Structural−Energetic Basis for Coupling between Equilibrium Fluctuations and Phosphorylation in a Protein Native Ensemble

Hemashree Golla, Adithi Kannan, Soundhararajan Gopi, Sowmiya Murugan, Lakshmi R Perumalsamy, and Athi N. Naganathan*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The functioning of proteins is intimately tied to their fluctuations in the native ensemble. The structural−energetic features that determine fluctuation amplitudes and hence the shape of the underlying landscape, which in turn determine the magnitude of the functional output, are often confounded by multiple variables. Here, we employ the FF1 domain from human p190A RhoGAP protein as a model system to uncover the molecular basis for phosphorylation of a buried tyrosine, which is crucial to the transcriptional activity associated with transcription factor TFII-I. Combining spectroscopy, calorimetry, statistical− mechanical modeling, molecular simulations, and *in vitro* phosphorylation assays, we show that the FF1 domain samples a diverse array of conformations in its native ensemble, some of which are phosphorylation-competent. Upon eliminating unfavorable charge− charge interactions through a single charge-reversal (K53E) or charge-neutralizing (K53Q) mutation, we observe proportionately lower phosphorylation extents due to the altered structural coupling, damped equilibrium fluctuations, and a more compact native ensemble. We thus establish a conformational selection mechanism for phosphorylation in the FF1 domain with K53 acting as a "gatekeeper", modulating the solvent exposure of the buried tyrosine. Our work demonstrates the role of unfavorable charge−charge interactions in governing functional events through the modulation of native ensemble characteristics, a feature that could be prevalent in ordered protein domains.
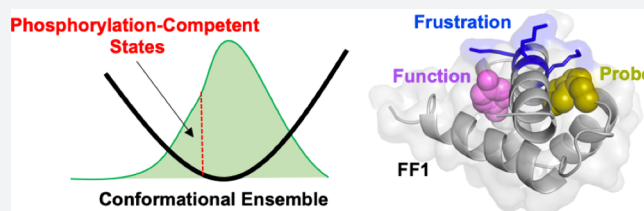
INTRODUCTION

Native dynamics of proteins encompass an array of motions, ranging from side-chain reorientations to the partial melting of helices and loops, and large-scale structural rearrangements involving domain movements. Such motions or fluctuations are a consequence of degeneracy in the number and nature of interactions between the protein and the solvent, the varied intramolecular interactions (within the protein chain), and the entropic stabilization afforded by disorder to different extents. Transitions between iso-energetic (free) states in the native ensemble are usually functionally driven, enabling ligand selectivity and binding, optimal enzymatic activity, post-translational modifications, and even regulation by exposing protease-accessible sites. Ever since the classic Monod− Wyman−Changeux (MWC) model of allostery was proposed,[1,2] numerous examples of proteins undergoing "conformational selection"[1] or "induced fit"[3] have been identified.[4−6] In the former, there exists a pre-equilibrium between the functional and nonfunctional states, while in the latter the binding of the ligand induces a conformational change to drive functional outcomes. A combination of the two has also been invoked to explain experimental data and all-atom simulations, particularly for disordered proteins.[7−12]

Quantifying the extent of structural fluctuations in the native ensemble can reveal insights into functional behaviors,[13−16]

while controlling them through targeted mutations help to decipher the underlying structural−energetic basis.[17−21] In this regard, probing for the conformational selection (CS) mechanism necessitates prior mapping of the conformational landscape that is, in turn, intricately tied to the folding mechanism. It is implicitly understood that proteins undergo enhanced dynamics in this landscape to populate a subset of conformations that are functionally competent.[22−25] Given the large degrees of freedom and varied interactions mediated by protein chains, large fluctuations can aid in sampling either a continuum of states (Figure 1A) or discrete states (Figure 1B). The molecular factors that govern these motions and hence the shape of the underlying free-energy profile can also be diverse and must be studied on a case-by-case basis. However, one common theme could be the role of surface charges given their ability to mediate long-distance interactions and their critical role in protein−ligand binding. Charged residues have been shown to play a dominant role in determining protein
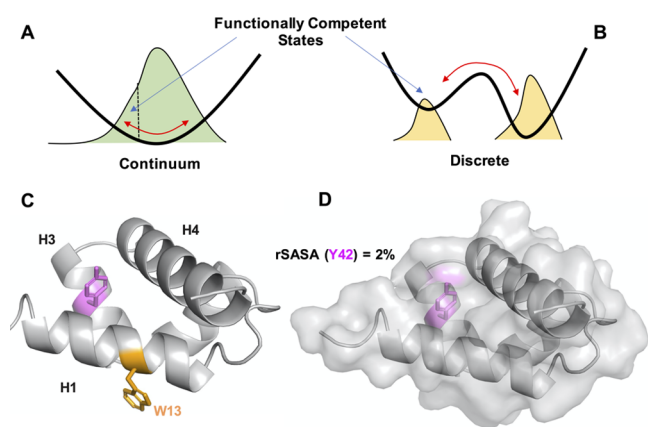
Figure 1. Functionally competent states can populate over (A) a single broad well (continuum) or (B) as a set of discrete conformations when visualized in a one-dimensional projection. The black curves are the free-energy profiles, while the filled areas represent the corresponding probabilities. (C) Cartoon and (D) surface representations of the p190A RhoGAP FF1 domain, respectively. The residues Y42 and W13 are shown in magenta and orange, respectively, in panel C.

stabilities,[26,27] folding mechanisms,[28−30] differences in conformational behaviors across paralogs and orthologs,[31] and even the magnitudes of protein diffusion coefficients within cells.[32]

We explore these issues in the current work by employing the first FF domain (FF1) from human p190A RhoGAP protein as a model system. The cytoplasmic p190A RhoGAP protein is comprised of a N-terminal GTPase domain, four FF domains in the middle, and a C-terminal RhoGAP domain (Supporting Figure S1) and participates in various cellular processes such as migration, invasion, and morphogenesis.[33] The structure of the FF1 domain was solved by multidimensional NMR spectroscopy by Macias and co-workers at 285 K (Figure 1C).[34] It is a 65 residue all-α protein characterized by an $\alpha1-\alpha2-\alpha3-\alpha4$ arrangement instead of the canonical $\alpha1-\alpha2-3_{10}-\alpha3$ architecture seen in other FF domains. Functionally, it binds to the transcription factor TFII-I in the cytoplasm. This interaction is eliminated upon the phosphorylation of tyrosine 308 (Y308; Y42, when the domain numbering starts from 1) in the third helix of the FF1 domain, resulting in the translocation of TFII-I to the nucleus.[35] The translocated TFII-I regulates the transcription of several genes involved in tumor suppression. The RhoGAP FF1 also controls translation by interacting with the eukaryotic initiation factor 3A (eIF3A), which solely depends on the phosphorylation of S296 (S30) of FF1.[36] Additionally, the tandem FF repeats are speculated to mediate or strengthen the downstream functions by scaffolding, while the control of interactions with the two proteins lies in FF1. Earlier studies on FF1 domain unfolding monitored by NMR spectroscopy reveal dramatic structural changes, with the peak dispersion in HSQC experiments decreasing with increasing temperature from 280 to 310 K and the peaks corresponding to the phosphorylation motif D41, Y42, and V43 entirely disappearing at 310 K. Comparing the spectral properties with those of the FF2 domain from another protein CA150, the authors concluded that the observed structural changes are unique to the FF1 domain from p190A RhoGAP.[34]

However, this work raises additional questions. First, the Y42 side-chain that is sandwiched between helix 1 (H1) and helix 4 (H4) is completely buried within the protein core, with a solvent accessible surface area (SASA) of just 2% (i.e., 98% buried; Figure 1D). How does Y42 get exposed to the solvent and then phosphorylated? Second, does the side-chain of Y42 exist in a pre-equilibrium between phosphorylation-competent and phosphorylation-incompetent conformations? Third, if this is the case, what structural−energetic features in the protein determine this pre-equilibrium and the degree of fluctuations? Fourth, can we engineer this pre-equilibrium through targeted mutations to control fluctuations and hence the extent of phosphorylation? Finally, can all these be probed and studied via ensemble thermodynamic measurements and simulations to provide a self-consistent picture? In this work, we employ an array of spectroscopic experiments, detailed statistical mechanical modeling, molecular simulations, and phosphorylation assays to show that the FF1 domain indeed samples a plethora of conformations in the native ensemble, some of which are phosphorylation-competent, while demonstrating the existence of a conformational selection mechanism using rationally designed mutants.

