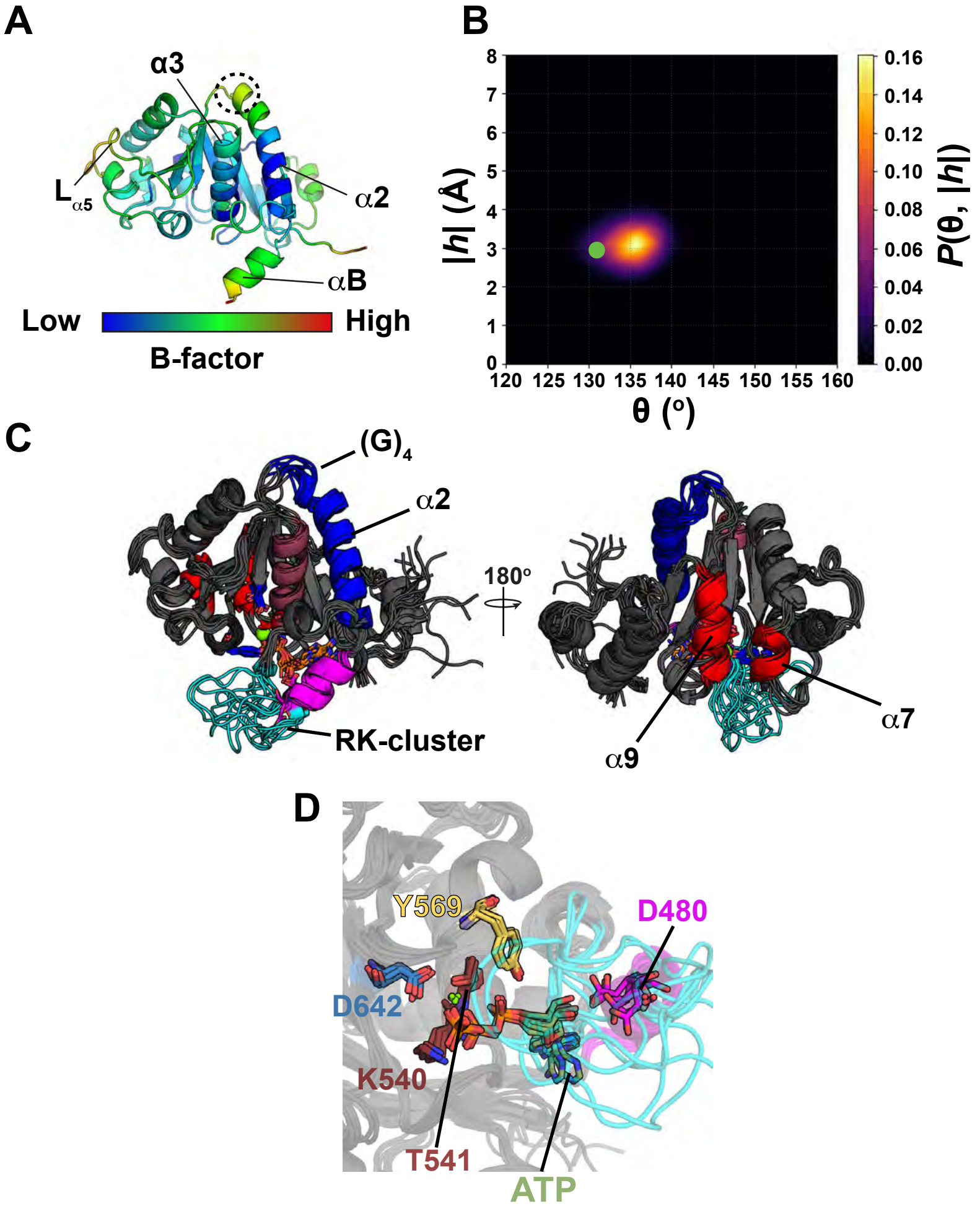
F_1-ATPase (1BMF)

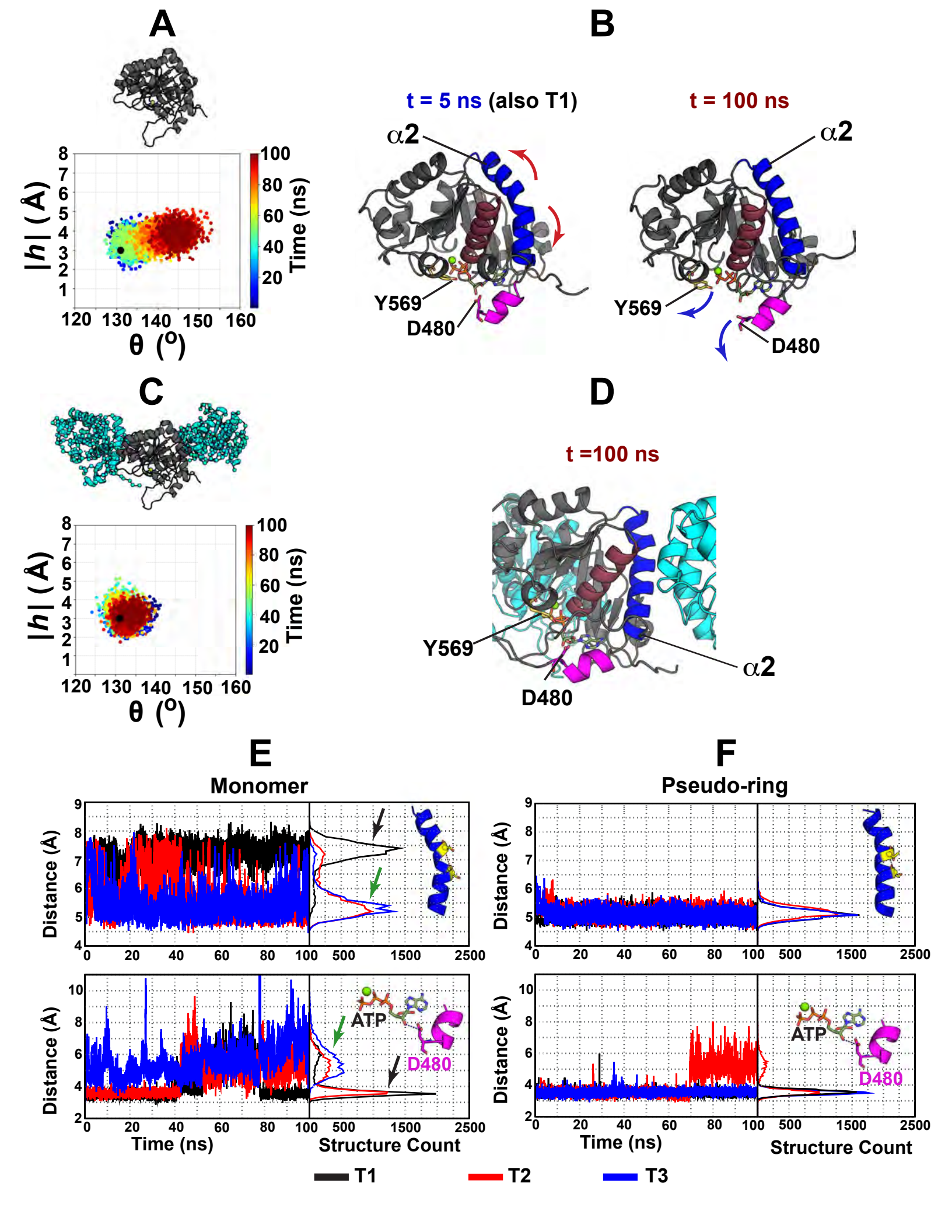

A
|h| (Å)
θ (°)
P(θ, |h|)
OS
CS
CS1
CS2
CS3
Y569
D480
Mg2+
ADP
OS
Y569
D480
Mg2+
ADP
B
|h| (Å)
θ(°)
P(θ, |h|)
C1
C2
C4
α3
α2
ADP
Y569
αB
L α5
α4
α3
90°
α2
α8
L α5
α9
α7
C
C544
Y569
D642
T541
D480
K540
ADP
C544
Y569
D642
T541
D480
Mg2+
K540
ADP
D
Density
Distance (Å)

