Kinetic modulation of a disordered protein domain by phosphorylation

Nathaniel Stanley[1], Santiago Esteban-Martín[2,*], and Gianni De Fabritiis[1,3,*]

[1]Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain.

[2]Joint BSC-IRB Research Programme in Computational Biology, Barcelona Supercomputing Center - BSC, Jordi Girona, 29, 08034 Barcelona, Spain.

[3]Institució Catalana de Recerca i Estudis Avançats, Passeig Lluis Companys 23, 08010 Barcelona, Spain

* Corresponding authors: gianni.defabritiis@upf.edu, santiago.esteban.martin@gmail.com

Phosphorylation is a major post-translational mechanism of regulation that frequently targets disordered protein domains, but it remains unclear how phosphorylation modulates disordered states of proteins. Here we determine the kinetics and energetics of a disordered protein domain, the kinase-inducible domain (KID) of the transcription factor CREB and that of its phosphorylated form pKID, using high-throughput molecular dynamics simulations. We identify the presence of a metastable, partially ordered state with a 60-fold slowdown in conformational kinetics that arises due to phosphorylation, kinetically stabilizing residues known to participate in an early binding intermediate. We show that this effect is only partially reconstituted by mutation to glutamate, indicating that the phosphate is uniquely required for the long-lived state to arise. This mechanism of kinetic modulation could be important for regulation beyond conformational equilibrium shifts.

Post-translational modifications in proteins participate in many fundamental cellular processes[1–3] and affect at least one-third of all eukaryotic proteins[4,5]. In particular, protein phosphorylation plays key roles in many signal transduction processes[4–8], preferentially targeting intrinsically disordered protein domains[4,5] (IDPs). IDPs remain largely unstructured under native conditions[9], resembling random-coil polymers akin to the unfolded states of proteins. Present in more than 50% of eukaryotic proteins[10,11], IDPs perform a plethora of biological functions[12–14] and are commonly associated with a variety of human diseases[15–18].

The effect of phosphorylation in proteins is manifold. For example, it can induce conformational changes[19], promote order-disorder transitions[20], and modulate binding via electrostatic interactions with partners[21]. However, it remains largely unexplored whether phosphorylation can regulate the conformational kinetics of proteins, and what effect this may have on molecular recognition with other partners. This is especially interesting for protein domains that do not have a native folded conformation, but also difficult to address as it requires the characterization of lowly populated, transient states.

In an attempt to understand how phosphorylation modulates disordered states of proteins, we determined the conformational kinetics and energetics of a disordered protein domain before and after phosphorylation at atomic resolution. We chose to study an experimentally well-characterized disordered fragment of the KID domain of transcription factor CREB[22] (residues 116-147). The KID domain includes the binding motif to the KIX domain of the coactivator CBP[23] (Fig. 1b). KID is known to be disordered in solution and to form two alpha helices upon binding[23], a process that involves binding intermediates[24]. Molecular recognition of KID is regulated via phosphorylation of serine 133 in the αB helix, which increases its binding affinity 40-fold[25] (binding of residues 119-147 to KIX). Interestingly, phosphorylation barely affects the fraction of folded αA and αB helices[26]. Regulation of binding affinity is mostly ascribed to phosphate electrostatic and hydrogen bonding interactions with KIX. However, mutation of serine 133 to a negatively charged residue such as glutamate (often considered to mimic interactions with amide NH, lysine and arginine residues[27]) cannot recapitulate pKID activity[28].

All-atom, explicit solvent micro to millisecond MD simulations[29–31] have in recent years significantly contributed to our understanding of disordered states of proteins[29,30,32,33]. Computational studies on KID protein using replica exchanged implicit solvent MD simulations[34] and short all-atom MD simulations[35] showed that KID is largely unstructured and phosphorylation barely affects its helical propensity. Moreover, coarse grain and short, high temperature all-atom simulations suggested that binding of KID to KIX initiates at the αB helix [35–38]. However, difficulties in resolving the micro to millisecond timescale using all-atom, unbiased MD simulations impeded the assessment of the effect of phosphorylation on long-time scales and thus on conformational kinetics.

Here, we use all-atom, explicit solvent molecular dynamics (MD) simulations and resolve equilibrium by performing 1.7 milliseconds of aggregated simulation time on the distributed computing network GPUGRID.net[39] with ACEMD[40]. We identify a metastable, partially ordered state with a 60-fold slowdown in conformational kinetics, exchanging at the multi-microsecond time regime, present only in phosphorylated KID. This post-translational modification kinetically locks residues in helix αB that participate in an early binding intermediate, while no significant change in the population of the ordered state was observed. Our results are in line with previous experimental NMR analyses. We suggest that long-lived states favored by phosphorylation are partially responsible for the increased binding affinity of pKID to KIX and propose that kinetic modulation of disordered protein domains may be an important mechanism of regulation by phosphorylation at the biochemical level beyond conformational equilibrium shifts.

Results

Induction of a slow conformational exchange process by phosphorylation. In order to study the effect of phosphorylation on the KID domain of CREB we performed all-atom, explicit solvent molecular dynamics simulations with 1.7 milliseconds of total aggregated data for the following systems: phosphorylated KID at serine 133 (pKID), KID domain, and a S133E mutant (gKID) used as a control (Fig. 1; see Methods). For each system, the data was analyzed by constructing a Markov state model of the entire ensemble of trajectories (see Methods). Markov state models have been used previously to successfully calculate several slow processes from ensemble MD simulations[41,42,43,44,45]. We used inter-residue Cα-Cα distances and ϕ/ψ backbone dihedral angles as general metrics to build the kinetic model. The conformational space was discretized into one thousand clusters and then projected on a 5-dimensional space using a new dimensionality reduction technique (time-sensitive independent component analysis, or TICA[46]), which identifies the slow coordinates in the dynamics.

The characteristic relaxation timescales of the first and second slowest processes for pKID, KID and gKID proteins, as determined from the Markov model, are reported in Table 1. The slowest relaxation process of pKID and KID differs by 60-fold (37 μs vs 0.6 μs). The second slowest process of pKID is also an order of magnitude slower than the slowest process observed in KID. In order to assess whether that slowdown in conformational kinetics was specific of the phosphate, we mutated the serine 133 to glutamate to partially mimic the negatively charged phosphate group[27]. The slowest process in gKID was found well below that of

pKID (Table 1). Therefore, phosphorylation induced a slow exchange process in the multi-microseconds time regime, a kinetic fingerprint of pKID that could not be recapitulated by S133E mutation.

Identification of a metastable state in phosphorylated KID. As the slowest process for all systems is well separated from the rest, we used a simple, two-state kinetic model to analyze conformational transitions in KID, called D (disordered) and O (ordered),

$$D \underset{1/\tau_{-1}}{\overset{1/\tau_1}{\rightleftharpoons}} O \qquad (1)$$
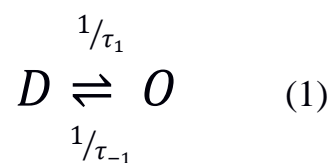
where τ_1 and τ_{-1} are the forward and backward mean first passage times. The relaxation time (τ_{ex}) is defined as $\tau_{ex} = (\tau_1^{-1} + \tau_{-1}^{-1})^{-1}$. We obtained a kinetic clustering over just two states using the Perron-cluster cluster analysis (PCCA)[47]. The slowest exchange process identified in pKID, that occurs at 37 ± 1 μs (Table 1), had a forward (τ_1) and backward (τ_{-1}) transition time of 463 ± 22 and 37.0 ± 0.8 μs (Fig. 2, Supplementary Fig. 1). The stationary (equilibrium) populations obtained from the Markov model are $95 \pm 1\%$ and $5 \pm 1\%$. Statistical errors were determined using a bootstrapping procedure of the data (see methods).

For the case of non-phosphorylated KID we found a large state that accounted for most of the population ($99 \pm 1\%$, Fig. 2). The slowest transitions in KID occurred at 0.6 ± 0.01 μs (Table 1), with forward and backward transition times of 5.7 ± 0.2 and 0.6 ± 0.1 μs (Fig. 2, Supplementary Fig. 1). Control simulations with gKID (S133E) protein also showed a largely populated state ($99 \pm 1\%$) and a slow exchange process of 1.8 ± 1 μs (Supplementary Fig. 1). Phosphorylation of KID resulted in a 60-fold slowdown exchange process ($\tau_{ex} \sim 37$ μs) as compared to KID domain ($\tau_{ex} \sim 0.6$ μs) and affected the forward (state D to state O) and backward (state O to state D) transitions times by a factor of 80 and 60, respectively.