■ METHODS

Purification of RhoGAP FF1 and Mutants. The plasmid pTXB1 (IMPACT, New England Biolabs, UK) harboring the codon-optimized gene for the RhoGAP FF1 domain protein sequence (SQQIATAKDKYEWLVSRIVKNHNENWLS-VSRKMQASPEYQDYVYLEGTQKAKKLFLQHIHRLKH-EHIER) was purchased from GenScript Inc. (New Jersey, USA). A single transformed colony was inoculated in 2 L of the Luria−Bertani (LB) medium with ampicillin (0.05%). Cells grown at 37 °C and 180 rpm to an optical density (600 nm) of ∼0.9 were induced with 0.5 mM IPTG, harvested after 16 h of growth at 16 °C, and lysed using a sonicator (Q500 Qsonica) in the sonication/affinity column buffer (20 mM Tris, 500 mM NaCl, 1 mM EDTA, and 1 mM PMSF, pH 8.5). The clarified lysate was loaded onto a chitin resin column, and the tagged intein was cleaved using 50 mM β-mercaptoethanol in an elution buffer (20 mM Tris and 50 mM NaCl, pH 8.5) after a 16 h incubation at room temperature. The eluted fractions containing the cleaved RhoGAP FF1 domain were loaded onto a cation column (BioRad; MiniPrep HighS, 5 mL cartridges). The column and elution buffers were 20 mM Tris, pH 7.5, combined with 50 mM NaCl and 1 M NaCl, respectively. The protein was gradient eluted at a salt concentration of 200−300 mM at a flow rate of 0.5 mL/min. The fractions containing the FF1 domain were pooled, frozen, and lyophilized. The lyophilized protein was dissolved in Milli-Q water, injected into a HiLoad 26/600 Superdex 75pg column, and eluted in 150 mM ammonium acetate, pH 8. The purity of the eluted fractions was assessed using SDS-PAGE, and the fractions were lyophilized for further use. The mutants (W13F, W13F/K53E, and W13F/K53Q, termed WT, K53E, and K53Q, respectively, for simplicity, and W13E) of the RhoGAP FF1 domain were generated by site-directed mutagenesis (see the Supporting Information) using the Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs). The mutants were transformed and purified using a protocol similar to that for the wild-type purification. All experiments were recorded in 20 mM sodium phosphate buffer, pH 7.0 (43 mM ionic strength). Buffers were freshly prepared with Milli-Q water, filtered, and degassed before

every experiment. Protein samples were filtered using a 0.22 μm syringe filter (Millipore), and absorbance was measured using a UV−visible spectrophotometer (Jasco Inc.). Concentrations were estimated with an extinction coefficient of 16960 M^{-1} cm^{-1} for the original construct (FF1), while that of 11460 M^{-1} cm^{-1} was used for the WT and K53E/K53Q/W13E variants.

Far- and Near-UV Circular Dichroism. Protein solutions at concentrations of ∼18 and 90 μM were used for far- and near-UV circular dichroism (CD) experiments, respectively. The experiments were recorded in a Jasco J-815 spectropolarimeter coupled to a Peltier system. For far- and near-UV CD, cuvettes with path lengths of 1 and 10 mm were used, respectively. Thermal unfolding was monitored by recording spectra at 5 K temperature intervals from 278 to 368 K. The protein sample was equilibrated for 2 min before the spectrum was recorded at each temperature.

Differential Scanning Calorimetry (DSC). The temperature dependence of the partial molar heat capacity was measured at pH 7.0 in 20 mM sodium phosphate buffer at a range of protein concentrations. Calorimetric experiments were performed using the VP-DSC microcalorimeter (Malvern Microcal VP, NL) at a scan rate of 1.5 K/min. The samples were degassed at room temperature prior to calorimetric measurements. Calorimetric cells were maintained at an excess pressure of 60 psi during the scan to prevent boiling at high temperatures. Buffer−buffer baselines were routinely acquired to ensure that the thermal history was maintained.

Fluorescence Spectroscopy. Fluorescence spectra were acquired using the Chirascan-plus qCD instrument (Applied Photophysics) coupled to a Peltier system. The emission spectra (300−550 nm) were obtained by exciting a ∼10 μM protein sample at 295 nm. The equilibrium thermal melt of the protein was recorded from 278 to 368 K at 5 K intervals. Fluorescence lifetime measurements were performed as described before.[37]

Wako−Saitô−Muñoz−Eaton (WSME) Model. A detailed description of the model can be found in earlier works.[38−41] Briefly, the WSME model discretizes the phase space accessible to a protein chain by assuming that every residue can sample either folded-like (binary variable *1*) or unfolded-like (binary variable *0*) conformations. The microstates will therefore be represented by strings of ones and zeros. The statistical weight of every microstate is determined by the stabilization free-energy contributions (van der Waals, electrostatics, and simplified solvation) derived from the native structure with Gō-like energetics (PDB ID 2K85)[34] and the conformational entropic penalty of fixing residues in the folded conformation. We employ a treatment that considers only single stretches or islands of ones (single sequence approximation, SSA), two islands of ones with no interactions between them (double sequence approximation, DSA), and two islands of ones that allow for interactions between them (DSA with loop, DSAw/L).[42] Native interactions were identified with a 5 Å distance cutoff, including nearest neighbors and excluding hydrogens, while charge−charge interactions were considered without assuming any distance cutoff. The final thermodynamic parameters, which were extracted by fitting the DSC curves to the model, are: van der Waals interaction energies per native contact of -44.9 ± 3.81 (WT) and -58.7 ± 1.48 J/mol (K53E); entropic penalties for fixing a residue in the native conformation of -10.25 ± 0.84 and -13.8 ± 0.32 J/mol·K per residue for all residues other

than proline, glycine, and nonhelical residues; and heat capacity changes per native contact of 0 and -0.25 ± 0.04 J/mol·K. All nonhelical and glycine residues were assigned an additional entropic penalty of -6.06 J/mol·K per residue given their larger degrees of freedom,[43] while the entropic cost of fixing proline was set to zero given its limited backbone flexibility. Free energy profiles and surfaces were constructed by accumulating the statistical weights of those microstates that exhibited a certain number of structured residues.

Replica Exchange Monte Carlo (REMC) Simulations. The structure of the FF1 domain from PDB ID 2K85[34] was used as a template to generate initial structures of the WT and K53E variants using PyMOL.[44] Replica exchange Monte Carlo (REMC) simulations were performed using the CAMPARI stand-alone package.[45] The structures were placed at the center of 100 Å spherical shells, and appropriate ions (five and three Cl^- ions for the WT and K53E variants, respectively) were added to neutralize each system. Additionally, 105 pairs of Na^+ and Cl^- ions were added to mimic experimental ionic strength conditions (43 mM). Both the simulations used the ABSINTH implicit solvent model with OPLS charges along with an energy function that accounted for the temperature-dependent dielectric constant and solvation free energies. The REMC simulations were run for 60 million steps over 22 temperature replicas that were equally spaced in the range of 280−450 K, with exchange attempts every 10^4 steps between consecutive temperature bins. The average exchange probability between the temperature bins was estimated to be ∼0.45. The coordinates are collected every 500 steps, and all the analyses were performed on the final 30 million steps.

***In Vitro* Phosphorylation Assays.** For the *in vitro* kinase assays, 3.75 μg of WT/other variants and 7.5 μg of K53E were used with myelin basic protein (MBP) as a positive control and "no substrate" as a negative control. Assays were performed in duplicates using 100 ng of purified PDGF receptor α kinase (Promega, USA) and the respective proteins in 30 μL of a reaction mixture containing the kinase assay buffer at pH 8 (50 mM HEPES, 10 mM $MgCl_2$, 2 mM $MnCl_2$, and 200 μM DTT), 0.25 μL of γ-^{32}P-labeled ATP, and 0.1 μL of nonlabeled ATP (10 mM stock). Samples were incubated at 298 or 310 K for 30 min. Caution! The radioisotope sample represents a health hazard. All studies were conducted in a Radioisotope Laboratory following necessary precautions. The assay samples were loaded on to 16% Tricine SDS-PAGE gels and then transferred to a PVDF membrane. The transfer efficiency was checked using a Ponceau stain followed by exposure to the storage phosphor screen and scanned using an Amersham Typhoon IP phosphorimager (GE Healthcare) after 24 h. The blot was further stained using Coomassie brilliant blue to ensure the equal loading of protein samples (this necessitated using twice the amount of K53E to ensure near-equal band intensities in the blot). All the images were quantified using both GelBandFitter tools[46] and ImageJ.[47] Phosphorylation extents were quantified by taking the ratio of intensities, measured as the area under the curves, between the phosphor- and Coomassie-stained bands to ensure the appropriate normalization. The percent relative phosphorylation was averaged across experiments and analysis tools.