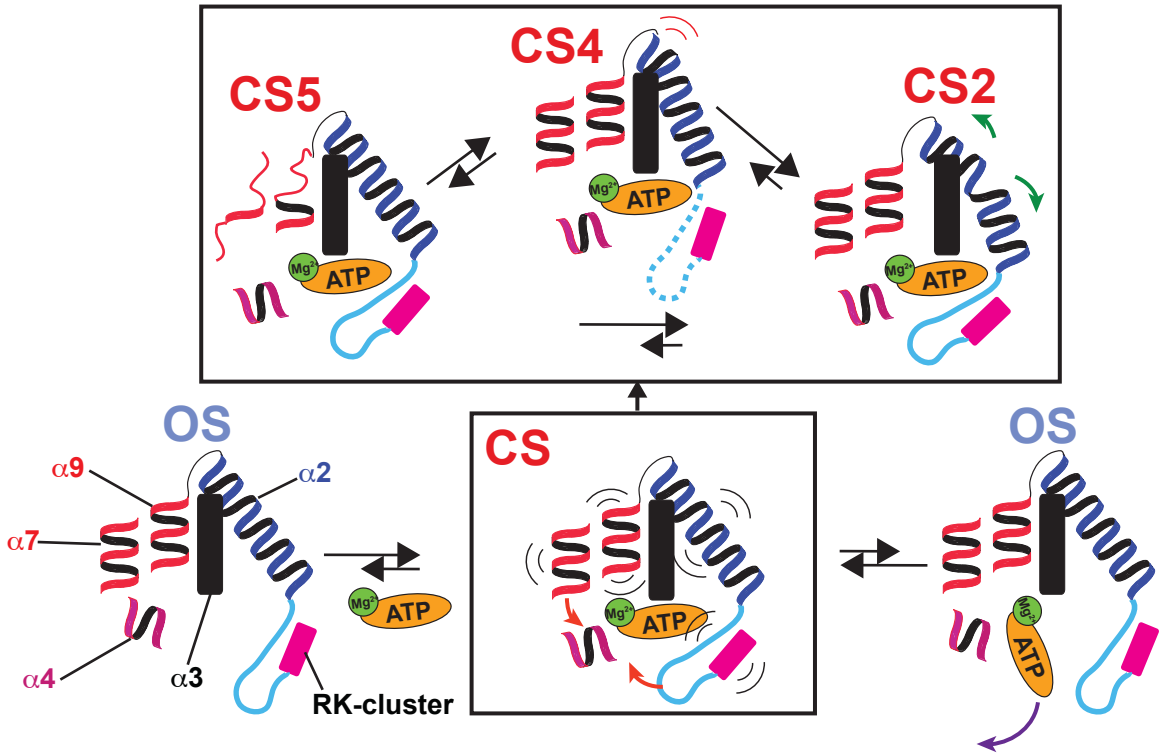

A
Time (min)
µcal/sec
kcal/mol
Molar Ratio
ADP
K_d= 0.3 ± 0.1 µM
ADP•Mg2+
K_d= 1.8 ± 0.5 µM
AMPPCP•Mg2+
K_d=148 ± 43 µM

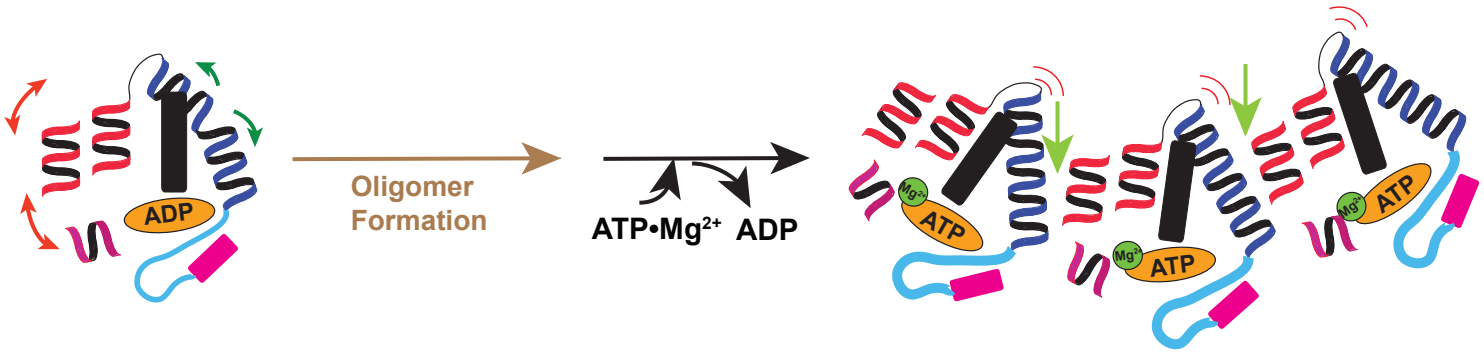

B
RK
α2
α3
α4 α5
α7α8
α9
ADP
ADP•Mg2+
Δδ (ppm)
Residue Number


C
α3
α2
α4
S512
RK-cluster
180°
α8
α5
α9
α7
C544


A
α3
L α5
α2
αB
Low
High
B-factor
B
|h| (Å)
θ (°)
P(θ, |h|)
C
(G)4
α2
180°
RK-cluster
α9
α7
D
Y569
D480
D642
K540
T541
ATP


A
B
t = 5 ns (also T1)
t = 100 ns
α2
Y569
D480
|h| (Å)
θ (°)
Time (ns)
C
D
t =100 ns
E
Monomer
F
Pseudo-ring
Distance (Å)
ATP
Time (ns)
Structure Count
T1
T2
T3


CS5
CS4
CS2
Mg2+
ATP
OS
CS
OS
α9
α2
α7
α4
α3
RK-cluster


ADP
Oligomer
Formation
ATP•Mg2+ ADP
Mg2+
ATP
Mg2+
ATP
Mg2+
ATP

SUPPLEMENTARY MATERIAL

Long-range dynamic correlations regulate the catalytic activity of the bacterial tyrosine kinase, Wzc

Fatlum Hajredini, Andrea Piserchio and Ranajeet Ghose

The environment variability analysis coupled to mean shift (EVA-MS) method

In order to perform the environment variability analysis (EVA) procedure on a structural ensemble, a multi-step approach was used. Using the crystal structure (chain A with the missing portion of the RK-cluster modeled in, see Materials and Methods for details) as reference, a set of nearest neighbor distances was determined for all residues within this structure using the following criteria: (1) for the amide nitrogen atom, N(*i*), of a given residue *i*, the carbonyl carbon atoms (C') of all non-flanking residues (i.e. excluding residues *i-1* and *i+1*) within 4 Å were identified and the corresponding distances with N(*i*) were determined; (2) for the C'(*i*) atom, the distances to the amide N atoms of all non-flanking residues within 4 Å were determined; (3) for the Cβ(*i*) atom, the distances of all Cβ atoms (Cα for glycine) for all non-flanking residues within 7 Å were determined. The first two sets of distances were chosen since they inform on the variation in backbone hydrogen-bonds and consequently of secondary structure within an ensemble; the third set of distances reflects sidechain packing. While distances involving additional sets of atoms may easily be included in the analysis, this three-bead coarse graining provides an adequate representation of local contacts in folded proteins at a reasonable computational cost. Note that excluding the glycine Cα and using a two-bead model for this residue had little influence on the results. These sets of distances, that define the local environment for each residue *i*, were stored in a reference distance map, $\chi_i(0)$, given by:

$$\chi_i(0) = \begin{pmatrix} i_1 & j_2 & d_0^{i_1,j_2} \\ i_1 & k_2 & d_0^{i_1,k_2} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ i_2 & l_1 & d_0^{i_2,l_1} \\ i_2 & m_1 & d_0^{i_2,m_1} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ i_3 & y_3 & d_0^{i_3,y_3} \\ i_3 & z_3 & d_0^{i_3,z_3} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix} \quad \text{(S1)}$$

