Supporting Information

De Simone et al. 10.1073/pnas.1112197108

SI Materials and Methods

Protein Expression and Purification. The gene for the acylphosphatase from *Drosophila melanogaster* (AcPDro2) was cloned into pGEX-4T1 (Amersham Biosciences), and the protein expressed as a fusion with GST in DH5α *Escherichia coli* cells by induction with 0.1 mM IPTG. The resulting protein was purified by glutathione-sepharose affinity chromatography and subjected to thrombin cleavage to remove GST. The purity of the cleaved protein was monitored using 15% SDS-PAGE and its exact mass defined using electrospray mass spectroscopy. The protein was stored in 50 mM acetate buffer containing 2 mM DTT at pH 5.5. Protein concentrations were determined spectrophotometrically using ε_{280} values of 1.09 mL mg^{-1} cm^{-1}. Protein samples containing ^{15}N or ^{13}C-^{15}N were expressed in DH5α *E. coli* cells and grown in minimal media containing 1 g/L $^{15}NH_4Cl$ and 1 g/L $^{15}NH_4Cl$, 3 g/L ^{13}C D-glucose, respectively.

Relaxation Experiments. Standard pulse sequences were used for T_1 and T_2 experiments (S1), except that the T_2 experiments incorporate a watergate sequence (S2) to improve water suppression. ^{15}N-^{1}H heteronuclear NOE experiments were carried out by modifying the standard pulse sequence with a watergate sequence (S3) and a water flip-back pulse, to minimize the effect of the slowly relaxing water magnetization on the NOEs measured for amides with rapidly exchanging protons (S4). Residue-specific heteronuclear ^{1}H-^{15}N NOE values were determined from the ratio of peak intensities in spectra recorded with and without saturation of the ^{1}H resonances. T_1 and T_2 values were obtained by fitting single exponential decays to the experimental data; the fitting of experimental data and the error analyses were performed with the program SPARKY. A model-free analysis of the T_1, T_2, and NOE data was used to define backbone amide order parameters S^2 (S5).

Structural Ensemble Refinement by H/D Exchange Data. The method employed for the refinement of the structural ensembles enforces a pseudoenergy term (S6) to the standard force fields employed in molecular dynamics. This term has the role of minimizing the discrepancy between calculated and experimental observables. Herein, the protection factors are accounted by the following phenomenological model (S7)

$$\ln P_i^{\mathrm{sim}}(C) = \beta_C N_i^C(C) + \beta_h N_i^h(C). \qquad \text{[S1]}$$

The protection factor of a residue i in a particular conformation, C, relative to an unstructured peptide is treated as the contribution from burial [measured as the number of heavy atoms within a distance of 6.5 Å from the amide nitrogen $N_i^c(C)$ under consideration and from hydrogen bonding to the amide N_i^h]. The weighting factors β_c and β_h have been previously calibrated (S7) and are applied to the two terms. The calculated protection factors (Eq. 5) are taken as averages over M replicas of the molecule; i.e.,

$$\overline{\ln P_i^{\mathrm{sim}}} = \frac{1}{M}\sum_k \ln P_i^{\mathrm{sim}}(C_k), \qquad \text{[S2]}$$

where the replicas have conformations C_k ($k = 1,\ldots,M$). Overall the pseudoenergy term is given by

$$\rho = \sum_i (\overline{\ln P_i^{\mathrm{sim}}} - \ln P_i^{\mathrm{exp}})^2, \qquad \text{[S3]}$$

where the protection factor restraints are applied as an average over an ensemble of conformations representing the states occupied by the protein.

Restrained sampling was performed by following a widely employed protocol (S7, S8). Ensemble simulations using four replicas were performed by using the CHARMM program with the CHARMM22 force field (S9). The calculations were initiated from the X-ray structure of AcPDro2 (Protein Data Bank ID 1URR) (S10) by immersing the protein in a 6-Å shell of TIP3 water molecules, using a boundary potential to prevent water molecules from escaping. All calculations used an atom-based truncation scheme with a list cutoff of 14 Å, a nonbonding cutoff of 12 Å,
with a Lennard–Jones smoothing function initiated at 10 Å. Covalent bonds were constrained with SHAKE. The initial velocities were randomly assigned from a Maxwell–Boltzmann distribution at 298 K with a different random seed for each of the four replicas. An initial equilibration simulation was run at 298 K, during which the agreement between calculated and experimental data, represented by their mean squared deviation "ρ," was allowed to converge. This objective was achieved by gently raising the restraint force constant. Subsequently, a series of 600 cycles of simulated annealing between 298 and 498 K was carried out to ensure that conformational space was sampled effectively. In the high-temperature sampling, the restraining force constant was set to a low value to allow extensive exploration of conformational space. Each cycle was carried out for 300 ps by using an integration step of 2 fs. The ensemble structures were extracted from the last 100 ps of each cycle (spaced 1 ps) at the sampling temperature of 298 K and the maximum restraining force and, to ensure convergence of the calculations, the initial 100 cycles were not considered in the analysis. As a result, for each analysis, 200,000 of sampled conformations where employed.