■ **RESULTS AND DISCUSSION**

Electrostatic Frustration Governs the Global Thermodynamic Behavior of the FF1 Domain. The RhoGAP FF1 domain harbors two tryptophan residues, W13 and W26, with
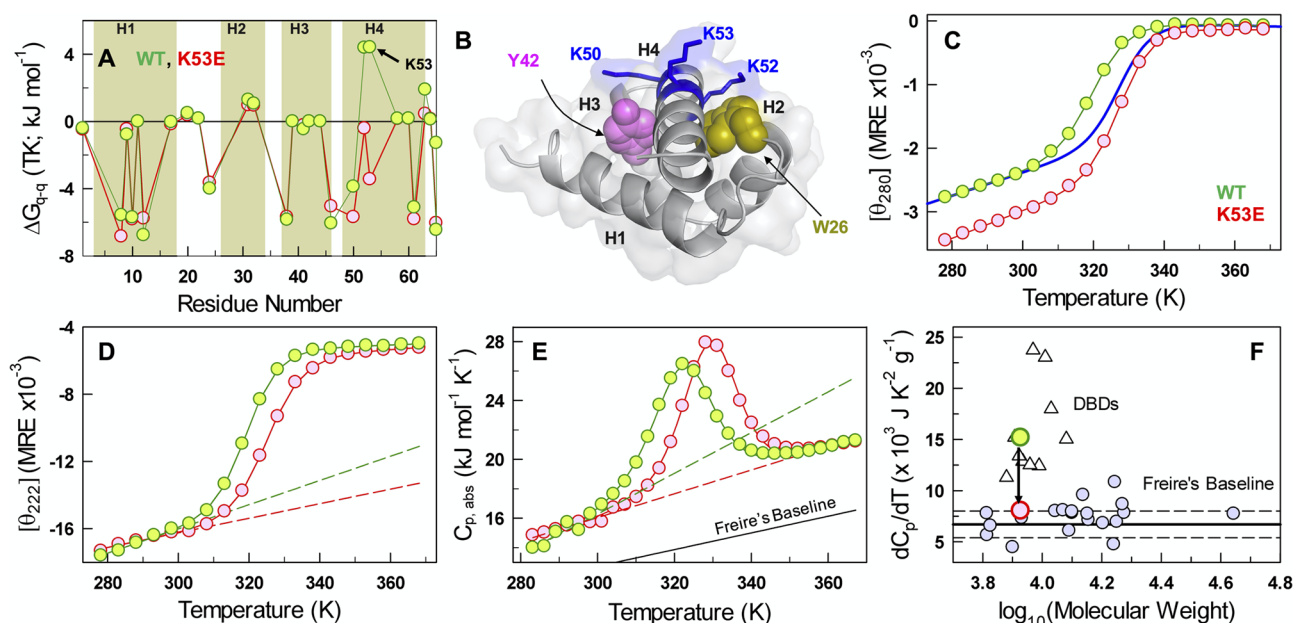
Figure 2. Electrostatic frustration and folding thermodynamics. (A) Tanford−Kirkwood electrostatic interaction free-energies as a function of residue number for the WT (green) and the K53E mutant (red), respectively. The shaded regions represent helical boundaries. (B) Structure of the FF1 domain highlighting the large degree of electrostatic frustration in the fourth helix. Thermal unfolding curves of the WT and the mutant from (C) near-UV CD at 280 nm, (D) far-UV CD at 222 nm, and (E) the heat capacity profiles. The dashed lines represent the corresponding pretransition baselines. The blue curve in panel C indicates the expected unfolding curve if there are no near-UV CD signal changes at the lowest temperature for the K53E mutant. (F) The slope of the pretransition baseline (dashed lines in panel E) compared with those of conformationally heterogeneous DNA binding domains (DBDs; triangles) and well-folded proteins (small filled circles). The continuous and dashed horizontal lines signal the Freire baseline slope and one standard deviation, respectively. Note that the mutation K53 → E53 changes the slope significantly.

the former fully exposed to the solvent in helix 1 (Figure 1C). The tryptophan at position 26 (W26) is highly conserved across the FF domains, not just in p190A RhoGAP tandem repeat FF domains but also across CA150 and FBP11 (Figure S2A). However, W13 is poorly conserved. Hence, we mutated it to phenylalanine (W13F), resulting in a construct with a single tryptophan to probe for structural changes with temperature. The W13F mutation reduces the melting temperature (T_m) by ∼3 K but importantly does not affect the slope of the pretransition baseline (Figure S2B and C). We employ this variant as the pseudo-WT, and we label it the WT for simplicity from here on.

To explore the extent to which charge−charge interactions are distributed on the surface of the WT FF1 domain, we calculated the Tanford-Kirkwood (TK) electrostatic interaction free-energies for every charged residue.[48,49] Figure 2A highlights that the majority of the electrostatic interaction free-energies are negative (favorable) except for two residues, namely K52 and K53 (green in Figure 2A). Mapping these residues onto the protein structure, we find that three residues K50, K52, and K53 are located spatially close to one another (Figure 2B). K50, which occludes Y42 from being fully exposed to the solvent, is in a favorable electrostatic environment (Figure S3). On the other hand, K52 and K53 display significant frustration (large positive interaction free-energies), with K52 being tightly packed against W26 and thus holding together helices 2 and 4 (H2 and H4) (Figure 2A). Given these observations, we hypothesize that the large repulsion between K52 and K53 determines the degree of structure in the fourth helix. In fact, a computational substitution of this lysine by glutamate (K53 → E53) fully eliminates this unfavorable interaction (red in Figure 2A) and stabilizes the mutant, with effective charge−charge interaction

energies of −38.3 and −52.3 kJ/mol for the WT and the K53E mutant, respectively. It is therefore possible that K53 acts as a "gatekeeper" not in the conventional sense of occluding access to the active site but instead by controlling the degree of structure in the fourth helix and hence the extent to which Y42 is exposed to the solvent through indirect effects on the native ensemble heterogeneity.

We quantify the degree of tertiary packing in the native ensemble of the WT and the K53E mutant (W13F/K53E is referred to as K53E) by monitoring the temperature dependence of the near-UV CD signal at 280 nm (Figure S4). The K53E mutant is more stable by 6 K compared to the WT (ΔT_m = 6 K, where T_m is the melting temperature from two-state fits) in accordance with the expectation from the Tanford−Kirkwood (TK) electrostatic interaction energy calculations. Interestingly, the WT exhibits a weaker near-UV CD signal in contrast to that of K53E, which is indicative of a less-packed hydrophobic core in the former. Note that the signal difference between the two variants is unexpected even at the lowest temperature (278 K), as any stabilization is expected to shift the curve only to the right (blue in Figure 2C). Monitoring the changes in the secondary structure by far-UV CD at 222 nm, we observe a trend where the native baseline is steeper for the WT than for the mutant (dashed lines in Figures 2D and S4). Studies on helical proteins have highlighted that such differences in the pretransition baselines are suggestive of differences in fluctuations in the native ensemble,[50] with the steeper baseline corresponding to a broader native ensemble.

Scanning calorimetry experiments provide an alternate avenue to quantify the extent of the native ensemble heterogeneity given the intimate connection between heat capacity and equilibrium enthalpic fluctuations.[51−53] In this
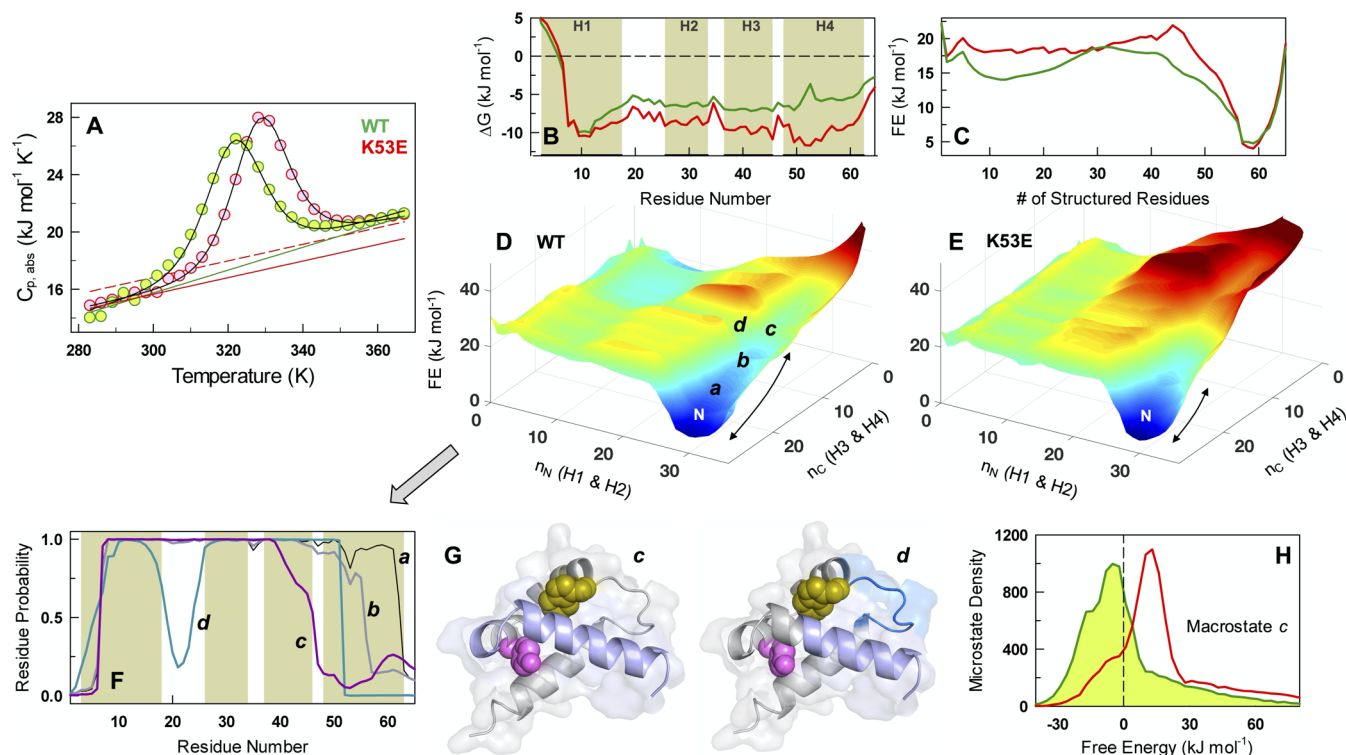
Figure 3. The WSME model predicts a broader native ensemble for the WT. (A) WSME model fits (curves) to the DSC data (circles) with the folded baselines as continuous lines and the unfolded baseline as a dashed line, respectively. (B) Residue stability profiles for WT (green) and K53E mutant (red) at 310 K. (C) One-dimensional free energy profiles at 310 K as a function of the number (#) of structured residues as the reaction coordinate. (D and E) Two-dimensional free-energy landscapes as a function of the number of structured residues in the N- and C-termini (n_H and n_C, respectively), highlighting the differences in the conformational landscapes of the WT and the K53E mutant. Note that the WT samples a broader native ensemble (double arrow) compared to the K53E mutant. "N" represents the fully folded native state. (F) Mean residue folding probability for the indicated macrostates in the conformational landscape (panel D). Low values of probability (<0.5) indicate unfolded or partially structured residues. (G) Mean structural features of the macrostates c and d shown in panels D and F. Macrostate c is characterized by a partially structured H4 (light blue) that can be sensed by both Y42 (magenta) and W26 (olive). Macrostate d exhibits a partially structured C-terminal in H4 (light blue), including the loop connecting H1 with H2 (navy blue). (H) The free-energy of all microstates encompassed within macrostate c for the WT (green) and the K53E mutant (red) at 310 K. Negative free-energies here represent disordered states.