The integers *j, k, l, m, y, z* (with *j, k, l, m, y, z* ≠ *i*-1 or *i*+1) index individual residues within the structure; the subscripts 1, 2 and 3 index backbone N, C' and sidechain Cβ (Cα for glycine) atoms, respectively. The interatomic distances e.g. $d_0^{i_1,j_2}$ were stored only if they were within the corresponding cutoff values mentioned above. Next, these *same* sets of distances were determined for all other structures (hereafter referred to as frames) within the ensemble and stored in corresponding distance maps, $\chi_i(t)$, one for each frame *t* (*t* > 0).

$$\chi_i(t) = \begin{pmatrix} i_1 & j_2 & d_t^{i_1,j_2} \\ i_1 & k_2 & d_t^{i_1,k_2} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ i_2 & l_1 & d_t^{i_2,l_1} \\ i_2 & m_1 & d_t^{i_2,m_1} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ i_3 & y_3 & d_t^{i_3,y_3} \\ i_3 & z_3 & d_t^{i_3,z_3} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix} \quad \text{(S2)}$$

The distances contained in the third column of the maps defined in Eq. S1 and Eq. S2, may be considered to represent the co-ordinates of a "pseudo-atom" in the *N*-dimensional space of contacts (the environment) for a given residue *i,* with *N* being the total number of contacts for that residue, as defined above, i.e. the number of rows in the $\chi_i(t)$. This pseudo-atom exists only in the positive orthant of this *N*-dimensional space. Using this representation, a variability function, V_i (*t*), can then be defined as:

$$V_i(t) = \frac{1}{N}\sum_{i_p,j_q \in \chi_i(0)}^{N} \sqrt{\left(\gamma_{i_p j_q}(t) \ - \gamma_{i_p j_q}(0)\right)^2} \quad \text{(S3)}$$

where *p*, *q* =1, 2, 3; |*p*-*q*|=1 for *p*=1, 2 and |*p*-*q*|=0 for *p*=3. $\gamma_{i_p j_q}(t)$ is a sigmoidal function that is given by:

$$\gamma_{i_p j_q}(t) = \frac{1-\left(\frac{d_t^{i_p,j_q}}{d_0^{i_p,j_q}}\right)^6}{1-\left(\frac{d_t^{i_p,j_q}}{d_0^{i_p,j_q}}\right)^{12}} \quad \text{(S4)}$$

Where $d_t^{i_p,j_q}$ represents the distance between atoms i_p and j_q in frame *t* (*t* > 0). $\gamma_{i_p j_q}(t)$ is similar in form to the switching functions introduced before (*44, 45*). This definition of $V_i(t)$ ensures that it is not dominated by very large deviations of individual contacts from the corresponding reference as would be the case if a Euclidian distance measure e.g. $\Delta d_t^{i_p,j_q} = \sqrt{(d_t^{i_p,j_q})^2 - (d_0^{i_p,j_q})^2}$ were used instead. Indeed, use of the latter is one of the drawbacks of using measures such as dRMSD values to estimate variability (*46*). While $V_i(t)$ has a limiting value of 0.5 (see Fig. S4A), this value is unlikely to be reached in practice since it would require all distances involving a particular residue to have very large deviations from the corresponding reference. Formally, $V_i(t) \to 0.5$ when all $d_t^{i_p,j_q} \to \infty$ (Eqs. S3, S4) suggesting a complete disengagement of residue *i* from the core; the alternative of all $d_t^{i_p,j_q} \to 0$ is not plausible. $V_i(t)$ reaches its minimum value of 0 when $d_t^{i_p,j_q} = d_0^{i_p,j_q}$ and $\gamma_{i_p j_q}(t) = \gamma_{i_p j_q}(0) = 0.5$ (see Fig. S4A).

$V_i(t)$ values provide a measure of all contacts (as per our earlier definition), and therefore of the local environment, for residue *i* in frame *t* relative to the reference i.e. frame 0. According to its definition in Eq. S3, $V_i(t)$ represents an average over all contacts for a given residue. The fact that individual $\gamma_{i_p j_q}(t)$, used in the calculation of $V_i(t)$, decay rapidly with increasing distance (see Fig. S4A), ensures that the $V_i(t)$ averaged over all i_p-j_q pairs depend strongly on the *number* of contacts that significantly deviate relative to the corresponding reference for a given frame *t* i.e. they are especially sensitive to a change in the co-ordination number across frames. Further, the fluctuations of $V_i(t)$ across the frames may be considered to measure the frame-dependent displacement of the *N*-dimensional pseudo-atom introduced above. This interpretation is useful in the analysis of changes in the local environment of individual residues and to identify correlated motions between sets of residues as described in the main text (also see Figs. S4B-E for an example of the analysis of correlated dynamics using $V_i(t)$ values).

In order to assess the contribution of the individual $V_i(t)$ represented by Eq. S3 relative to all residues in a given structure, the following per-residue scaled variability can be defined as

$$V_{i,scaled}(t) = \frac{e^{\kappa[V_i(t) - \overline{V(t)}]} - 1}{\sum_{i=1}^{M} e^{\kappa[V_i(t) - \overline{V(t)}]} - 1} \quad \text{(S5)}$$

Where *M* is the total number of residues in a structure. $\overline{V(t)}$ in Eq. S5 defines the average fluctuation for all residues over the entire structure in frame *t*. Thus, $\overline{V(t)}$ would be biased towards the most dominant modes of variation within the structure. Provided that a majority of the residues over all structures that comprise the ensemble (i.e. for all *t*) undergo small fluctuations rather than very large rearrangements, then $\overline{V(t)}$, to a large extent, defines the structural noise represented by these small fluctuations. The exponential scaling of the deviation of each $V_i(t)$ from the average ensures that the $V_{i,scaled}(t)$ are further biased in the favor of residues with significant variations with respect to the average for that frame i.e. have a largest number of distances in $\chi_i(t)$ that deviate substantially from the corresponding ones in $\chi_i(0)$. The degree of this bias may be empirically tuned through the factor κ, taken be to be 10 in the present case. Next, this procedure is repeated and the overall per-residue variability across all frames, $V_{i,norm}$, is obtained as

$$V_{i,norm} = \frac{\sum_{t=1}^{F} V_{i,scaled}(t)}{\sum_{i=1}^{M}\sum_{t=1}^{F} V_{i,scaled}(t)} \quad \text{(S6)}$$

F is the total number of frames. Thus, like the $V_{i,scaled}(t)$ used to calculate them, the $V_{i,norm}$ values have a weak dependence on small fluctuations that do not significantly alter the local contacts of a given residue. Instead, the $V_{i,norm}$ values are exquisitely sensitive to the reorganization of the local environment across residues and across frames. Thus, the $V_{i,norm}$ values, that represent the fractional contribution of a given residue to the total variability across the ensemble, provide an excellent measure of conformational variability.

An additional advantage of the use of $V_{i,norm}$ is that it provides a simple way to parse regions with different degrees of variability and assign a ranking or hierarchy using an iterative approach. This entails repeating the entire procedure starting with the construction of the $\chi_i(0)$ and subsequently of each $\chi_i(t)$, and ending with the calculation of $V_{i,norm}$, after removing residues that satisfy the following condition: $V_{i,norm} \geq \gamma \left|V_{i,norm}\right|_{max}$ from all frames (a value of 0.5 was used for γ in the present case); $\left|V_{i,norm}\right|_{max}$ represents the maximum value after an application of the EVA procedure. This procedure helps to reduce, in a rigorous fashion, the obscuring effects of

the most highly varying clusters of residues e.g. disordered loops and termini, revealing those with lesser degrees of variability.