Error Analysis of the Free Energy Calculations. Statistical errors on the free energies values, calculated from the Boltzmann populations of the ensembles as projected onto specific reaction coordinates, were analyzed. A detailed description of the statistical error herein employed can be found in ref. S11. Briefly, we expressed the statistical error associated to the free energy value $G(Z_k,\Delta)$ as a confidence interval (S12). For a given volume Δ in the phase space defined by the reaction coordinates Z_k, the observed population is defined as

$$P(Z_k,\Delta) = \frac{n_k}{N}, \qquad \text{[S4]}$$

where N is the total number of sampled configurations (200,000 in each sampling) and n_k is the number of protein configurations populating the volume Δ. The confidence interval in the population is therefore expressed as

$$P\pm(Z_k,\Delta)=\frac{P(Z_k,\Delta)+\frac{\Phi^2}{\sqrt{N}}\pm\sqrt{P(Z_k,\Delta)(1-P(Z_k,\Delta))+\frac{\Phi^2}{4N}}}{1+\frac{\Phi^2}{N}}. \quad [S5]$$

The confidence interval in the free energy is therefore given by

$$G\pm(Z_k,\Delta)=-\frac{1}{\beta}\ln(P\pm(Z_k,\Delta)). \quad [S6]$$

The estimate of the free energy error is then given by the difference between the upper and lower bounds, $G_+(Z_k,\Delta)$ and $G_-(Z_k,\Delta)$:

$$G_+(Z_k,\Delta)-G_-(Z_k,\Delta)=-\frac{1}{\beta}\ln\left(\frac{P(Z_k,\Delta)+\frac{\Phi^2}{\sqrt{N}}-\sqrt{P(Z_k,\Delta)(1-P(Z_k,\Delta))+\frac{\Phi^2}{4N}}}{P(Z_k,\Delta)+\frac{\Phi^2}{\sqrt{N}}+\sqrt{P(Z_k,\Delta)(1-P(Z_k,\Delta))+\frac{\Phi^2}{4N}}}\right). \quad [S7]$$

1. Farrow NA, et al. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15 N NMR relaxation. *Biochemistry* 33:5984–6003.
2. Piotto M, Saudek V, Sklenar V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661–665.
3. Grzesiek S, Bax A (1993) The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J Am Chem Soc* 115:12693–12594.
4. Loria JP, Rance M, Palmer AG (1999) Relaxation-compensated Carr-Purcell-Meiboom-Gill experiments for characterizing kinetic processes by NMR spectroscopy. *J Am Chem Soc* 121:2331–2332.
5. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559.
6. De Simone A, Richter B, Salvatella X, Vendruscolo M (2009) Toward an accurate determination of free energy landscapes in solution states of proteins. *J Am Chem Soc* 131:3810–3811.
7. Best RB, Vendruscolo M (2006) Structural interpretation of hydrogen exchange protection factors in proteins: Characterization of the native state fluctuations of CI2. *Structure* 14:97–106.
8. Gsponer J, et al. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci USA* 103:99–104.
9. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30:1545–1614.
10. Zuccotti S, et al. (2004) Three-dimensional structural characterization of a novel *Drosophila melanogaster* acylphosphatase. *Acta Crystallogr D Biol Crystallogr* 60:1177–1179
11. Kobrak MN (2003) Systematic and statistical error in histogram-based free energy calculations. *J Comput Chem* 24:1437–1446.
12. Larsen R, Marx M (1986) *An Introduction to Mathematical Statistics and Its Applications* (Prentice Hall, Englewood Cliffs), 2nd Ed, NJ:278.