regard, it is informative to compare the slope and the intercept of the pretransition regions from heat capacity profiles to the Freire baseline (FB). The FB accounts for protein size-effects and is derived from the native baselines of well-folded systems; hence, any deviation from the FB is expected to arise from enhanced conformational fluctuations.[54] Both the WT and the mutant display higher heat capacity values, which are distinct from the Freire baseline even at the lowest temperatures (Figure 2E), but do they exhibit different degrees of conformational heterogeneity? The pretransition region is steeper for the WT in accordance with far-UV CD measurements (dashed lines in Figure 2E). The connection to conformational fluctuations can also be explicitly made from the slope of the pretransition baselines. Specifically, the DNA binding domains (DBDs) that display greater structural polymorphism have been shown to exhibit higher pretransition slopes compared to those of well-folded proteins[55] (Figure 2F). We find that the pretransition heat capacity slope of the WT is similar to those of the DBDs, but the mutant K53E data point to a substantially smaller slope and hence fall within the expectation for reasonably well-folded proteins. A fit assuming a two-state model results in unphysical crossing baselines for the WT, while it explains the K53E thermogram with near-parallel baselines (Figure S5). Taken together, it is clear that the WT is characterized by weak tertiary packing in the

hydrophobic core and exhibits a broader native ensemble with enhanced conformational fluctuations. Thus, the K53E mutation plays a dual role by not only increasing the stability but also modulating the characteristics of the system in the native ensemble.

Differential Native Ensemble Heterogeneity from Statistical Mechanical Modeling. In this section, we employ the WSME model to quantify the differences in the conformational behavior between the WT and the mutant. The version of the WSME model employed here accounts for microstates with single and two stretches of folded residues while also allowing for interactions across the structured islands (see Methods). We thus assume a large ensemble of 1, 369, 499 microstates where the energetics of every microstate are determined by the native structure (Gō-like model), including van der Waals interactions, electrostatics, and simplified solvation, apart from sequence- and structure-dependent conformational entropy. The model captures the overall features of the heat capacity profiles very well (Figure 3A) but still misses out in accounting for the steep pretransition slope of the WT (note the lowest temperature points in green in Figure 3A). Thus, the model predictions discussed below represent only a lower estimate of the conformational heterogeneity in the WT.
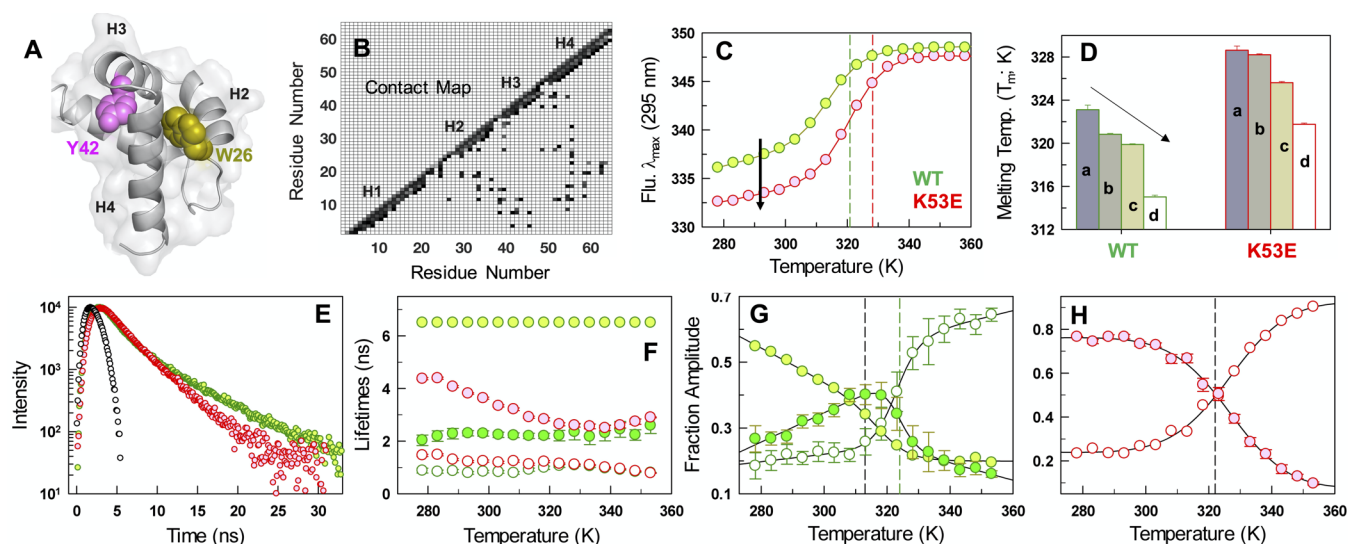
Figure 4. Probing structural changes in helix 4. (A) Structure and (B) contact map of the FF1 domain showing the crucial interactions mediated by H4, whose changes can be monitored by fluorescence parameters of W26. (C) Changes in the H4–H2 interface as monitored by the fluorescence emission maximum of W26 (λ_{max}; excitation at 295 nm) as a function of temperature. The vertical dashed lines signal the melting temperature from near-UV CD at 280 nm. The black arrow denotes the lower emission maximum for the K53E mutant. (D) Melting temperatures derived from two-state fits to the unfolding curves from (a) DSC, (b) near-UV CD at 280 nm, (c) far-UV CD at 222 nm, and (d) fluorescence emission maximum changes. Note that the differences going from global (DSC) to local probes (λ_{max}) are maintained between the WT and the mutant. (E) Fluorescence intensity transients at 298 K, with the IRF shown in black. (F) Fluorescence lifetimes of the WT (green) and the K53E mutant (red) as a function of temperature. Corresponding amplitudes of the lifetimes for (G) the WT and (H) the K53E mutant. The vertical dashed lines in panel G signal the apparent inflection point from the amplitudes of the longest (black dashed) and shortest lifetime components (green dashed), respectively.

It is instructive to discuss the effect of the K53E mutation on the overall residue-level stability profile of the FF1 domain. In the WT, helix 1 appears to be the most stable secondary structure (green and more negative in Figure 3B), and helix 4 appears to be the least stable (less negative in Figure 3B). The mutation K53E flips this pattern by enhancing the stability of helix 4, as expected from the TK calculations. Note that any mutation that uniformly influences all residues would enhance the stability without changing the relative ordering of the stability pattern. The one-dimensional free-energy profiles as a function of the number of structured residues (the reaction coordinate) are markedly different (Figure 3C). The native ensemble is predicted to be broader for the WT and have a shallow slope, while the K53E mutant displays a steeper slope and hence a narrow native well.