In order to achieve clustering and analysis of multi-dimensional probability density functions (PDFs) involving $V_i(t)$ values of the most highly varying residues based on $V_{i,norm}$ analysis, an unsupervised machine learning approach based on the mean shift (MS) algorithm (*47*) was used. MS determines regions of high density in an *N*-dimensional data set. MS is commonly used in machine learning tasks involving visual-perception and feature space decompostion (*48-50*). In all cases EVA-MS clustering was conducted using the LPCM library (*51, 52*) in the R programing environment using a bandwidth parameter for the MS of 0.005. A simple illustration of MS clustering on $V_i(t)$ using a 2-dimensional PDF is shown in Fig. S4E.

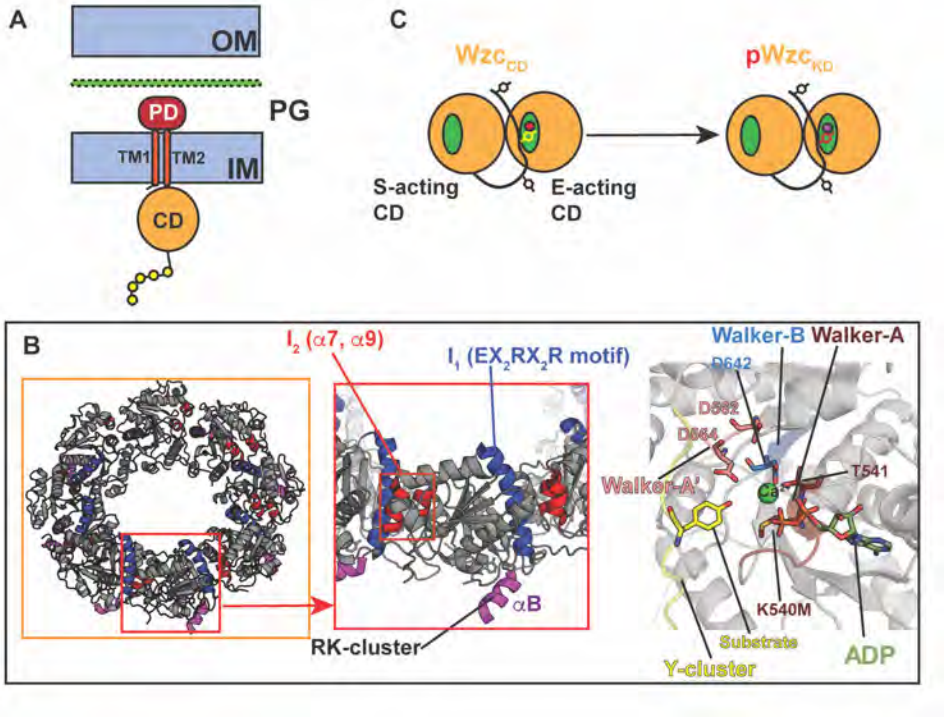


Fig. S1. Representation of a typical BY-kinase from a Gram-negative bacterium.
(A) Schematic representation of the domain organization of a BY-kinase. The BY-kinase localizes to the inner membrane (IM). The outer membrane, (OM) and the peptidoglycan layer (PG), are also shown schematically. The kinase comprises a periplasmic domain (PD, red), a two-pass transmembrane domain (with transmembrane helices, TM1 and TM2, shown in dark orange), a cytosolic catalytic domain (CD, shown in light orange), and a C-terminal tail comprising of several tyrosine residues (shown as yellow spheres) forming the Y-cluster. BY-kinases function as part of a multiprotein complex that spans both the inner and outer membranes. **(B)** The isolated CD of *E. coli* (K12) Wzc (Wzc_{CD}) forms an octameric ring in a front-to-back arrangement (PDB: 3LA6, left panel). An expansion of the inter-monomer interfaces (middle panel) shows the first interface (I_1) comprising helix α2 (dark blue) that carries a conserved $^{508}EX_2RX_2R^{514}$ motif. I_1 interacts with helices α7 and α9 that form the second interface (I_2, red) on a neighboring protomer. Also shown is the RK-cluster (purple) that comprises a helical N-terminus (αB) but is disordered, with missing electron density, at its C-terminus. The right panel displays the catalytic site and the conserved catalytic elements therein. The Walker-A (dark red; ^{533}**G**VSPSI**GKT**^{541}), Walker-A' (salmon; ^{558}VLLI**D**C**D**M**R**^{566}) and Walker-B (blue; ^{639}VLI**D**TP**P**^{645}) motifs are shown; sidechains of key residues are shown (underlined on the corresponding sequence, conserved residues are in bold) and labeled. The active site contains a Ca^{2+} ion (green) *in lieu* of the catalytic Mg^{2+}. Also shown is a substrate tyrosine residue (yellow) from the Y-cluster of the adjoining monomer inserted into the active site. **(C)** Intermolecular auto-phosphorylation of a Y-cluster tyrosine (yellow) of a substrate-acting (S-acting) Wzc_{CD} inserted into the active-site of an enzyme-acting (E-acting) Wzc_{CD}, is illustrated schematically. The process generates a phosphorylated tyrosine (orange) at the expense of ATP (red), with a concurrent conversion of the latter to ADP (brown).

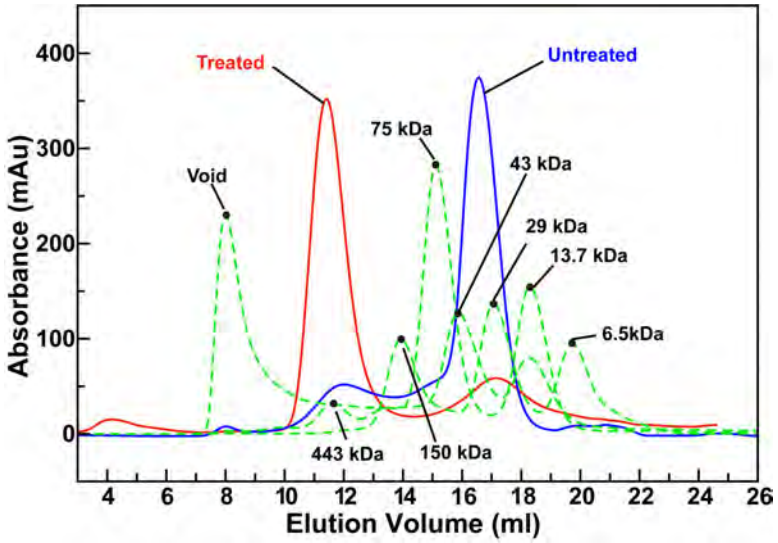


Fig. S2. Oligomeric states of Wzc_{CD}.
Wzc_{CD} is heterogeneously phosphorylated on the Y-cluster when expressed in *E. coli* and is largely monomeric in solution in the highly phosphorylated state (blue, untreated) even at high concentrations. Y-cluster dephosphorylation by incubation with λ-phosphatase results in a shift in population towards a high molecular weight oligomer (red, treated). Gel-filtration traces using a Superdex 200 10/300 (GE Healthcare Biosciences) are shown. Also shown are traces for molecular weight markers (green dashed lines). The λ-phosphatase-treated species appears at a molecular weight that is ~1.6-times larger than that expected for an octamer. This is likely to be the result of the unique hydrodynamic properties of the flat, disc-shaped structure.