Phosphorylation kinetically locks key residues involved in binding. To further investigate the slowdown observed on conformational exchange in pKID as compared to both KID and gKID proteins, we computed the autocorrelation function for helix folding and unfolding at the residue level and determined the characteristic relaxation times (Fig. 3a, see Methods, Supplementary Fig. 3). We found that residues 131-136 at the N-terminal part of αB helix undergo slow multi-microsecond conformational exchange ($\tau_{ex} \sim 40$ μs) and, therefore, participate in the slowest process observed in pKID (Table 1). NMR experiments identified an early intermediate complex in which N-terminal αB region is

engaged with KIX[24]. This region maps on top of residues undergoing slow exchange in phosphorylated KID (Fig. 3a). It is plausible, therefore, that the state O identified in this work and the early binding intermediate detected by NMR have common structural details.

Structural analysis. As expected for an intrinsically disordered protein domain, the most populated state in pKID, KID and gKID is largely unstructured (Supplementary Fig. 2). We observed an increase in the population of the minor state from $1 \pm 1\%$ to $5 \pm 1\%$ when serine 133 was phosphorylated, and no change upon S133E mutation (Fig. 2). In all cases the state was characterized by an increase in helical content of residues in αB helix (Supplementary Fig. 2). This minor state is referred herein as ordered or state O. This effect was most notable for the case of pKID, in which ca. 80% of the conformers in this state had the N-terminal residues 131-135 of αB helix folded. An interesting structural feature of states O and D is that they group conformations in which residue 131 is either forming part of the αB helix or unfolded, respectively. This structural signature separates states that exchange slowest in all three systems. Interestingly, NMR relaxation experiments showed that residue 131 in pKID folds one to two orders of magnitude slower than the rest of residues in the complex with KIX[24]. The helicity of residues in αB helix tends to increase in this state with respect to the disordered state D (Supplementary Fig. 2). Overall, the slow exchange processes involved disordered (state D) - ordered (state O) transitions.

Inspection of state O in pKID revealed additional features beyond secondary structure formation (Fig. 3b). The phosphate moiety, located within residues undergoing slow exchange, serves as a staple that locks pKID into a conformation with the αA helix partially folded (ca 35%) and the N-terminal of αB helix mostly folded (ca 80%) (Supplementary Fig. 2). In addition, a turn bridging helix αA and αB is held by electrostatic interactions and hydrogen bonds between the phosphate in helix αB and lysine and arginine residues in the C-terminal region of αA helix. Note that similar states are also observed for KID and gKID proteins, but these states exchange fast with the rest of the conformations.

Other types of secondary structural arrangements were also present, such as beta sheets, but these were lowly populated (<0.1%) and exchanged rapidly with other states (faster than 500 nanoseconds for the case of pKID, for example).

Comparison to NMR. We assessed the structural properties of our equilibrium simulations by a direct comparison against backbone chemical shifts (CS) measured by NMR spectroscopy for KID and pKID proteins[26]. CS allows mapping secondary structure propensities in disordered states of proteins. Overall,

results were found in agreement with experimental CS (Fig. 4, $C\alpha/C\beta$ RMSD$_{NMR-MD}$ for pKID and KID is 0.6/0.4 and 0.5/0.4 ppm, error in the prediction algorithm is ca 0.5 ppm for SHIFTX2[48] and ca 1 ppm for SPARTA+[49]). A deviation between calculated and measured $C\alpha$ CS was observed for residues 120-128 (Fig. 4a) indicating an increased population of helix αA present under experimental conditions (288K versus 315K used for simulations). However, there is a known α-helical induction of structure in KID at low temperatures[50], so the difference might be due to this or to limitations of the force fields. We note that CS could not be calculated for pSer133, which was then mutated to serine, affecting prediction of neighboring residues 132 and 134. Comparison of CS changes induced by phosphorylation (Fig. 4b) showed a similar pattern between calculated and measured CS. These findings suggested common structural re-arrangements of KID upon phosphorylation as determined by simulation and NMR.

Experimental NMR measurements detected similar inter-proton NOE distance patterns for the phosphorylated and non-phosphorylated forms of the domain except in the vicinity of the phosphorylation site. Six weak or very weak NOEs that involved residues 127, 131, 134 and 137 were only observed for pKID[26]. Overall, NOE distances are similar for KID and pKID (4.8±0.6 Å). NOEs involving I127H^N-R130H^N and I131H^N-R133H^N nuclei showed smaller distances for KID (Δ_{dist}= 0.3±0.1 Å). NOE distances involving residues R131Hα-Y134H^N and Y134Hα-I137H^N were by contrast smaller for the case of pKID (Δ_{dist} = 0.5±0.1 Å). These additional NOEs bridged up to 7 residues. This NOE pattern suggested an increase in helical content near the N-terminus of the αB helix, in agreement with NMR.

Discussion

We have described the conformational kinetics and energetics of an intrinsically disordered protein domain, the kinase-inducible domain (KID) of the transcription factor CREB, before and after a post-translational phosphorylation. We identified the presence of a metastable, partially ordered state with at least 60-fold slowdown in conformational exchange that arises due to phosphorylation, involving folding and unfolding of residues in the N-terminal region of αB helix. Previous NMR studies investigating conformational dynamics of KID in its unbound state[24] can detect exchange processes in the ca. 0.3–10 ms time window[51] and, therefore, the process we observe in this work (tens of microseconds) remained hidden.

Phosphorylation induced a minor shift (in absolute terms) in the equilibrium distribution of folded αA and

αB helices, in agreement with NMR data[24]. However, an 80- and 60-fold increase in the folded and unfolded residence time for residues at the N-terminal αB helix was observed. These N-terminal αB residues were found by NMR experiments to participate in an early binding intermediate[24]. Binding of kinetically locked regions can give other transient interactions time to form, potentially increasing the number of productive binding events and thus the overall affinity for the binding partner. Similarly, kinetically locked regions that participate in molecular recognition could frustrate the unbinding process. This would result in a shift of the binding equilibrium towards the bound state.

We devised a toy kinetic model to show that, at least in principle, a slowdown in conformational kinetics of both forward and backward transition rates by the same amount can translate into a 40-fold increase in binding affinity (Fig. 5). In an analogy to the binding mechanism of KID (D) to KIX (X), in our model the bound complex (XD) is formed via an intermediate[24]. A key element of the model is the presence of productive (XD_p) and nonproductive (XD_{np}) intermediates, for which phosphorylation induces a 100-fold slowdown in conformational exchange. It is therefore reasonable to propose that the slowdown in conformational exchange we observed offers an additional mechanism to increase binding affinity. The fact that the glutamate mutation cannot sustain long-lived states and disrupts[25] the binding is consistent with this proposed mechanism of modulation.

Previous studies have shown that pSer133 contributes to binding via specific intermolecular contacts, and a loss of affinity is seen when mutating specific residues in KIX[25,28,52]. The exact atomistic mechanism and whether these contacts are also sufficient alone for binding is less clear. In our kinetic model, multiple routes are possible. A mutation destabilizing some important contacts would reduce the residence time of the bound form ($1/k_{-5}$ in the kinetic diagram). The effect can be easily simulated, and a 10 times increase in k_{-5} corresponds to 10 times lower affinity. However, the additional kinetic route is also mathematically viable, i.e. the binding affinity can be modulated controlling the time spent in productive vs non-productive intermediates. Both routes can play a role in general, and usually there is a complementary effect in such complex processes. The fact that it is generally possible to modulate binding of a disordered domain in a kinetic way is novel and practical. In our case, it was indeed suggested by the fact that we see a slow-down in conformational kinetics of pKID which is not seen in gKID.

It has also been proposed that pSer133 could reduce the conformational entropy of the unbound state of KID[34,53]. This is compatible with the results presented here.