The differences are amplified when the large ensemble is projected onto two coordinates (Figure 3D and E), namely the number of structured residues in the N- (includes residues from H1 and H2; n_N) and C-termini halves of the structure (residues from H3 and H4; n_C). A larger sea of blue and hence lower free energies are visible for large portions of the landscape in the direction of n_C and closer to the native state (N) for the WT. This is indicative of enhanced conformational heterogeneity, which is in agreement with the higher heat capacity slopes of the WT (region around "N" marked in Figure 3D). In contrast, the landscape features of the K53E mutant are suggestive of a more compact native ensemble (smaller area of blue around "N" in Figure 3E), which is in excellent agreement with different experimental probes (*vide supra*). It is possible to identify the structured protein regions in each of the macrostates *a*, *b*, *c*, and *d* marked in Figure 3D by accumulating the statistical weights of microstates that satisfy the specific criterion of a fixed number of residues structured in the N- and C-terminal halves. The C-terminal H4

progressively unfolds going from *a* to *b* to *c*, with *d* being a macrostate characterized by the partial unfolding of H4 that in turn promotes additional unfolding in the loop that connects H1 and H2 because of their spatial proximity (Figure 3G). We would like to highlight that the macrostates identified above are (by definition) large ensembles, as exemplified in Figure 3H for the macrostate *c*. The distribution of microstate free-energies indicates that many microstates exhibit partial unfolding of H4 (negative free-energy values in Figure 3H). The situation is reversed for the mutant K53E, which is predominantly folded with a relatively fewer number of microstates exhibiting an unfolded-like status in H4. We thus propose that macrostate *c* is the "phosphorylation-competent state", where the partial unfolding of H4 contributes to the transient exposure of Y42 to the solvent.

A similar weak packing and partial structure in H4 was also reported for the FF domain from HYPA/FBP11.[56−58] However, this appears as an on-pathway intermediate (*i.e.*, separated from the native state via a barrier), while here we show that the conformations with disorder in the fourth helix of the RhoGAP FF1 domain populate among the continuum of states in the native ensemble. To explore this discrepancy, we employ the WSME model and predict the HYPA/FBP11 FF domain free-energy profile. A clear intermediate with a partial structure in the fourth helix was observed in the resulting one- and two-dimensional projections (Figure S6), which agrees with previous works while also attesting to the model's predictive capabilities. It is therefore evident that disorder or a partial structure in the fourth helix of FF domains can arise via different mechanisms whose molecular origins need to be explored individually, while this feature is implicitly accounted for by the WSME model.

Local Experimental Probes and Helix 4 Stability. Probing for partial unfolding in helix 4 (H4) is challenging in
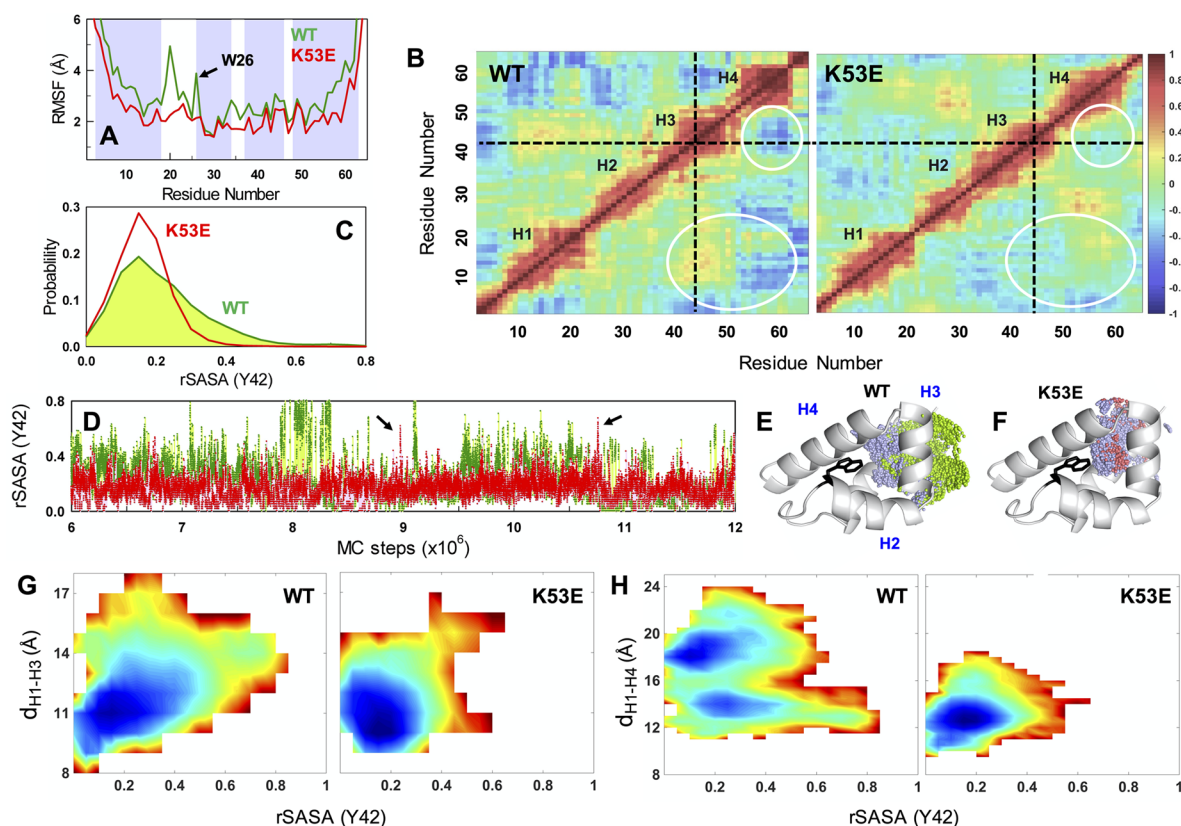
287

Figure 5. Simulations support experimental observations of conformational heterogeneity in the WT. (A) Root-mean-square fluctuations (RMSF) as a function of the residue number. Shaded regions represent the secondary structure elements. Note that the loop connecting H1 and H2 exhibits a larger RMSF. The mean RMSFs of W26 are 2.1 and 3.9 Å in the K53E mutant and the WT, respectively. (B) Ensemble-derived structural correlations point to dramatic differences in the coupling patterns (white circles) between the WT and the K53E mutant. (C) The WT displays a broader distribution of the relative solvent accessible surface area (rSASA) of Y42 compared to that of the mutant. (D) rSASA of Y42 as a function of MC steps. Note that specific snapshots in the K53E mutant display higher rSASA (arrows). (E and F) Positions of the Y42 hydroxyl groups mapped onto the structure. The WT exhibits a larger proportion of exposed hydroxyl groups (green in panel E) compared to the mutant (red in panel F). Light blue circles signal buried hydroxyl groups. W26 is shown in black to highlight that it is sensitive to the observed structural changes. (G and H) Distribution of distances between H1−H3 and H1−H4 as a function of the tyrosine 42 solvent exposure. The density plots are colored in the spectral scale, ranging from blue (high probability) to red (low probability).

the RhoGAP FF1 domain, as NMR spectroscopy (a method suitable for identifying partially structured states) points to a dynamic conformational exchange between multiple conformations and resonances that disappear with temperature modulations.[34] To overcome this experimental barrier, we exploit the signal properties of tryptophan W26, which is packed at the interface of the helices H4 and H2 (Figure 4A and B). Since tryptophan residues are extremely sensitive to their local environment, it is possible to explicitly observe structural changes involving this local region by measuring not just the fluorescence intensities but also the fluorescence emission maximum wavelength (λ_{max}), as well as the number and magnitude of fluorescence lifetimes and their amplitudes.

The λ_{max} values, upon excitation of W26 at 295 nm (probing only for the tryptophan properties), were found to be 336 and 333 nm (± 0.5 nm) for the WT and the K53E mutant, respectively. The higher λ_{max} for the WT, indicative of more tryptophan exposure to the solvent due to electrostatic frustration, is consistent with the weaker near-UV CD signal at 280 nm (Figure 2C). Unlike near-UV CD data, emission maxima are independent of the protein concentration and provide solid evidence that the mutation K53E reorganizes the ensemble or the relative side-chain position of W26 through the modulation of charge−charge interactions. Remarkably,

the melting temperature (T_m) extracted from the temperature dependence of λ_{max} was found to be 7 K lower for both the proteins compared to those extracted from near-UV CD or DSC experiments (vertical dashed lines in Figure 4C indicate the T_m from near-UV CD; Figure S4). A distinct trend is also visible in the magnitude of melting temperatures when going from global to local probes for both the proteins, *i.e.*, T_m follows the order DSC > near-UV CD > far-UV CD > fluorescence (295 nm), with a maximum difference of 7 K between DSC and fluorescence (Figure 4D). This observation is unexpected and suggests that both the WT and the mutant exhibit weak thermodynamic coupling between the different structural elements. However, differences in melting temperatures alone do not prove if the native ensemble of the WT is more heterogeneous than that of the K53E mutant.