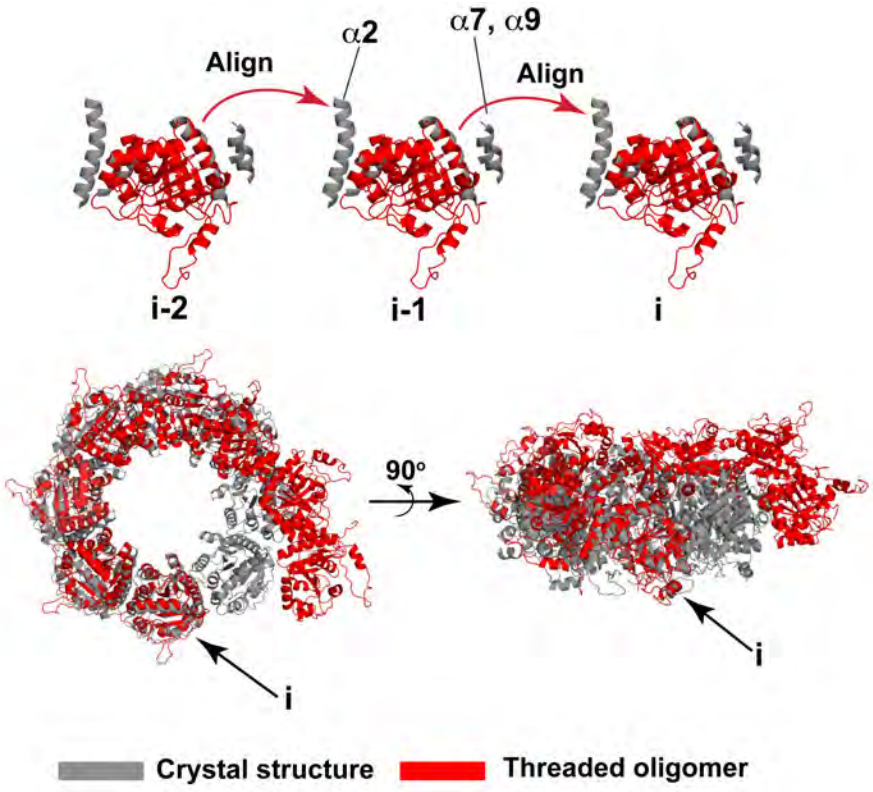


Fig. S3. Demonstration of the incompatibility of the OS obtained in the REST2 simulations of apo-Wzc$_{CD\Delta C}$ with the geometry of the octameric ring.
The top panel illustrates the procedure used to thread a representative structure from the OS into the crystallographic octamer. A small segment comprising α2 (I_1: 504-525) and its contacting α7 and α9 (I_2: residues 615-622 and 650-656) on the adjacent monomer was generated from the crystal structure (grey helices). Two copies of this segment were aligned to a representative structure of the OS (red) either through I_2 giving rise to a dangling I_1, or through I_1 to give rise to a dangling I_2. Eight copies of this composite were generated and threaded via the dangling parts in the following way: α2 of monomer i-1 was aligned to the dangling α2 of monomer i. Next, the procedure was repeated where the α2 of the i-2 monomer was aligned to the dangling α2 of the monomer i-1. This process was repeated 8 times to generate the structure shown in red in the bottom panel. This procedure ensured that the relative orientation of I_1 and I_2 (and their respective contacts) are the same as in the crystal structure. As is obvious from the bottom panel, an open and twisted octamer (in red) results if all constituent monomers are in the OS. This demonstrates that the OS is not compatible with the closed octameric ring (in grey) seen in the crystal structure of Wzc$_{CD}$. However, unlike for the CS, obtained in the other simulations, no drastic distortions (bending of α2 or unfolding of α7/α9) are seen in the OS.

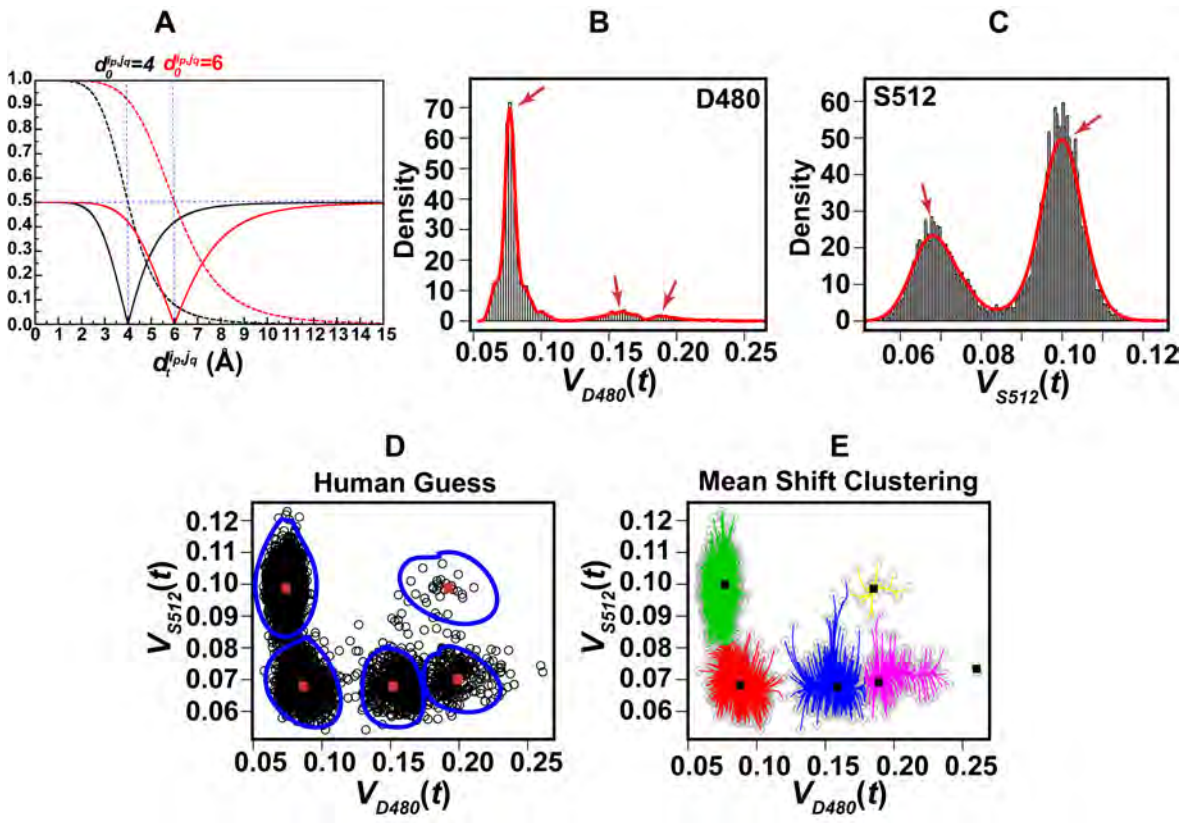


Fig. S4. Illustration of the EVA-MS approach.
(A) Plots illustrating the variation of $\gamma_{i_p j_q}$ (dashed lines; see Eq. S4) and V_i (solid lines; see Eq. S3) with $d_t^{i_p,j_q}$ for two different values of $d_0^{i_p,j_q}$ (4 Å in black and 6 Å in red). When $d_t^{i_p,j_q} = d_0^{i_p,j_q}$, $\gamma_{i_p j_q}$ = 0.5 and V_i reaches its minimum value of 0. V_i is asymmetric about $d_t^{i_p,j_q} = d_0^{i_p,j_q}$ with a steep increase towards its maximum value of 0.5 for $d_t^{i_p,j_q} < d_0^{i_p,j_q}$ and a more gradual increase for $d_t^{i_p,j_q} > d_0^{i_p,j_q}$. **(B)-(E)** A simple illustration of the EVA-MS approach using the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations as an example. Residues D480 **(B)** and S512 **(C)** were determined to be highly varying through the application of the EVA procedure (high $V_{i,norm}$ values). The $V_i(t)$ values for D480 and S512 display multimodal distributions that are indicative of discrete states being sampled in each case. **(D)** A 2-dimensional projection of the joint probability distribution function (PDF) involving $V_i(t)$ values of D480 and S512 parses into distinct clusters whose borders and centers can be easily visually identified (blue ellipses and red squares, respectively) at least in a qualitative sense. These regions are also detected (and quantitatively characterized) using the mean shift (MS) algorithm and are indicated by different colors **(E)**. Visual detection beyond 3-dimensions i.e. the calculation of joint PDFs for four or more residues, is not possible, even qualitatively. The MS algorithm is able to cluster and detect modes and the corresponding central tendencies in a general N-dimensional PDF. The advantage of the MS algorithm over other clustering methods e.g. k-means is that it does not require the number of clusters to be known *a priori*.