Phosphorylation of KID could change conformational entropy due to altered kinetics, but compensate in conformational enthalpy producing similar overall populations.

With disordered domains taking part in more than 50% of proteins responsible for signaling in the cell[54], this kinetic mechanism of modulation highlights a possible further mechanism by which post-translation modifications may affect disordered domains and their interactions with binding partners.

Methods

Simulation setup. The sequence corresponding to residues 116-147 of CREB was used in this work (Fig. 1). Initial coordinates were built using VMD software[55]. CHARMM22*[56] force field was used for the peptides. The phosphate group on pKID (bearing two negative charges) was patched using that provided in the CHARMM27[57] force field with serine parameters from CHARMM22*. TIP3P model was used for the water molecules.

The peptides were placed in a cubic water box of 64-angstrom sides and equilibrated for 2 ns at 315K in the NPT ensemble at 1 bar. The peptides were simulated at 500K for 120 nanoseconds to decorrelate from the initial conformation at constant volume (no cis conformations of peptide bonds were observed). For each of the three simulated systems, 100 starting structures were taken from each nanosecond of the last 100 ns of the high temperature run. For each of those 100 starting structures, 10 replicas were submitted to GPUGRID.net[39] using ACEMD[40]. An additional 200 simulations were later performed for each system to ensure adequate sampling of the process described, for a total of 1200 for each system. Production runs lasted 480 ns and were performed at 315K in the NVT ensemble. For all simulations, the particle mesh Ewald (PME) algorithm[58], rigid hydrogen bonds, hydrogen mass repartitioning[59], and a time step of 4 fs were employed.

Markov state model analysis. We built a Markov state model from the molecular simulations trajectories. MSMs have been successfully used to reconstruct the equilibrium and kinetic properties in a large number of molecular systems[41,44,60–62]. By determining the frequency of transitions between conformational states we construct a master equation which describes the dynamics between a set of conformational states. Relevant states are determined geometrically by clustering the simulation data onto a metric space (e.g. contact maps). The projected space used in this work is formed by the distances between all 32 Cα pairs plus ϕ/ψ backbone dihedral angles of the KID peptide. This high-dimensional data was then projected into five dimensions using the time-sensitive independent component analysis

(TICA) method[46] with a 2 ns lag-time, which selects the slowest varying variables. The data was then clustered into 1000 states using the k-means clustering algorithm. Results remained consistent to changes in the number of clusters (e.g. 10^4) and on the projected dimensions (e.g. 10 dimensions).

The master equation is then built as

$$\dot{P}_i(t) = \sum_{j=1}^{N}\left[k_{ij}P_j(t) - k_{ji}P_i(t)\right] = K_{ij}P_j(t), \qquad (2)$$

where $P_i(t)$ is the probability of state i at time t, and k_{ij} are the transition rates from j to i, and $\mathbf{K} = (K_{ij})$ is the rate matrix with elements $K_{ij} = k_{ij}$ for $i \neq j$ and $K_{ii} = -\sum_{j \neq i} k_{ji}$. The master equation $d\mathbf{P}/dt = \mathbf{KP}$ has solution with initial condition $\mathbf{P}(0)$ given by $\mathbf{P}(t) = \mathbf{T}(t)\mathbf{P}(0)$, where we defined the transition probability matrix $T_{ij}(t)$ $= (\exp[\mathbf{K}t])_{ij} = p(i,t|j,0)$, i.e. the probability of being in state i at time t, given that the system was in state j at time 0. In practical terms, $p_{ij}(\Delta t)$ is estimated from the simulation trajectories for a given lag time Δt using a maximum likelihood estimator compatible with detailed balance[63]. The eigenvector $\boldsymbol{\pi}$ with eigenvalue 1 of the matrix $\mathbf{T}(\Delta t)$ corresponds to the stationary, equilibrium probability. Higher eigenvectors correspond to exponentially decaying relaxation modes[62] for which the relaxation timescale is computed by the eigenvalue as $\tau_s = \frac{\Delta t}{\log(\lambda_s)}$, where λ_s is to the largest eigenvalue above 1. For long enough lag times Δt the model will be Markovian, however every process faster than Δt is lost. Therefore the shortest lag is chosen for which the relaxation timescales do not show a dependence on the lag time Δt anymore (See Supplementary Fig. 1). In our case, we chose a 100 ns lag time as it showed the least dependence for the slowest processes in all systems. Furthermore, the initial N microstates can be lumped together into macrostates using kinetic information from the MSM eigenvector structure[47]. This allows one to obtain a limited number of important, kinetically distinct states. Mean first passage times and commitor probabilities can also be calculated to obtain the relevant kinetics of the system[64].

Autocorrelation analysis. A MSM-based trajectory of 10 ms effective time (100 ns lag time) was built for each system by sampling conformations according to transition probabilities between microstates (MSM trajectories are distributed upon request to the authors). Relaxation times were calculated by fitting exponentials to autocorrelation functions (ACF). The fitting procedure included three exponentials. We used a time constant < 100 ns for the first, which is the lag time used to build the kinetic model, to account for the initial fast decay in the ACF. The fitting was limited to the first 100 μs of the ACF. See Supplementary Methods for details.

Error estimation: We estimated errors for all properties using a bootstrapping technique. We performed 10 independent runs in which 20% of the trajectories were randomly eliminated and a new MSM was built before properties were recalculated.

NMR calculation. We calculated expected chemical shifts from the simulation using 10,000 frames of the MSM trajectory for each of the three simulations. Calculations with SHIFTX2[48] and SPARTA+[49] provided consistent results. Average NOE distances (r) were calculated from the MSM trajectories as $<r^{-6}>^{-1/6}$.

Kinetic model. We built a kinetic model based on five coupled reactions (Fig. 5 and Supplementary Methods). As with a previously described mechanism for KID binding to KIX[24], we used an intermediate step. Unlike previous models, our model contains an additional state for both the unbound and intermediate states, such that a binding competent and incompetent conformer (termed productive and non-productive, respectively) can be accounted for. We label the exchange between these states reaction 1 (rate $k_{\pm 1}$) in the unbound state and reaction 2 in the intermediate state (rate $k_{\pm 2}$). Kinetic rates for the other step reactions 3, 4 and 5 (rates $k_{\pm 3}$, $k_{\pm 4}$, $k_{\pm 5}$) were taken from those determined by NMR. The coupled equations were solved numerically using MATLAB. The equilibrium dissociation constant was calculated as $K_d = [D_{free}][X_{free}]/[XD]$.

References

1. Grotenbreg, G. & Ploegh, H. Chemical biology: dressed-up proteins. *Nature* **446,** 993–995 (2007).
2. Geiss-Friedlander, R. & Melchior, F. Concepts in sumoylation: a decade on. *Nat. Rev. Mol. Cell Biol.* **8,** 947–956 (2007).
3. Morrison, R. S. *et al.* Proteomic analysis in the neurosciences. *Mol. Cell. Proteomics MCP* **1,** 553–560 (2002).
4. Xie, H. *et al.* Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* **6,** 1917–1932 (2007).

5. Iakoucheva, L. M. *et al.* The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **32,** 1037–1049 (2004).

6. Hubbard, M. J. & Cohen, P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends Biochem. Sci.* **18,** 172–177 (1993).

7. Pawson, T. & Scott, J. D. Signaling through scaffold, anchoring, and adaptor proteins. *Science* **278,** 2075–2080 (1997).

8. Hunter, T. & Karin, M. The regulation of transcription by phosphorylation. *Cell* **70,** 375–387 (1992).

9. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293,** 321–331 (1999).

10. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. Intrinsic protein disorder in complete genomes. *Genome Inform. Workshop Genome Inform.* **11,** 161–171 (2000).

11. Oldfield, C. J. *et al.* Comparing and combining predictors of mostly disordered proteins. *Biochemistry (Mosc.)* **44,** 1989–2000 (2005).

12. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradović, Z. Intrinsic Disorder and Protein Function†. *Biochemistry (Mosc.)* **41,** 6573–6582 (2002).

13. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6,** 197–208 (2005).

14. Dyson, H. J. Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys.* **44,** 467–518 (2011).

15. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradović, Z. & Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **323,** 573–584 (2002).

16. Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K. & Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry (Mosc.)* **45,** 10448–10460 (2006).

17. Raychaudhuri, S., Dey, S., Bhattacharyya, N. P. & Mukhopadhyay, D. The Role of Intrinsically Unstructured Proteins in Neurodegenerative Diseases. *PLoS ONE* **4,** e5566 (2009).

18. Uversky, V. N. *et al.* Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* **10 Suppl 1,** S7 (2009).

19. Jiang, Z. G. & McKnight, C. J. A phosphorylation-induced conformation change in dematin headpiece. *Struct. Lond. Engl. 1993* **14,** 379–387 (2006).

20. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291,** 2429–2433 (2001).

21. Lubman, O. Y. & Waksman, G. Dissection of the energetic coupling across the Src SH2 domain-tyrosyl phosphopeptide interface. *J. Mol. Biol.* **316,** 291–304 (2002).

22. Chrivia, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365,** 855–859 (1993).

23. Radhakrishnan, I. *et al.* Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. *Cell* **91,** 741–752 (1997).

24. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447,** 1021–1025 (2007).

25. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R. & Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277,** 42241–42248 (2002).

26. Radhakrishnan, I., Pérez-Alvarado, G. C., Dyson, H. J. & Wright, P. E. Conformational preferences in the Ser133-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Lett.* **430,** 317–322 (1998).

27. Mandell, D. J. *et al.* Strengths of hydrogen bonds involving phosphorylated

amino acid side chains. *J. Am. Chem. Soc.* **129,** 820–827 (2007).

28. Parker, D. *et al.* Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation. *Mol. Cell* **2,** 353–359 (1998).

29. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **334,** 517–520 (2011).

30. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 5915–5920 (2013).

31. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134,** 3787–3791 (2012).

32. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17874–17879 (2013).

33. Henry, E. R., Best, R. B. & Eaton, W. A. Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17880–17885 (2013).

34. Ganguly, D. & Chen, J. Atomistic Details of the Disordered States of KID and pKID. Implications in Coupled Binding and Folding. *J. Am. Chem. Soc.* **131,** 5214–5223 (2009).

35. Chen, H.-F. Molecular Dynamics Simulation of Phosphorylated KID Post-Translational Modification. *PLoS ONE* **4,** e6516 (2009).

36. Turjanski, A. G., Gutkind, J. S., Best, R. B. & Hummer, G. Binding-Induced Folding of a Natively Unstructured Transcription Factor. *PLoS Comput Biol* **4,** e1000060 (2008).

37. Ganguly, D. & Chen, J. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. *Proteins* **79,** 1251–1266 (2011).

38. Umezawa, K., Ikebe, J., Takano, M., Nakamura, H. & Higo, J. Conformational Ensembles of an Intrinsically Disordered Protein pKID with and without a KIX Domain in Explicit Solvent Investigated by All-Atom Multicanonical Molecular Dynamics. *Biomolecules* **2,** 104–121 (2012).

39. Buch, I., Harvey, M. J., Giorgino, T., Anderson, D. P. & De Fabritiis, G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. *J. Chem. Inf. Model.* **50,** 397–403 (2010).

40. Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **5,** 1632–1639 (2009).

41. Sadiq, S. K., Noé, F. & Fabritiis, G. D. Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc. Natl. Acad. Sci.* 201210983 (2012). doi:10.1073/pnas.1210983109

42. Huang, X. *et al.* Constructing multi-resolution markov state models (msms) to elucidate RNA hairpin folding mechanisms. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 228–239 (2010).

43. Noé, F. *et al.* Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proc. Natl. Acad. Sci.* **108,** 4822 –4827 (2011).

44. Pan, A. C. & Roux, B. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129,** 064107 (2008).

45. Bowman, G. R., Huang, X. & Pande, V. S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods San Diego Calif* **49,** 197–201 (2009).

46. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139,** 015102–015102–13 (2013).

47. Deuflhard, P. & Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear Algebra Its Appl.* **398,** 161–184 (2005).

48. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50,** 43–57 (2011).

49. Shen, Y. & Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **48,** 13–22 (2010).

50. Hua, Q. X., Jia, W. H., Bullock, B. P., Habener, J. F. & Weiss, M. A. Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry (Mosc.)* **37,** 5858–5866 (1998).

51. Ban, D., Sabo, T. M., Griesinger, C. & Lee, D. Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra-tc Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy. *Molecules* **18,** 11904–11937 (2013).

52. Volkman, H. M., Rutledge, S. E. & Schepartz, A. Binding Mode and Transcriptional Activation Potential of High Affinity Ligands for the CBP KIX Domain. *J. Am. Chem. Soc.* **127,** 4649–4658 (2005).

53. Solt, I., Magyar, C., Simon, I., Tompa, P. & Fuxreiter, M. Phosphorylation-induced transient intrinsic structure in the kinase-inducible domain of CREB facilitates its recognition by the KIX domain of CBP. *Proteins Struct. Funct. Bioinforma.* **64,** 749–757 (2006).

54. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18,** 756–764 (2008).

55. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

56. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100,** L47–49 (2011).

57. Mackerell Jr., A. D., Feig, M. & Brooks III, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).

58. Harvey, M. J. & De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory Comput.* **5,** 2371–2377 (2009).

59. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20,** 786–798 (1999).

60. Buchete, N.-V. & Hummer, G. Peptide folding kinetics from replica exchange molecular dynamics. *Phys. Rev. E* **77,** 030902 (2008).

61. Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **108,** 10184–10189 (2011).

62. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134,** 174105–174105–23 (2011).

63. Bowman, G. R., Beauchamp, K. A., Boxer, G. & Pande, V. S. Progress and challenges in the automated construction of Markov state models for full protein systems. *J. Chem. Phys.* **131,** 124101 (2009).

64. Singhal, N., Snow, C. D. & Pande, V. S. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **121,** 415–425 (2004).

Author Contributions

N.S. and G.D.F. conceived of the experiment. N.S. prepared and performed the simulations. N.S. and S.E.M. performed the analysis. N.S., S.E.M, and G.D.F. wrote the manuscript.

System	1st slowest process (µs)	2nd slowest process (µs)
pKID	37.0 ± 0.8	6.0 ± 1.2
KID	0.6 ± 0.01	0.2 ± 0.01
gKID	1.8 ± 0.01	0.4 ± 0.01