To explore this question, we measured the fluorescence lifetimes of W26, as lifetimes are exquisitely sensitive to the ensemble properties.[37,59] In fact, the WT and the K53E mutant display distinct transients even at 298 K (Figure 4E). The WT FF1 domain unusually exhibits three fluorescence lifetimes, while the K53E mutant exhibits two lifetimes, which are conventionally observed for tryptophan residues (Figures 4F and S7). The temperature dependence of the corresponding amplitudes is equally complex for the WT. The amplitude of

the longer lifetime (~6.7 ns) depends steeply on temperature, with an apparent midpoint of 314 K (filled light green in Figure 4G), while the amplitude of the shortest lifetime (~1 ns) shows an inflection point at 324 K (open circles in Figure 4G). The intermediate lifetime component's (~2−3 ns) amplitude increases, peaks at 319 K, and decreases at higher temperatures (bright green in Figure 4G). This is suggestive of an intermediate-like state whose population generally shows a parabolic profile upon destabilizing perturbations. The K53E mutant, on the other hand, displays a behavior expected for the transition between two conformational substates, with a T_m of 322 K (Figure 4H). The intricate temperature dependence of the amplitudes firmly establishes that the WT samples additional states in the native ensemble that are either less populated in the mutant K53E or invisible from the viewpoint of the probe employed.

Transient Exposure of the Buried Tyrosine 42 and Structural Polymorphism from Simulations. We provide further evidence for the conformational heterogeneity of the WT by performing replica exchange Monte Carlo (REMC) simulations using the ABSINTH implicit solvent force-field (see Methods). The WT displays an enhanced conformational flexibility compared to the mutant (Figure 5A), with the residues in the region 18−26 displaying large fluctuations. Interestingly, this corresponds to the loop that connects H1 to H2 and interacts with the C-terminal half of H4, thus resembling the partially unstructured macrostate *d* from the WSME model simulations (Figure 3F and G). The mutation K53E modulates the overall thermodynamic coupling between residues with the native ensemble, which can be observed from the changes in ensemble-averaged inter-residue cross-correlation coefficients (Figure 5B). The circled regions in Figure 5B indicate that the cross-correlations between H1 and H4 are reversed, while those between H3 and H4 that are negative in the WT vanish to zero in the mutant.

These pervasive changes in inter-residue couplings in turn modulate the relative solvent exposure of tyrosine 42 (Y42). Y42 is more solvent-exposed in the WT, with the mutation K53E restricting the conformational flexibility and hence the Y42 solvent exposure (Figure 5C). This is more evident in the plot of the Y42 solvent exposure as a function of MC steps and the corresponding structural snapshots (Figure 5D−F). We would like to highlight that the mutation K53E does not completely abrogate the structural opening event but instead significantly reduces its probability, hence minimizing the Y42 exposure on average (also see Figure 3H and the associated discussion). The observed differences in cross-correlation values are better explained in terms of interhelical distances between H1 and H3 ($d_{H1−H3}$) and between H1 and H4 ($d_{H1−H4}$) (Figure 5G and H, respectively). The WT is characterized by a larger distribution of $d_{H1−H3}$ and $d_{H1−H4}$ distances, which promotes the exposure of Y42 to the solvent. Thus, REMC simulations demonstrate that the elimination of electrostatic frustration via the K53E mutation modulates the coupling between the various structural elements, reduces the native ensemble heterogeneity, and minimizes Y42 solvent exposure.

Conformational Selection and Phosphorylation. Experiments and simulations therefore highlight that the WT FF1 domain samples multiple conformations in equilibrium, some of which are phosphorylation-competent (Figures 3H and 5D). The populations of these phosphorylation-competent states are reduced by introducing the K53E mutation. We tested this

prediction by carrying out phosphorylation assays with γ-^{32}P-labeled ATP at 298 and 310 K with platelet-derived growth factor (PDGF) receptor α-kinase. The consensus recognition motif for the kinase on the substrate is [X−X−Y*−V−F−I] with a preference for an acidic residue at the $n − 1$ position and hydrophobic residues at the $n + 1$, $n + 2$, and $n + 3$ positions,[60] where n is the tyrosine position that is phosphorylated and marked with a star. The WT FF1 domain harbors the site [QDYVYL] in the third helix, which is phosphorylated.[34,35] In the current work, the resulting bands from the phosphor image were appropriately normalized to the total amount of protein transferred to the membrane to quantify the extent of relative phosphorylation (see Methods and Figure S8). We find that the mutant K53E is consistently less phosphorylated than the WT by <50% both at 298 and 310 K (Figure 6A and B; $p < 0.02$), thus validating the expectations from experiments and predictions from simulations that enhanced fluctuations in the WT drive phosphorylation.
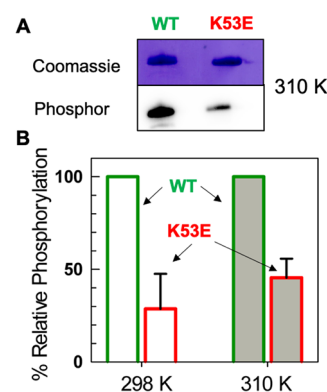

Figure 6. *In vitro* phosphorylation assays. (A) Coomassie-stained and phosphor images of WT and K53E at 310 K. Note the larger intensity for the band corresponding to the WT in the phosphor image. (B) Relative phosphorylation extents at 298 and 310 K for the WT (blue) and the K53E mutant (red).

Tuning Phosphorylation Extents via the K53Q Mutation. The K53E mutation not only eliminates frustration but also stabilizes the region of the protein around the K53 due to favorable interactions with the adjacent positively charged residues (Figure 2A). The inference from this observation is that mutating K53 to an uncharged residue should eliminate the frustration but not provide additional stabilization or modulation of the native ensemble beyond that afforded by the loss of frustration (Figure S9A). Such a mutant should exhibit a graded dependence in its conformational behavior between the WT and the K53 mutant. True to this expectation, we observe that (i) the K53Q mutant is less stable than the K53E mutant but more stable than the WT (T_m values of 319.9, 323.3, and 325.6 K for the WT, K53Q, and K53E, respectively, by far-UV CD; Figure S9B), (ii) K53Q exhibits a far-UV CD pretransition between those of the WT and the K53E mutant (Figure S9B), (iii) there is a smaller increase in the tertiary structure as evidenced by near-UV CD (Figure S9C), (iv) there is a similar decrease in the value λ_{max} compared to that of the K53E mutant (Figure S9D), and (v) the a slightly flatter pretransition as evidenced by DSC (Figure S9E) but not a two-state-like behavior as seen in the K53E mutant. *In vitro* kinase assays show proportionately higher phosphorylation extents for

the K53Q mutant at both 298 and 310 K when compared to those of the K53E mutant (Figure S9F), thus highlighting the tunable nature of the native ensemble in the FF1 domain.

A natural question that follows whether there are other residues that could be mutated to control activity. It is relatively easier to identify frustration involving charged residues through the TK calculation (Figure 2A, for example), but this is not the case for uncharged residues. One avenue to explore the extent to which every residue is frustrated, irrespective of the charge status, is through the web server Frustratometer,[61] which accounts for how favorable a particular residue environment is relative to all possible combinations of pairwise interactions at every site along the sequence. The server output reveals that there are only two regions in the FF1 domain that are highly frustrated. The most frustrated region is around the phosphorylation site Y42, which is consistent with earlier works that show functional regions are frustrated.[62,63] In addition, the server predicts a second site adjacent to the functional stretch and around K52 as the second-most frustrated region in the protein (Figure S10). Mutating K53 to Q and E progressively eliminates frustration around the second site exactly as predicted by the TK calculations, which is also consistent with experiments (Figure S10). We further chose two independent controls, namely the original FF1 protein (Figure S2, without the W13F mutation) and W13E, with the mutation to E observable in FF3 and FF4 domains (Figure S2, the third and fourth sequences, respectively). In both the cases, we observed a steep pretransition in the heat capacity profiles indicative of large conformational fluctuations in the temperatures between 280 and 310 K (Figure S11A). The phosphorylation extents at 310 K are very similar for both the variants, comparable to the WT, and importantly higher than those of the K53Q and K53E mutants (Figure S11B), further validating our hypothesis and experiments.

CONCLUSIONS

We have employed a collection of spectroscopic and calorimetric probes at different levels of resolution to show that the p190A RhoGAP FF1 domain samples diverse conformations in its native ensemble, some of which are phosphorylation-competent. Evidence is also presented from two complementary computational models, the statistical mechanical WSME model, and REMC simulations, which are internally consistent with each other and with the experiments. Enhanced fluctuations in the WT native ensemble are primarily determined by a single residue, K53, which is located in the fourth helix. K53 mediates unfavorable charge—charge interactions with positively charged residues in its neighborhood, thus locally destabilizing the structure to allow the hydroxyl group of Y42 to be exposed to the solvent. Accordingly, the K53E mutation significantly reduces the subset of conformations that are phosphorylation-competent due to decreased fluctuations, which was observed directly via simulations and indirectly by employing global and local structural probes. Finally, *in vitro* phosphorylation assays explicitly highlight that the K53E mutant is <50% phosphorylated compared to the WT, as expected from the decreased average solvent exposure of Y42.