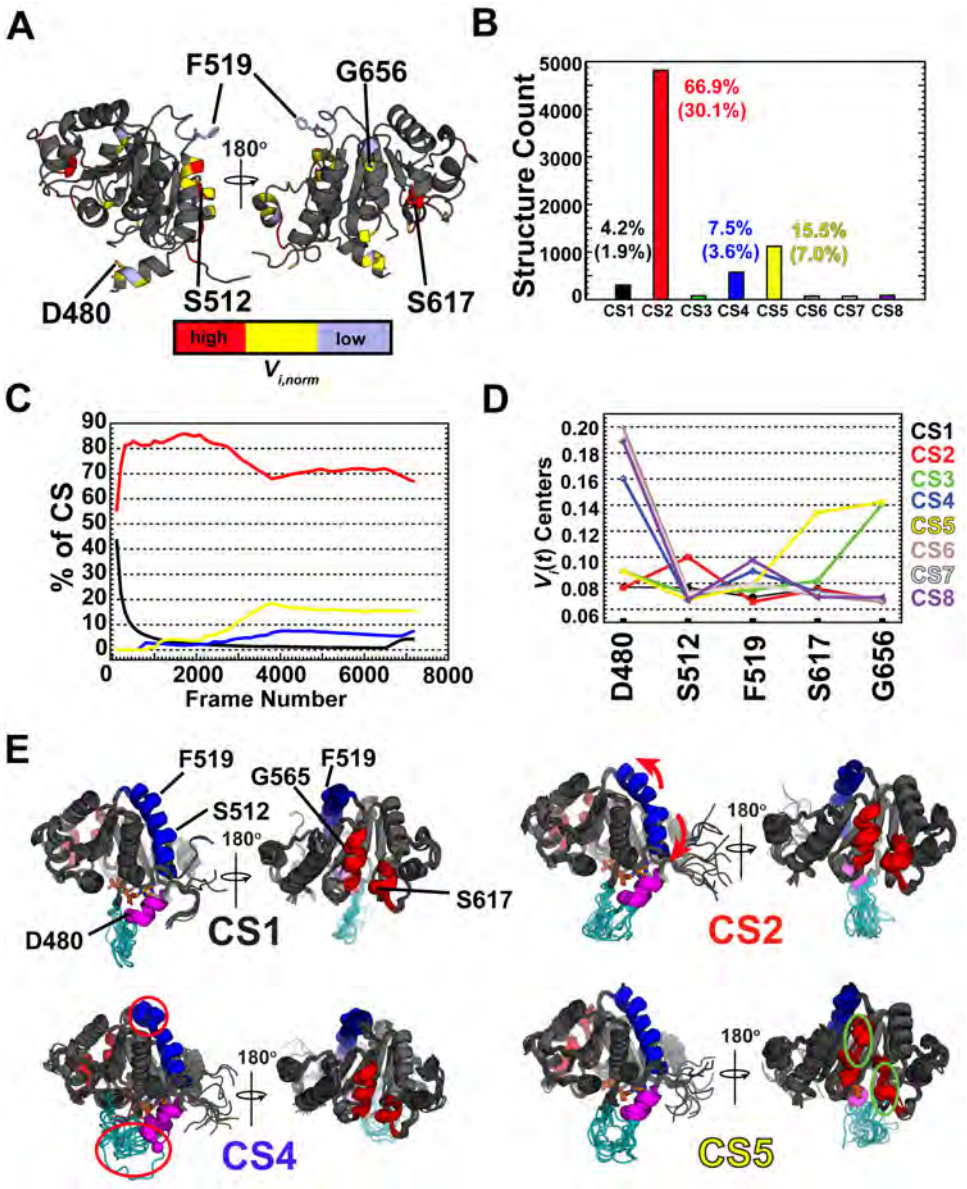


Fig. S5. EVA-MS decomposition of the CS of $Wzc_{CD\Delta C}$•ATP•Mg^{2+}.
EVA-MS decomposition of the conformations that comprise the CS (see Fig. 2A) in the REST2 simulations of $Wzc_{CD\Delta C}$•ATP•Mg^{2+}; the crystal structure was used as reference in the analysis. **(A)** The most varying residues (based on $V_{i,norm}$ values, excluding termini and loops, see Eq. S6) are colored red (EVA iteration 5), yellow (EVA iteration 6) and light-blue (EVA iteration 7), in order of their relative degree of variance, on a ribbon representation of a Wzc_{CD} monomer. Residues that show multimodal behavior (indicative of parsing into discrete states) in their $V_i(t)$ distributions were found to be D480 (RK cluster, αB), S512 (α2), F519 (the C-terminal end of α2), S617 (α7) and G656 (α9). The $V_i(t)$ values of these residues were used to construct a 5-dimensional PDF. Decomposition of the PDF using MS clustering produced 47 distinct clusters of which only 4, clusters CS1, CS2, CS4 and CS5, account for ~94% of all structures within the CS. **(B)** The number of structures in each of the 8 major clusters (CS1-CS8) is shown as a bar-plot.

The percent population of each cluster within the CS is indicated adjacent to the corresponding bar plot; the percent population over the entire ensemble is shown in parentheses. **(C)** Evolution of the populations of the clusters CS1, CS2, CS4 and CS5 expressed as percent of the total structures within the CS. The x-axis denotes the frame number in chronological order (note that this is discontinuous since the analysis includes structures found within the CS only rather than in the entire ensemble). The percent occurrence is calculated every 1 ns (i.e. for every 100 frames). The relative stabilization of the populations suggests adequate sampling i.e. convergence. **(D)** 2-dimensional projections of the major clusters decomposing the contributions of each of the 5 residues. The x-axis indexes the residue, and the y-axis denotes the projection of the 5-dimensional PDF onto each residue. Thus, each line represents a unique cluster and the value across the various residues indicate regions where structural evolution has occurred. For example, CS1 (black) is characterized by low values of $V_i(t)$ across all residues i.e. it is very similar to the crystal structure. CS2 (red) is characterized by evolution on S512 (α2) but is otherwise similar to the crystal structure. **(E)** Superposition of 10 randomly selected structures from each of CS1, CS2, CS4 and CS5 are shown. The probe residues are shown as spheres. CS1 is very similar to the crystal structure. CS2 is characterized by a bending about the center of α2 (red arrows, probed by the S512). CS4 is characterized by joint evolution about D480 (RK-cluster) and F519 (C-terminal end of α2) and the structures show a highly disordered RK-cluster and a high degree of distortion at the C-terminal end of α2 (red circles). CS5 is characterized by joint evolution in S617 (α7) and G656 (α9). The corresponding structures indicate that in this conformation, α7 and α9 are partially unfolded (green circles). It is evident that use of this approach allows the simple identification of the correlation (e.g. potential allosteric coupling) between remote regions.