Table 1 | Timescales of 1st and 2nd slowest processes for the systems studied. See also Supplementary Fig. 1. Errors were determined by a bootstrapping technique (see Methods).
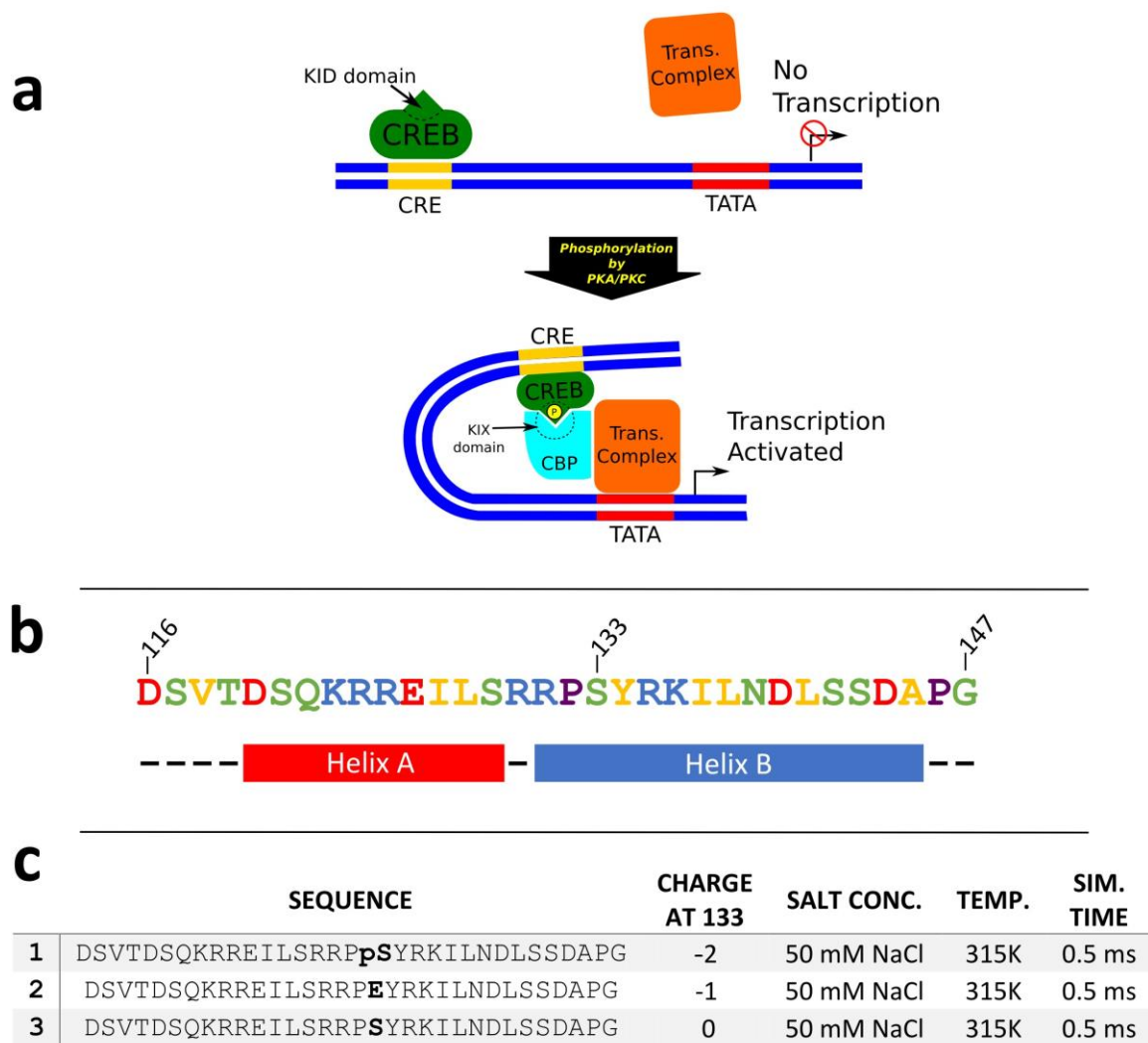
Figure 1 | System Overview. (a) Schematic representation of CREB activation. Upon phosphorylation of CREB at serine 133 at the kinase-inducible domain (KID), the KIX domain of the co-activator CBP binds to pKID with 100-fold increased affinity[25] and promotes transcriptional activation by recruiting the transcription machinery. (b) KID sequence with residues color-coded as positive (blue) and negatively (red) charged, polar (green) and hydrophobic (yellow). Helical regions formed in the complex with KIX[23] are shown. (c) Summary of simulations performed in this work.
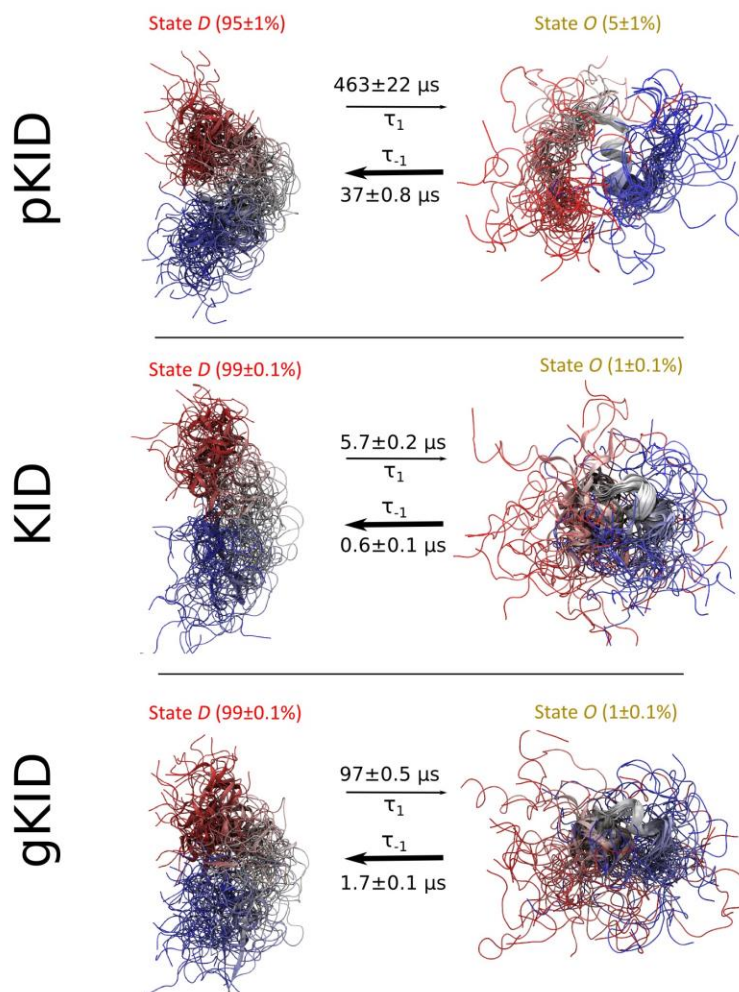
Figure 2 | Kinetic modulation of the energy landscape in a disordered domain via phosphorylation. Conformational exchange between state *D* (disordered) and state *O* (ordered) is shown together with forward (τ_1) and backward (τ_{-1}) transition times (see also Supplementary Fig. S1). Conformations were aligned using residues with the lowest Cα-Cα distance variance. Error estimates were determined using a bootstrap technique (see methods). N- to C-terminal is color coded red and blue.
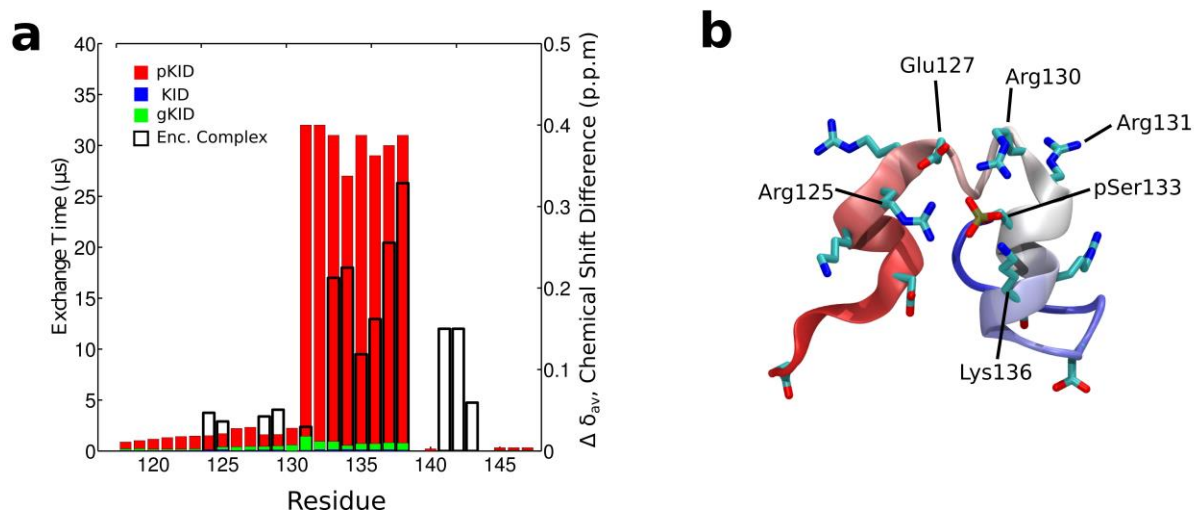