"Gatekeeper" residues determine the extent of accessibility of a ligand or substrate to the active site and even the selectivity. Here, we demonstrate that the position K53 acts as a gatekeeper by controlling the extent of the solvent accessibility of Y42. It does so not by physical occlusion or local effects but by modulating the degree of structure and structural correlations in the native ensemble. Ensemble-derived structural cross-correlation maps accordingly highlight that the WT and the mutant display vastly different coupling extents that are not localized but instead spread across the entire structure. A ~7 K difference in melting temperatures was also observed between local and global probes for both the WT and the K53E mutant (Figure 4). This is evidence for larger dynamics in the native ensemble than that reported in the current work, which in turn decouples different structural regions in the protein even in the more stable K53E mutant. The mutation, therefore, does not fully eliminate phosphorylation but reduces the extent by <50%. The proposed conformational selection mechanism and the dynamic exposure of buried tyrosine, driven by conformational fluctuations in the native ensemble, could be a generic mechanism for phosphorylation of buried residues in agreement with recent simulations.[64]

Mutational tuning of activity is a commonly employed strategy to identify functionally coupled sites. Modulating packing interactions, through mutations in the hydrophobic core, has the drawback of reducing stability and sometimes even completely unfolding the protein. However, the long-range nature of charge—charge interactions provides the specific advantage of enabling the control of structure at a distant site without introducing a mutation close to the active site or the ligand binding region or in the hydrophobic core. Our study shows that such long-range electrostatic frustration in FF1 serves as a basis for the partial unfolding-coupled phosphorylation, further influencing processes vital for normal cell survival. The neutralization of this frustration by a single charged residue mutation is shown to affect stability, ensemble fluctuations, long-range correlations, and hence function. In fact, recent experiments have revealed that the unfolding mechanism of the protein Hha can be switched from C-terminal first to N-terminal first via the elimination of a single unfavorable electrostatic interaction.[31] Given the abundance of charged residues on the protein surface, it is reasonable to expect that such electrostatic interaction-driven control of fluctuations is prevalent in ordered proteins. Our findings should thus enable the design of protein-based molecular switches and functions through the precise modulation of long-range interactions.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.1c01548.

> Gene sequence of the FF1 domain; primer sequences; domain organization and sequence alignment of FF domains from different proteins; far- and near-UV CD thermal unfolding curves of W13F and the FF1 domain; cartoon depicting the interaction of the residue K50 with E46; far- and near-UV CD and fluorescence spectra of WT and K53 mutant; two-state fits to DSC curves; predictions from the WSME model; representative fluorescence intensity transients, and residual plots; phosphor- and Coomassie-stained membrane images at 310 K; detailed thermodynamic and functional characterization of the K53Q mutant; predictions from the

Frustratometer Web server; and control experiments (PDF)

AUTHOR INFORMATION

Corresponding Author

Athi N. Naganathan − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India;* ⊙ orcid.org/0000-0002-1655-7802; Phone: +91-44-2257 4140; Email: athi@iitm.ac.in

Authors

Hemashree Golla − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India*

Adithi Kannan − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India*

Soundhararajan Gopi − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India;* ⊙ orcid.org/0000-0002-7511-2571

Sowmiya Murugan − *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai 600036, India*

Lakshmi R Perumalsamy − *Department of Biomedical Sciences, Sri Ramachandra Institute of Higher Education and Research, Chennai 600116, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscentsci.1c01548

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

WSME, Wako−Saitô−Muñoz−Eaton
NMR, nuclear magnetic resonance
REMC, replica exchange Monte Carlo
DSC, differential scanning calorimetry
CD, circular dichroism

REFERENCES

(1) Monod, J.; Wyman, J.; Changeux, J. P. On Nature of Allosteric Transitions - a Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88−118.

(2) Changeux, J. P. Allostery and the Monod-Wyman-Changeux Model after 50 Years. *Annu. Rev. Biophys.* **2012**, *41*, 103−133.

(3) Koshland, D. E., Jr.; Nemethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5* (1), 365−385.

(4) Fenwick, R. B.; Esteban-Martin, S.; Salvatella, X. Understanding Biomolecular Motion, Recognition, and Allostery by Use of

Conformational Ensembles. *Eur. Biophys. J.* **2011**, *40* (12), 1339−1355.

(5) Kiefhaber, T.; Bachmann, A.; Jensen, K. S. Dynamics and Mechanisms of Coupled Protein Folding and Binding Reactions. *Curr. Opin. Struct. Biol.* **2012**, *22* (1), 21−29.

(6) Liu, J.; Nussinov, R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLOS Comput. Biol.* **2016**, *12* (6), e1004966.

(7) Hammes, G. G.; Chang, Y.-C.; Oas, T. G. Conformational Selection or Induced Fit: A Flux Description of Reaction Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13737−13741.

(8) Wang, Q.; Zhang, P.; Hoffman, L.; Tripathi, S.; Homouz, D.; Liu, Y.; Waxham, M. N.; Cheung, M. S. Protein Recognition and Selection through Conformational and Mutually Induced Fit. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (51), 20545−20550.

(9) Naganathan, A. N.; Orozco, M. The Conformational Landscape of an Intrinsically Disordered DNA-Binding Domain of a Transcription Regulator. *J. Phys. Chem. B* **2013**, *117*, 13842−13850.

(10) Greives, N.; Zhou, H.-X. Both Protein Dynamics and Ligand Concentration Can Shift the Binding Mechanism between Conformational Selection and Induced Fit. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (28), 10197−10202.

(11) Chu, W.-T.; Chu, X.; Wang, J. Binding Mechanism and Dynamic Conformational Change of C Subunit of PKA with Different Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (38), E7959−E7968.

(12) Sen, S.; Udgaonkar, J. B. Binding-Induced Folding under Unfolding Conditions: Switching between Induced Fit and Conformational Selection Mechanisms. *J. Biol. Chem.* **2019**, *294* (45), 16942−16952.

(13) Bahar, I.; Lezon, T. R.; Yang, L. W.; Eyal, E. Global Dynamics of Proteins: Bridging between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.

(14) Marsh, J. A.; Teichmann, S. A.; Forman-Kay, J. D. Probing the Diverse Landscape of Protein Flexibility and Binding. *Curr. Opin. Struct. Biol.* **2012**, *22* (5), 643−650.

(15) Bowman, G. R.; Geissler, P. L. Equilibrium Fluctuations of a Single Folded Protein Reveal a Multitude of Potential Cryptic Allosteric Sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (29), 11681−11686.

(16) Sikosek, T.; Chan, H. S. Biophysics of Protein Evolution and Evolutionary Protein Biophysics. *J. R. Soc. Interface* **2014**, *11* (100), 20140419.

(17) Levin, A. M.; Bates, D. L.; Ring, A. M.; Krieg, C.; Lin, J. T.; Su, L.; Moraga, I.; Raeber, M. E.; Bowman, G. R.; Novick, P.; Pande, V. S.; Fathman, C. G.; Boyman, O.; Garcia, K. C. Exploiting a Natural Conformational Switch to Engineer an Interleukin-2 "Superkine. *Nature* **2012**, *484* (7395), 529−533.

(18) López, C. J.; Yang, Z.; Altenbach, C.; Hubbell, W. L. Conformational Selection and Adaptation to Ligand Binding in T4 Lysozyme Cavity Mutants. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (46), E4306−15.

(19) Michielssens, S.; Peters, J. H.; Ban, D.; Pratihar, S.; Seeliger, D.; Sharma, M.; Giller, K.; Sabo, T. M.; Becker, S.; Lee, D.; Griesinger, C.; de Groot, B. L. A Designed Conformational Shift to Control Protein Binding Specificity. *Angew. Chem., Int. Ed. Engl.* **2014**, *53* (39), 10367−10371.

(20) Maciag, J. J.; Mackenzie, S. H.; Tucker, M. B.; Schipper, J. L.; Swartz, P.; Clark, A. C. Tunable Allosteric Library of Caspase-3 Identifies Coupling between Conserved Water Molecules and Conformational Selection. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (41), E6080−E6088.

(21) Pádua, R. A. P.; Sun, Y.; Marko, I.; Pitsawong, W.; Stiller, J. B.; Otten, R.; Kern, D. Mechanism of Activating Mutations and Allosteric Drug Inhibition of the Phosphatase SHP2. *Nat. Commun.* **2018**, *9* (1), 4507.

(22) Kumar, S.; Ma, B. Y.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci.* **2000**, *9* (1), 10−19.

(23) Vega, S.; Abian, O.; Velazquez-Campoy, A. On the Link between Conformational Changes, Ligand Binding and Heat Capacity. *Biochim. Biophys. Acta* **2016**, *1860* (5), 868−878.

(24) Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem. Rev.* **2016**, *116* (11), 6516−6551.