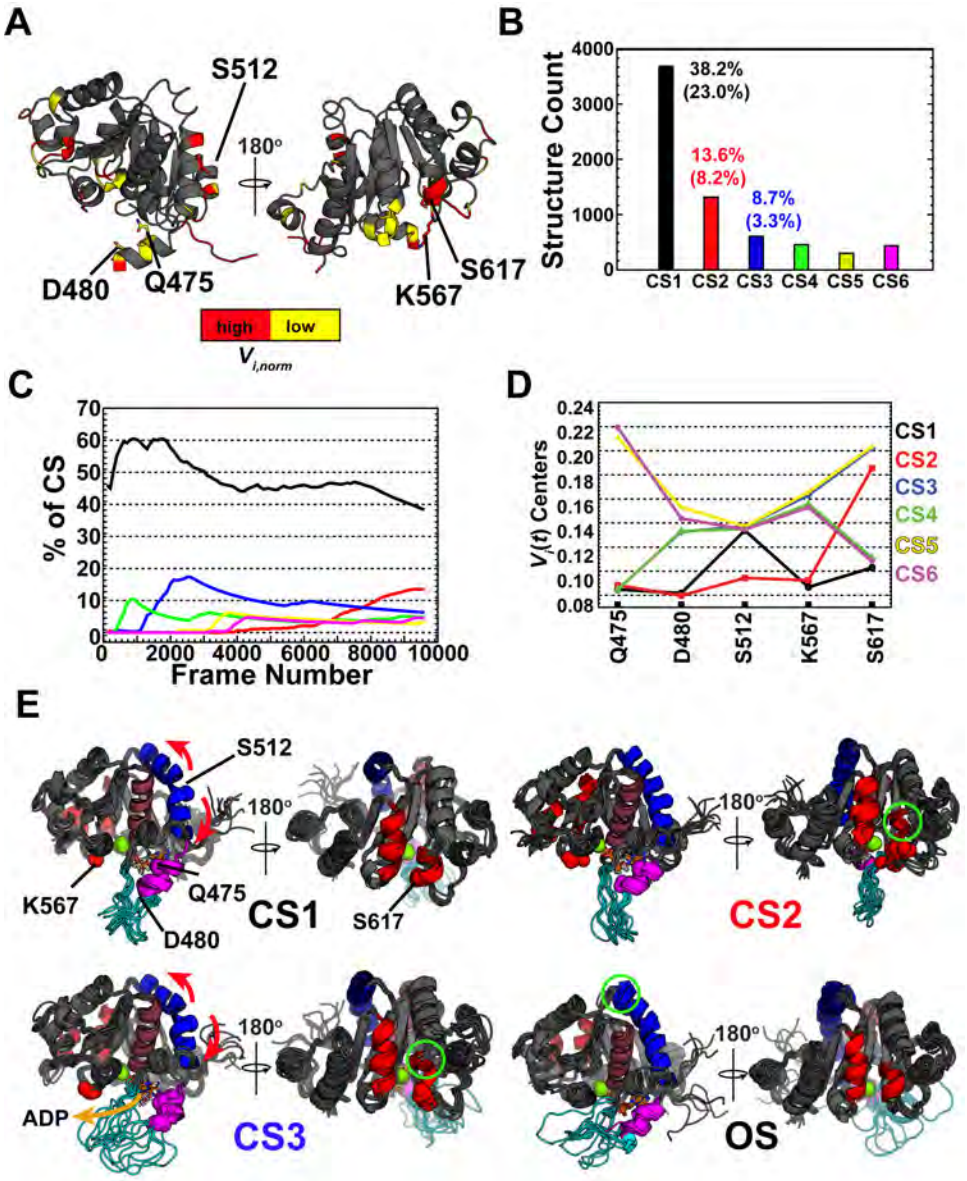


Fig. S6. EVA-MS decomposition of the CS of $Wzc_{CD\Delta C}$•ADP•Mg^{2+}.
EVA-MS decomposition of the conformations that comprise the CS (see Fig. 4A) in REST2 simulations of $Wzc_{CD\Delta C}$•ADP•Mg^{2+}; the crystal structure was used as reference in the analysis. **(A)** EVA determined top varying residues (disregarding loops and termini) are shown in red (EVA iterations 2 and 3) and yellow (EVA iteration 4). Residues with multimodal $V_i(t)$ distributions Q475 and D480 (αB of the RK-cluster), S512 (α2), K567 ($L_{\alpha 4}$), and S617 (α7) were used to construct a 5-dimensional PDF. **(B)** MS clustering produces a total of 607 clusters of which only 6 are significantly populated; the percentage of structures within each of these six major clusters (CS1-CS6) of the CS are shown as in Fig. S5B. **(C)** Evolution of the percent occupancy of each of the 6 major clusters are shown as in Fig. S5C. **(D)** Decomposition of the contribution of each residue to the 6 major clusters. **(E)** Conformations of 10 randomly selected structures from CS1, CS2 and CS3 and the OS are shown. CS1, the most populated CS cluster that accounts of ~38% of all structures in the CS, shows structural evolution on α2 characterized by a bending at S512

(red arrows). The second most populated cluster (CS2) is characterized by an unfolding of α7 (green circle; reported by S617). CS3 shows large contributions across all residues, except Q475. In this conformation, there is bending of α2 (red arrows), partial unfolding of α7 (green circle), disorder in the RK-cluster and the disengagement of ADP (orange arrow). Also shown are structures from the OS that are characterized by disorder in the RK-cluster and at the C-terminal end of α2 (green circle).