Figure 3 | Identification of residues undergoing slow conformational exchange in phosphorylated KID. a) Residue specific relaxation times of helix folding/unfolding process (filled bars) derived from autocorrelation functions. Empty bars show chemical shifts changes in pKID that map residues participating in an early binding intermediate as detected by NMR[24]. b) Example of a kinetically locked conformation in phosphorylated KID. The peptide backbone is shown in cartoon representation. Heavy atoms of phosphorylated serine 133 and several charged residues are shown as sticks.
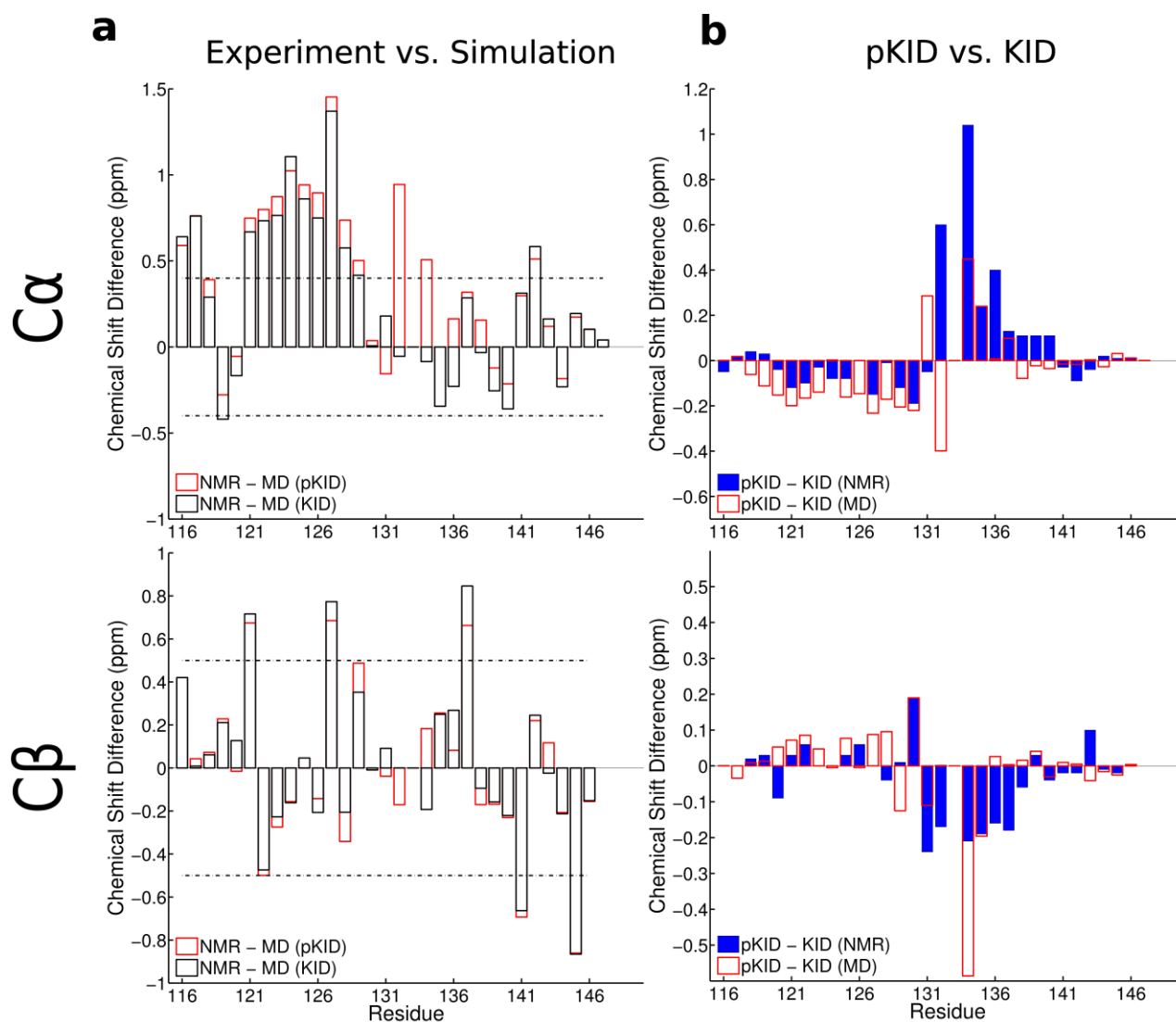
Figure 4 | Comparison against NMR data. a) Chemical shift (CS) difference between experimentally measured (NMR, 288K) and calculated (MD, 315K). Calculations were performed with SPARTA+[49] and SHIFTX2[48], which rendered equivalent results. Dashed lines correspond to the intrinsic error in the prediction estimated from SHIFTX2 (Cα and Cβ errors are 0.4 and 0.5 ppm, for SPARTA+ the estimated errors are ca 1ppm). Cα/Cβ RMSD for pKID and KID is 0.6/0.4 and 0.5/0.4 ppm. b) Experimental (NMR, blue) and calculated (MD, red) CS differences between pKID and KID systems. Note that calculations in pKID were performed with serine 133 instead of phosphorylated serine 133, affecting prediction of neighboring residues 132 and 134.
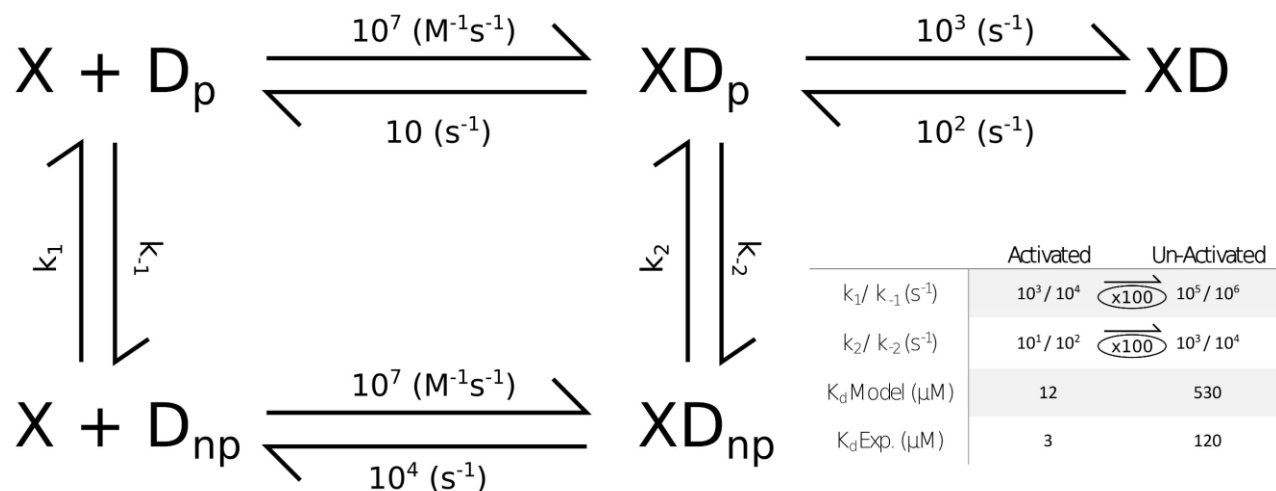
$$X + D_p \xrightleftharpoons[10 \ (s^{-1})]{10^7 \ (M^{-1}s^{-1})} XD_p \xrightleftharpoons[10^2 \ (s^{-1})]{10^3 \ (s^{-1})} XD$$