(25) Law, A. B.; Sapienza, P. J.; Zhang, J.; Zuo, X.; Petit, C. M. Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J. Am. Chem. Soc.* **2017**, *139*, 3599−3602.

(26) Szilagyi, A.; Zavodszky, P. Structural Differences between Mesophilic, Moderately Thermophilic and Extremely Thermophilic Protein Subunits: Results of a Comprehensive Survey. *Structure* **2000**, *8* (5), 493−504.

(27) Sanchez-Ruiz, J. M.; Makhatadze, G. I. To Charge or Not to Charge? *Trends Biotechnol.* **2001**, *19* (4), 132−135.

(28) Halskau, O.; Perez-Jimenez, R.; Ibarra-Molero, B.; Underhaug, J.; Muñoz, V.; Martinez, A.; Sanchez-Ruiz, J. M. Large-Scale Modulation of Thermodynamic Protein Folding Barriers Linked to Electrostatics. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (25), 8625−8630.

(29) Tripathi, S.; Garcìa, A. E.; Makhatadze, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J. Phys. Chem. B* **2015**, *119* (41), 13103−13112.

(30) Tsai, M. Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, Structure Prediction, and the Energy Landscapes for Protein Folding and Binding. *Protein Sci.* **2016**, *25* (1), 255−269.

(31) Narayan, A.; Gopi, S.; Lukose, B.; Naganathan, A. N. Electrostatic Frustration Shapes Folding Mechanistic Differences in Paralogous Bacterial Stress Response Proteins. *J. Mol. Biol.* **2020**, *432* (17), 4830−4839.

(32) Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V.; Danielsson, J.; Oliveberg, M. Physicochemical Code for Quinary Protein Interactions in Escherichia Coli. *Proc. Natl. Acad. Sci. U S A* **2017**, *114* (23), E4556−e4563.

(33) Tcherkezian, J.; Lamarche-Vane, N. Current Knowledge of the Large RhoGAP Family of Proteins. *Biol. Cell* **2007**, *99* (2), 67−86.

(34) Bonet, R.; Ruiz, L.; Aragón, E.; Martín-Malpartida, P.; Macias, M. J. NMR Structural Studies on Human P190-A RhoGAPFF1 Revealed That Domain Phosphorylation by the PDGF-Receptor Alpha Requires Its Previous Unfolding. *J. Mol. Biol.* **2009**, *389* (2), 230−237.

(35) Jiang, W.; Sordella, R.; Chen, G.-C.; Hakre, S.; Roy, A. L.; Settleman, J. An FF Domain-Dependent Protein Interaction Mediates a Signaling Pathway for Growth Factor-Induced Gene Expression. *Mol. Cell* **2005**, *17* (1), 23−35.

(36) Parasuraman, P.; Mulligan, P.; Walker, J. A.; Li, B.; Boukhali, M.; Haas, W.; Bernards, A. Interaction of P190A RhoGAP with EIF3A and Other Translation Preinitiation Factors Suggests a Role in Protein Biosynthesis. *J. Biol. Chem.* **2017**, *292* (7), 2679−2689.

(37) Narayan, A.; Gopi, S.; Fushman, D.; Naganathan, A. N. A Binding Cooperativity Switch Driven by Synergistic Structural Swelling of an Osmo-Regulatory Protein Pair. *Nat. Commun.* **2019**, *10* (1), 1995.

(38) Wako, H.; Saito, N. Statistical Mechanical Theory of Protein Conformation. 2. Folding Pathway for Protein. *J. Phys. Soc. Jpn.* **1978**, *44* (6), 1939−1945.

(39) Muñoz, V.; Eaton, W. A. A Simple Model for Calculating the Kinetics of Protein Folding from Three-Dimensional Structures. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (20), 11311−11316.

(40) Naganathan, A. N. Predictions from an Ising-like Statistical Mechanical Model on the Dynamic and Thermodynamic Effects of Protein Surface Electrostatics. *J. Chem. Theory Comput.* **2012**, *8* (11), 4646−4656.

(41) Gopi, S.; Lukose, B.; Naganathan, A. N. Diverse Native Ensembles Dictate the Differential Functional Responses of Nuclear Receptor Ligand-Binding Domains. *J. Phys. Chem. B* **2021**, *125* (14), 3546−3555.

(42) 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.* **2013**, *110* (44), 17880−17885.

(43) Rajasekaran, N.; Gopi, S.; Narayan, A.; Naganathan, A. N. Quantifying Protein Disorder through Measures of Excess Conformational Entropy. *J. Phys. Chem. B* **2016**, *120*, 4341−4350.

(44) PyMOL, ver. 1.2; Schrödinger LLC, 2009.

(45) Vitalis, A.; Pappu, R. V. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.* **2009**, *30* (5), 673−699.

(46) Mitov, M. I.; Greaser, M. L.; Campbell, K. S. GelBandFitter−a Computer Program for Analysis of Closely Spaced Electrophoretic and Immunoblotted Bands. *Electrophoresis* **2009**, *30* (5), 848−851.

(47) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* **2012**, *9* (7), 671−675.

(48) Tanford, C.; Kirkwood, J. G. Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* **1957**, *79*, 5333−5339.

(49) Ibarra-Molero, B.; Loladze, V. V.; Makhatadze, G. I.; Sanchez-Ruiz, J. M. Thermal versus Guanidine-Induced Unfolding of Ubiquitin. An Analysis in Terms of the Contributions from Charge-Charge Interactions to Protein Stability. *Biochemistry* **1999**, *38*, 8138−8149.

(50) Naganathan, A. N.; Muñoz, V. Thermodynamics of Downhill Folding: Multi-Probe Analysis of PDD, a Protein That Folds Over a Marginal Free Energy Barrier. *J. Phys. Chem. B* **2014**, *118* (30), 8982−8994.

(51) Cooper, A. Thermodynamic Fluctuations in Protein Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73* (8), 2740−2741.

(52) Kaya, H.; Chan, H. S. Polymer Principles of Protein Calorimetric Two-State Cooperativity. *Proteins* **2000**, *40* (4), 637−661.

(53) Muñoz, V.; Sanchez-Ruiz, J. M. Exploring Protein Folding Ensembles: A Variable Barrier Model for the Analysis of Equilibrium Unfolding Experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646−17651.

(54) Gomez, J.; Hilser, V. J.; Xie, D.; Freire, E. *Heat-Capacity of Proteins. Proteins* **1995**, *22* (4), 404−412.

(55) Privalov, P. L.; Dragan, A. I. Microcalorimetry of Biological Macromolecules. *Biophys. Chem.* **2007**, *126*, 16−24.

(56) Jemth, P.; Gianni, S.; Day, R.; Li, B.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Demonstration of a Low-Energy on-Pathway Intermediate in a Fast-Folding Protein by Kinetics, Protein Engineering, and Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (17), 6450−6455.

(57) Jemth, P.; Day, R.; Gianni, S.; Khan, F.; Allen, M.; Daggett, V.; Fersht, A. R. The Structure of the Major Transition State for Folding of an FF Domain from Experiment and Simulation. *J. Mol. Biol.* **2005**, *350* (2), 363−378.

(58) Korzhnev, D. M.; Vernon, R. M.; Religa, T. L.; Hansen, A. L.; Baker, D.; Fersht, A. R.; Kay, L. E. Nonnative Interactions in the FF Domain Folding Pathway from an Atomic Resolution Structure of a Sparsely Populated Intermediate: An NMR Relaxation Dispersion Study. *J. Am. Chem. Soc.* **2011**, *133* (28), 10974−10982.

(59) Lakshmikanth, G. S.; Sridevi, K.; Krishnamoorthy, G.; Udgaonkar, J. B. Structure Is Lost Incrementally during the Unfolding of Barstar. *Nat. Struct. Biol.* **2001**, *8*, 799−804.

(60) Pinna, L. A.; Ruzzene, M. How Do Protein Kinases Recognize Their Substrates? *Biochim. Biophys. Acta* **1996**, *1314* (3), 191−225.

(61) Parra, R. G.; Schafer, N. P.; Radusky, L. G.; Tsai, M.-Y.; Guzovsky, A. B.; Wolynes, P. G.; Ferreiro, D. U. Protein Frustratometer 2: A Tool to Localize Energetic Frustration in Protein Molecules, Now with Electrostatics. *Nucleic Acids Res.* **2016**, *44* (W1), W356−60.

(62) Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Frustration, Function and Folding. *Curr. Opin. Struct. Biol.* **2018**, *48*, 68−73.

(63) Freiberger, M. I.; Guzovsky, A. B.; Wolynes, P. G.; Parra, R. G.; Ferreiro, D. U. Local Frustration around Enzyme Active Sites. *Proc. Natl. Acad. Sci. U S A* **2019**, *116*, 4037−4043.

(64) Henriques, J.; Lindorff-Larsen, K. Protein Dynamics Enables Phosphorylation of Buried Residues in Cdk2/Cyclin-A-Bound P27. *Biophys. J.* **2020**, *119* (10), 2010−2018.