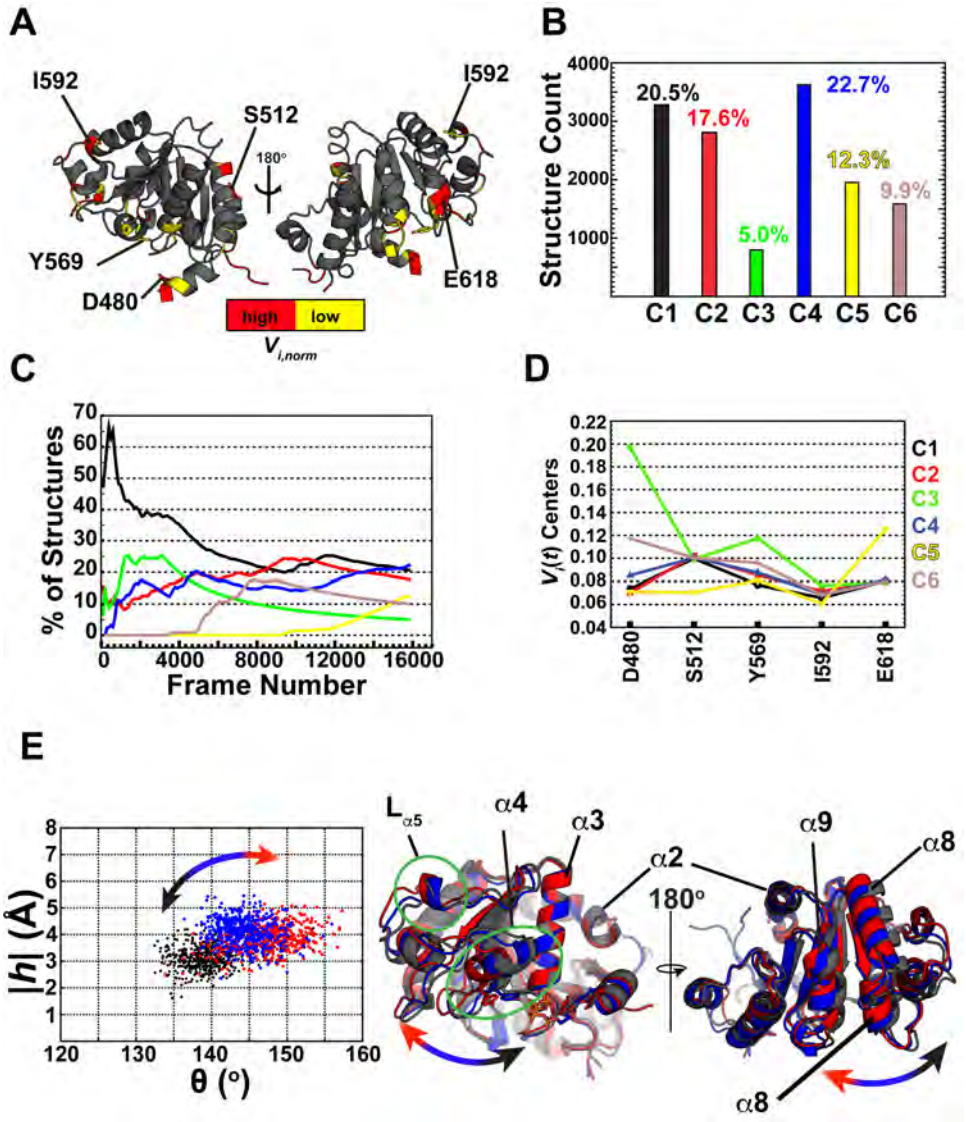


Fig. S7. EVA-MS decomposition of the Wzc_{CDAC}•ADP REST2 ensemble.
(A) EVA based determination of most varying residues are shown in red (EVA iteration 3) and yellow (EVA iteration 4). Residues with multimodal $V_i(t)$ distributions were found to be D480 (αB in the RK-cluster), S512 (α2), Y569 (α4), I592 ($L_{\alpha5}$), and E618 (α7). **(B)** EVA-MS clustering of the 5-dimensional PDF yielded 6 major clusters (95 clusters in total) that accounted for ~88% of structures in the ensemble. **(C)** Percentage of occurrence of each cluster as a function of frame number calculated as in Fig. S5C. **(D)** 2-dimensional decomposition of the major clusters showing the contributions of the probe residues. The three largest clusters (C4, C1, C2 in descending order of occupancy) are characterized by the bending of α2 and differ only slightly in the structural evolution across the RK-cluster, α4, and $L_{\alpha5}$. Cluster C5 is characterized by the unfolding of α7. **(E)** Projection of 500 structures drawn randomly from each of the clusters C1 (black), C2 (red), and C4 (blue) onto the cylindrical coordinate system. Representative structures superimposed on α3 are shown on the right. It can be seen that the structures belonging to C1, C2 and C4 display somewhat different rigid body fluctuations of α2, $L_{\alpha5}$, α7, and α8 relative to α3 and thus occupy different positions, on average, in $|h|$-θ space.

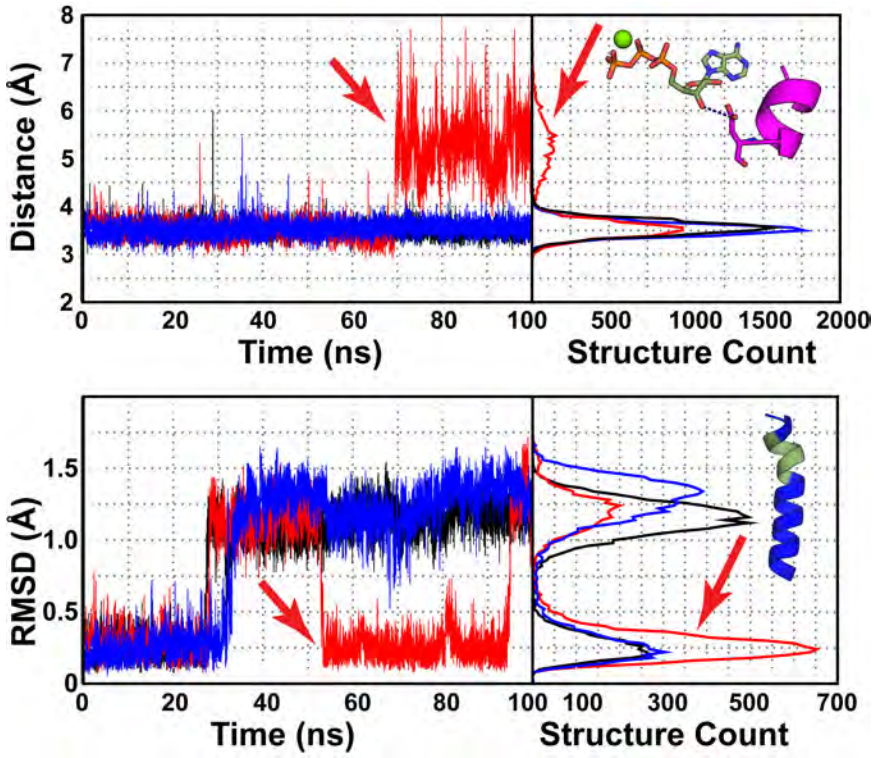


Fig. S8. **The degree of disorder at the C-terminus of α2 correlates with the optimal engagement of ATP.**
Unlike in the case of the monomer (see Fig. 7), the contact between ATP and D480 is maintained throughout the entire 100 ns in two of the three trajectories of the ATP•Mg^{2+} complex of the pseudo-ring (top panel, black and blue trajectories). The latter half of these trajectories are characterized by increased disorder at the C-terminus of α2 (bottom panel) as indicated by the increased RMSD for the Cα atoms of residues 517-522 (colored green in the ribbon shown on the right panel). It is notable that the C-terminus of α2 is also disordered in the crystal structure of Wzc_{CD} (Fig. 6A). In one of the trajectories (red), there is re-introduction of order at the C-terminus of α2 (as evident from the reduced RMSD values); this increased order results in some degree of disengagement of ATP from D480 (indicated by the red arrows).

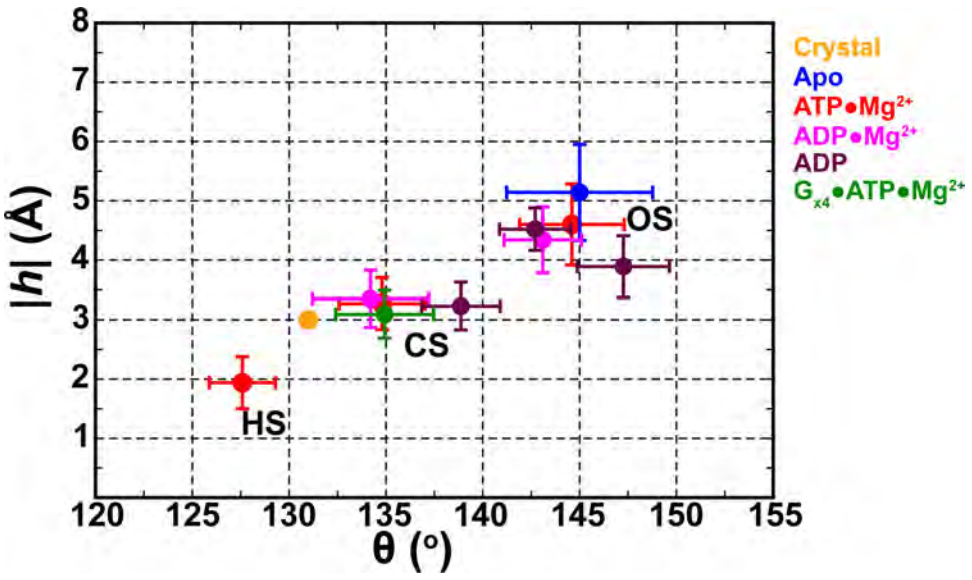


Fig. S9. Summary of major global conformational states seen in all REST2 simulations. Apo-$Wzc_{CD\Delta C}$ (blue), $Wzc_{CD\Delta C}$•ATP•Mg^{2+} (red), $Wzc_{CD\Delta C}$•ADP•Mg^{2+} (pink), $Wzc_{CD\Delta C}$•ADP (dark red; in this case, clusters C1, C2 and C4 are shown, see Fig. S7 for details) and $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+} (green). Also shown for comparison is the conformation seen in the crystal structure (orange dot). The centers of the distributions and the corresponding standard deviations (widths) are indicated by the filled circles and error bars, respectively. The apo simulations only populate the OS and the $Wzc_{CD\Delta C}$-$(G)_4$•ATP•Mg^{2+} simulations only populate the CS. While CS represents a crystal-like state with the proper co-ordination of ATP and Mg^{2+}, it does not represent a fully active conformation. The HS that is only populated in the $Wzc_{CD\Delta C}$•ATP•Mg^{2+} simulations likely represents a reactive conformation. In addition to the appropriately aligned catalytic elements, the HS also includes an ordered RK-cluster and R490 therein that likely serves in a role that is equivalent to the signature lysine of a deviant Walker-A motif or the arginine-finger in the AAA^+ family of NTPases.