k_1 k_{-1} k_2 k_{-2}

$$X + D_{np} \xrightleftharpoons[10^4 \ (s^{-1})]{10^7 \ (M^{-1}s^{-1})} XD_{np}$$

	Activated		Un-Activated
k_1/ k_{-1} (s^{-1})	$10^3 / 10^4$	→x100	$10^5 / 10^6$
k_2/ k_{-2} (s^{-1})	$10^1 / 10^2$	→x100	$10^3 / 10^4$
K_d Model (μM)	12		530
K_d Exp. (μM)	3		120

Figure 5 | Illustrative example on the potential role of conformational kinetics in overall binding affinity. In this example unbound species exchange between two states, labeled D_p and D_{np}. Phosphorylation induces a 100-fold slowdown in conformational kinetics, affecting forward (k_1) and backward (k_{-1}) rates by the same amount. A key element of the model is binding via productive (XD_p) and nonproductive (XD_{np}) intermediates, for which phosphorylation induces a 100-fold conformational exchange slowdown (k_2 and k_{-2}). Unbinding of non-productive (XD_{np}) intermediate is faster than unbinding of XD_p (1000-fold, a leakage route). Exchange for productive and nonproductive intermediates is set 100 times slower that in their free state. This example results in 40-fold increased binding affinity for activated (e.g. phosphorylated) protein.
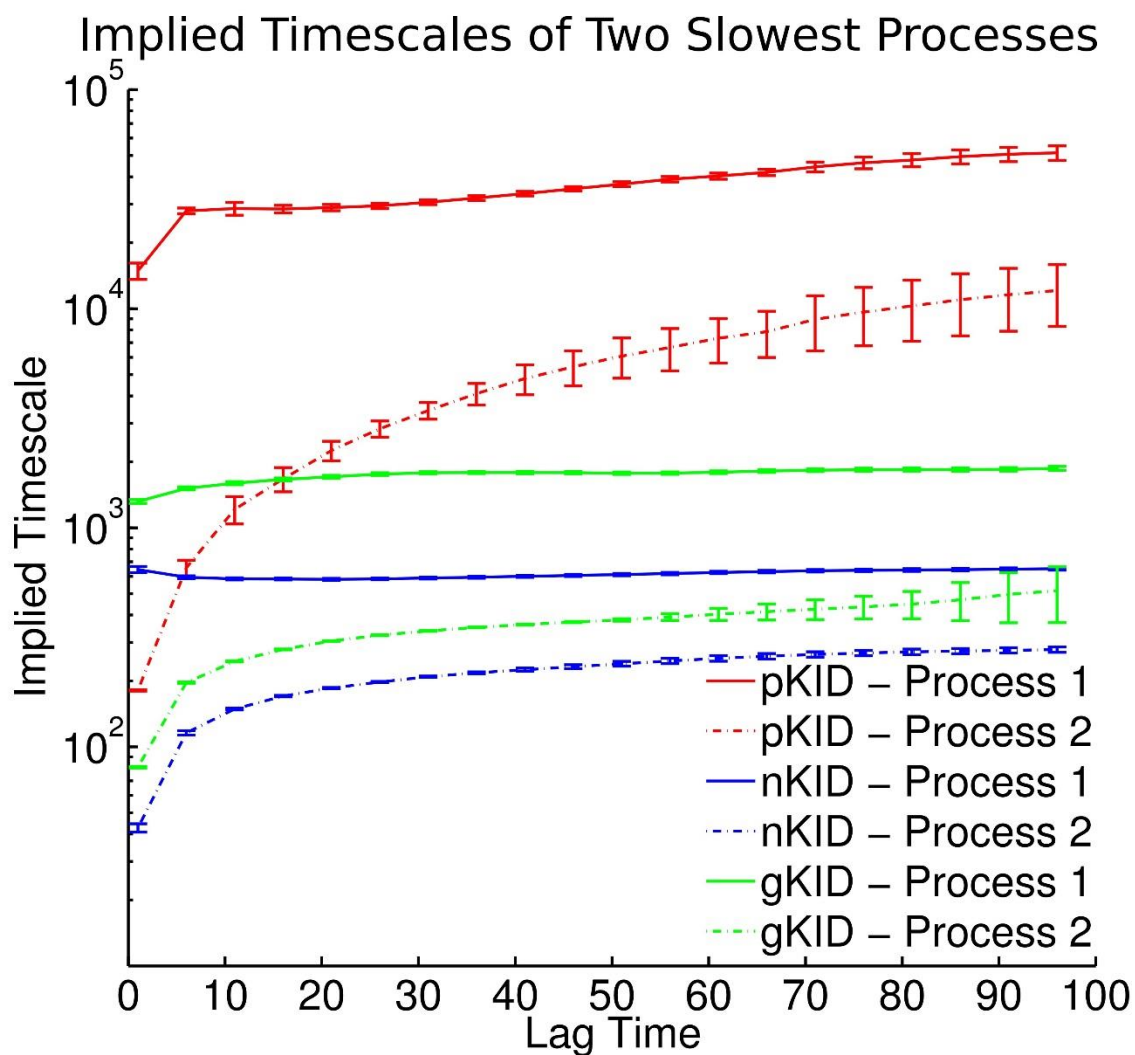
Supplementary Figure 1 | Implied timescales. Convergence of timescales for the two slowest processes for pKID (red), KID (blue), and gKID (green) as determined from the Markov State Model (see Methods). Error bars represent the s.e.m and were determined by ten rounds of bootstrapping in which 20% of the data was randomly removed. Times are in nanoseconds.

Helicity by Residue
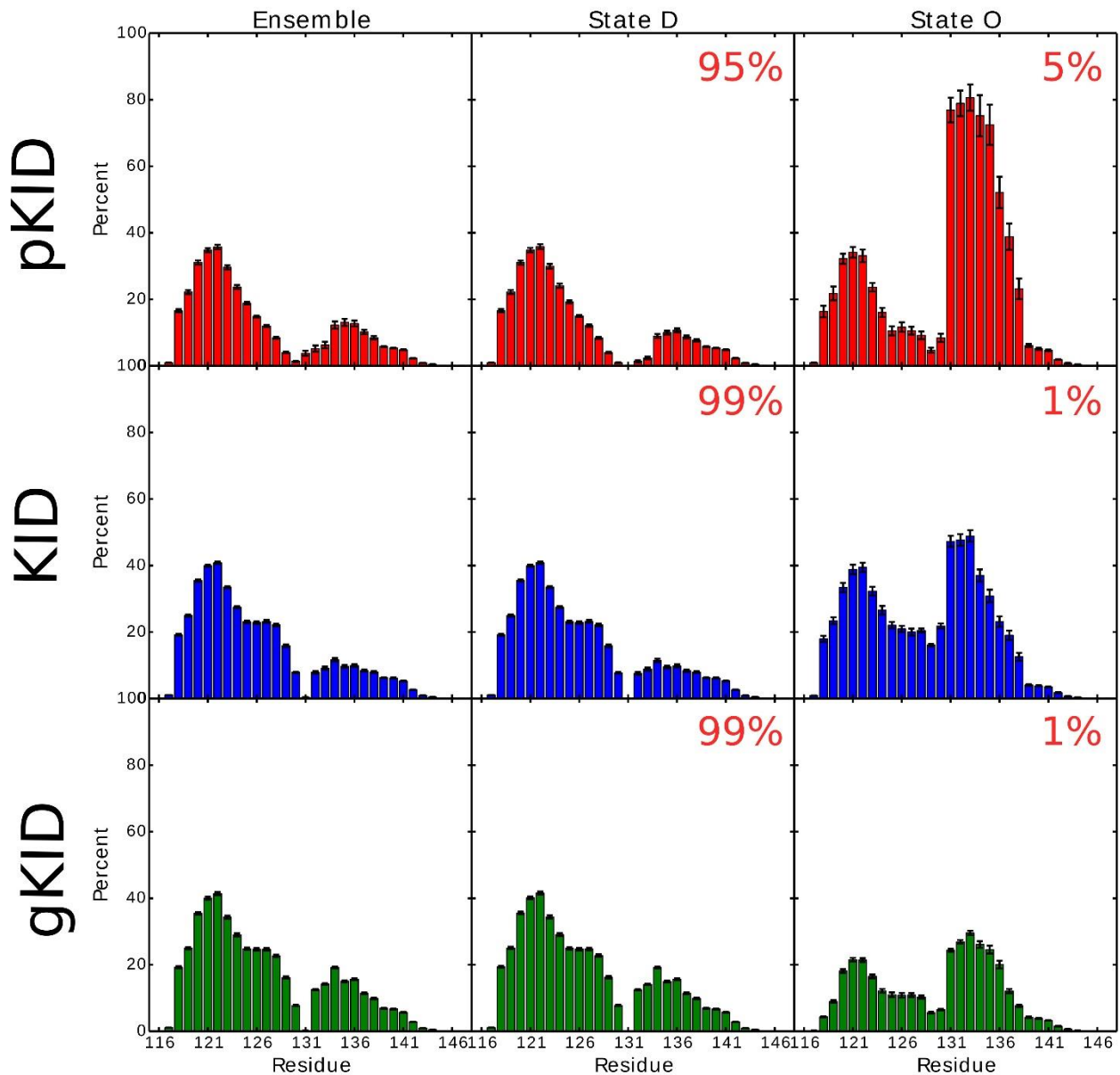
Supplementary Figure 2 | Percent helicity by residue. Plots in the left column show the percent helicity of each residue for the entire ensemble of pKID, KID and gKID systems. The center and right columns show helicity for state D (bulk or disordered state) and state O (ordered state), respectively. The population of state D and state O are shown in red in the upper right corner. Error bars represent the s.e.m and were determined by ten rounds of bootstrapping in which 20% of the data was randomly removed.
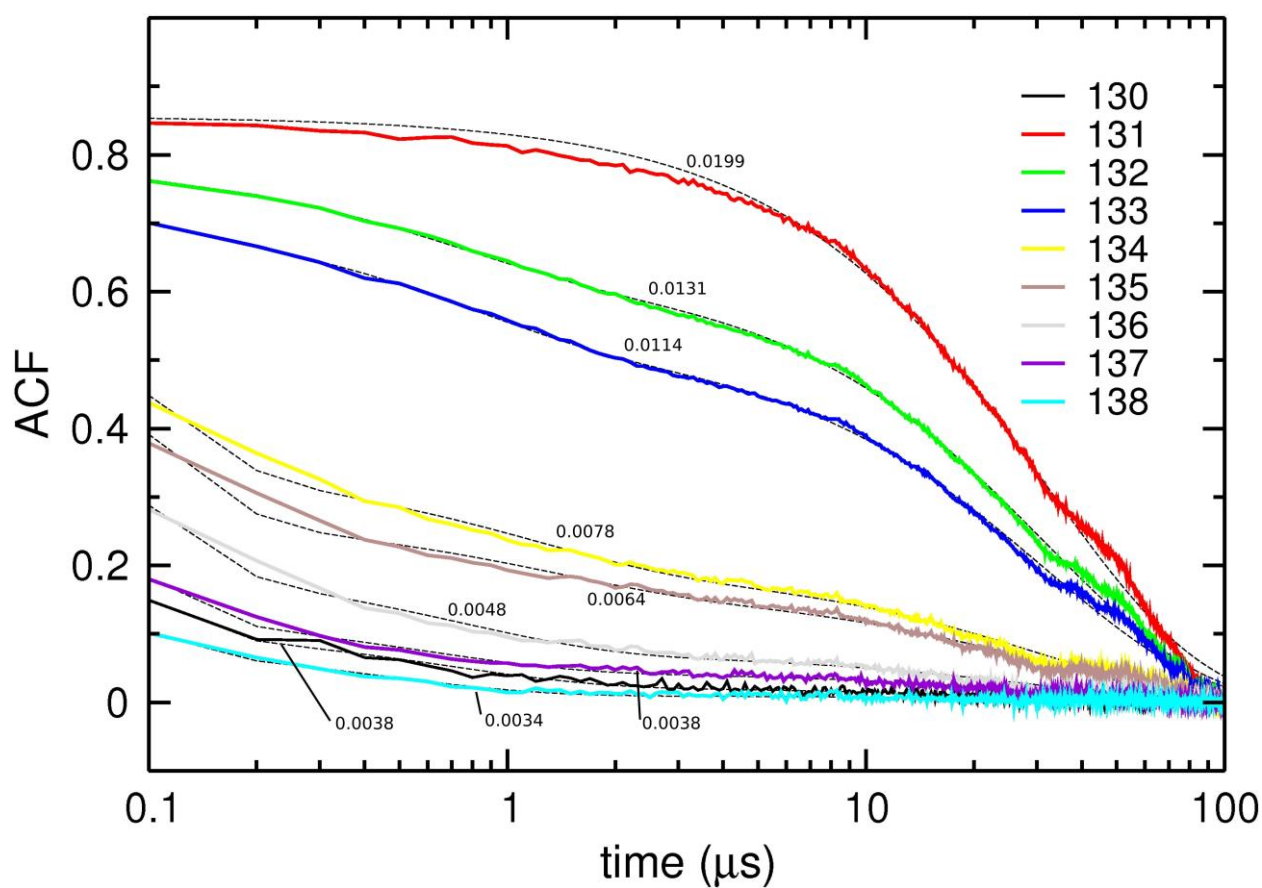
Supplementary Figure 3 | Autocorrelation functions and fits for residues in helix αB in pKID. Fitting of ACFs for residues 130 to 138 is shown as dashed lines and their corresponding RMSD are also given. See also Figure 3a in main text and Supplementary Methods.

Supplementary Methods

Construction and assessment of a kinetic model. The kinetic model we use consists of five coupled, reversible reactions. The following equations describe velocities through each of the five reactions:

$$v_1 = k_1[D_\text{p}] - k_{-1}[D_\text{np}] \qquad (1)$$

$$v_2 = k_2[XD_\text{p}] - k_{-2}[XD_\text{np}] \qquad (2)$$

$$v_3 = k_3[X][D_\text{p}] - k_{-3}[XD_\text{p}] \qquad (3)$$

$$v_4 = k_4[X][D_\text{np}] - k_{-4}[XD_\text{np}] \qquad (4)$$

$$v_5 = k_5[XD_\text{p}] - k_{-5}[XD] \qquad (5)$$

Differential equations for the concentrations of each of the species in the system can be written in terms of these velocities:

$$\frac{d[X]}{dt} = -v_3 - v_4 \qquad (6)$$

$$\frac{d[D_\text{p}]}{dt} = -v_3 - v_1 \qquad (7)$$

$$\frac{d[XD_\text{p}]}{dt} = v_3 - v_2 - v_5 \qquad (8)$$

$$\frac{d[D_\text{np}]}{dt} = v_1 - v_4 \qquad (9)$$

$$\frac{d[XD_\text{np}]}{dt} = v_4 + v_2 \qquad (10)$$

$$\frac{d[XD]}{dt} = v_5 \qquad (11)$$

We used MATLAB and Kinpy (https://code.google.com/p/kinpy/) to solve the set of coupled equations numerically from arbitrary starting concentrations until a steady state was reached.

The dissociation constant, K_d, can then be calculated as:

$$K_\text{d} = \frac{[D_\text{free}][X_\text{free}]}{[XD]} \qquad (12)$$

where $[D_\text{free}] = [D_\text{p}] + [D_\text{np}]$ and $[X_\text{free}] = [X]$.

We investigated the impact a slowdown in conformational exchange between productive and non-productive conformers would have on affinity. A 10-, 100- and 1000-fold slowdown in both the forward and backward rates for $k_{\pm1}$ and $k_{\pm2}$ resulted in 10, 40 and 80 -fold change in binding affinity (K_d), respectively. We used a 100-fold slowdown thereafter since it was most similar in timescale to the effect seen in this work.

We checked the sensitivity of the overall model by modifying individual rates for each of the reactions by a factor of 2 while keeping all other rates the same. The 40-fold change in affinity remained unaffected when these individual rates were altered. The affinities for activated and non-activated systems results 14 ± 7 μM and 568 ± 268 μM.

Autocorrelation analysis. We generated ten MSM trajectories of 10 ms each in order to assess the robustness of the autocorrelation analysis. ACFs were computed for each of these trajectories. Fitting was performed using 3 exponential functions. The first exponential was bounded to 0.1 microseconds as upper limit to account for the initial fast decay, which is determined by the resolution of the MSM (100 ns lag time). The remaining timescales and amplitudes were left unconstrained.

Residues 116 to 130 and 137 to 147 do not have relaxation processes at the tens of microseconds with amplitudes greater than 0.05. Only residues 131 to 136 present relaxation processes around 40 microseconds with amplitudes 0.70, 0.53, 0.45, 0.17, 0.14, 0.05, respectively. A second relaxation process occurs at around 0.35 microseconds with small amplitudes for all residues. The initial decay corresponds to relaxation times around 2 ns with amplitudes 0.29, 0.47, 0.54, 0.82, 0.85, 0.83.

Estimated errors for residues 131 to 136 for the slowest time scale were around 15.5 microseconds. The standard deviation observed comes from variability in the ACFs derived for each of the 10 MSM trajectories and not from difficulties in fitting. Using 3 exponentials sufficed to fit the ACFs computed for each residue in each of the 10 MSM trajectories generated i.e. using 2 or 4 exponential functions left the time scale of the slowest process barely affected. The RMSD average over the 10 trajectories in the fitting for residues 131 to 136 were, respectively, 0.0405, 0.0341, 0.0302, 0.0148, 0.0128, 0.0057.