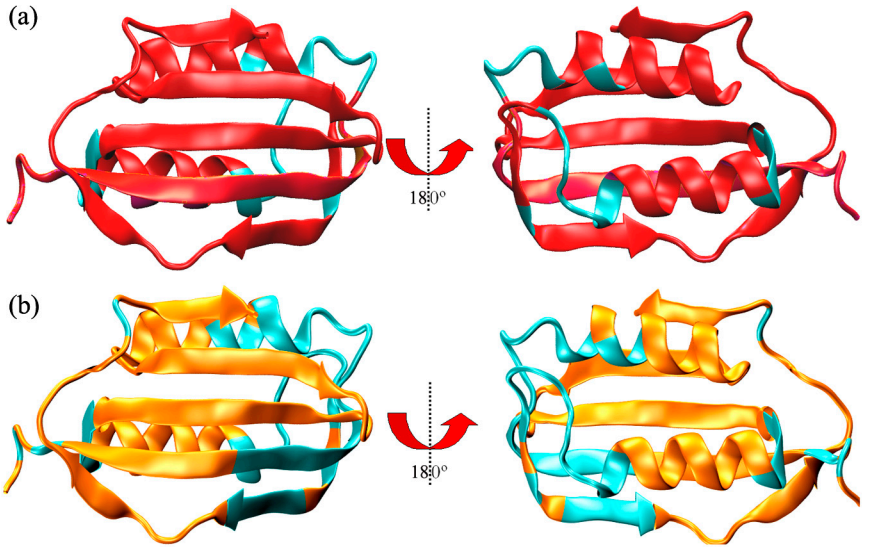


Fig. S1. AcPDro2 assignment. (*A*) Phosphate-bound state (condition *C*). Residues with assigned resonances are shown as red ribbons. Residues when resonances are missing in the spectra or overlapping in the heteronuclear single quantum coherence (HSQC) spectra are marked by cyan ribbons. (*B*) Ligand-free state (condition *A*). Residues with assigned resonances are shown as orange ribbons. Residues when resonances are missing in the spectra or overlapping in the HSQC spectra are marked by cyan ribbons. The conditions employed in this work are denoted A (acetate buffer and 0% TFE, trifluoroethanol), B (acetate buffer and 5% TFE), C (phosphate buffer and 0% TFE), and D (phosphate buffer and 5% TFE).

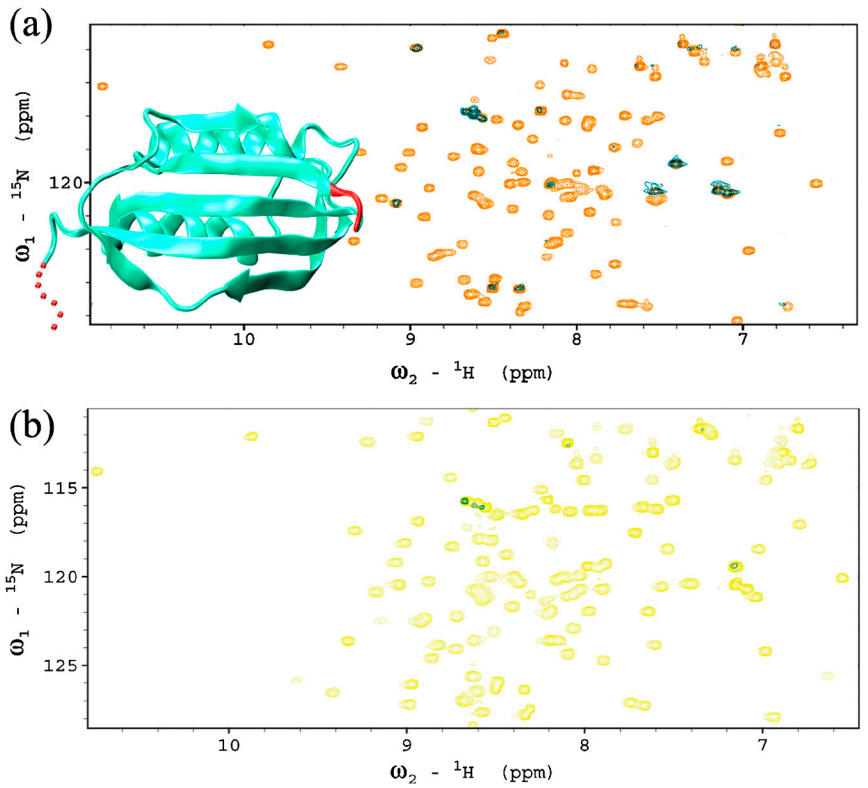


Fig. S2. Phase-modulated clean chemical exchange (CLEANEX-PM) spectra of AcPDro2. (*A*) Protein in acetate buffer, ligand-free state (condition *A*). ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) (orange) and CLEANEX spectra (green peaks). Backbone amides showing CLEANEX peaks are drawn on the representative structure (red), and are localized in the unstructured N terminus (residues A2, G3, S4, G5, V6) and in the loop connecting strands S2 and S3 (residues T46, R47, and D48). (*B*) Phosphate-bound state of the protein (condition *C*). All CLEANEX peaks belong to side chains.

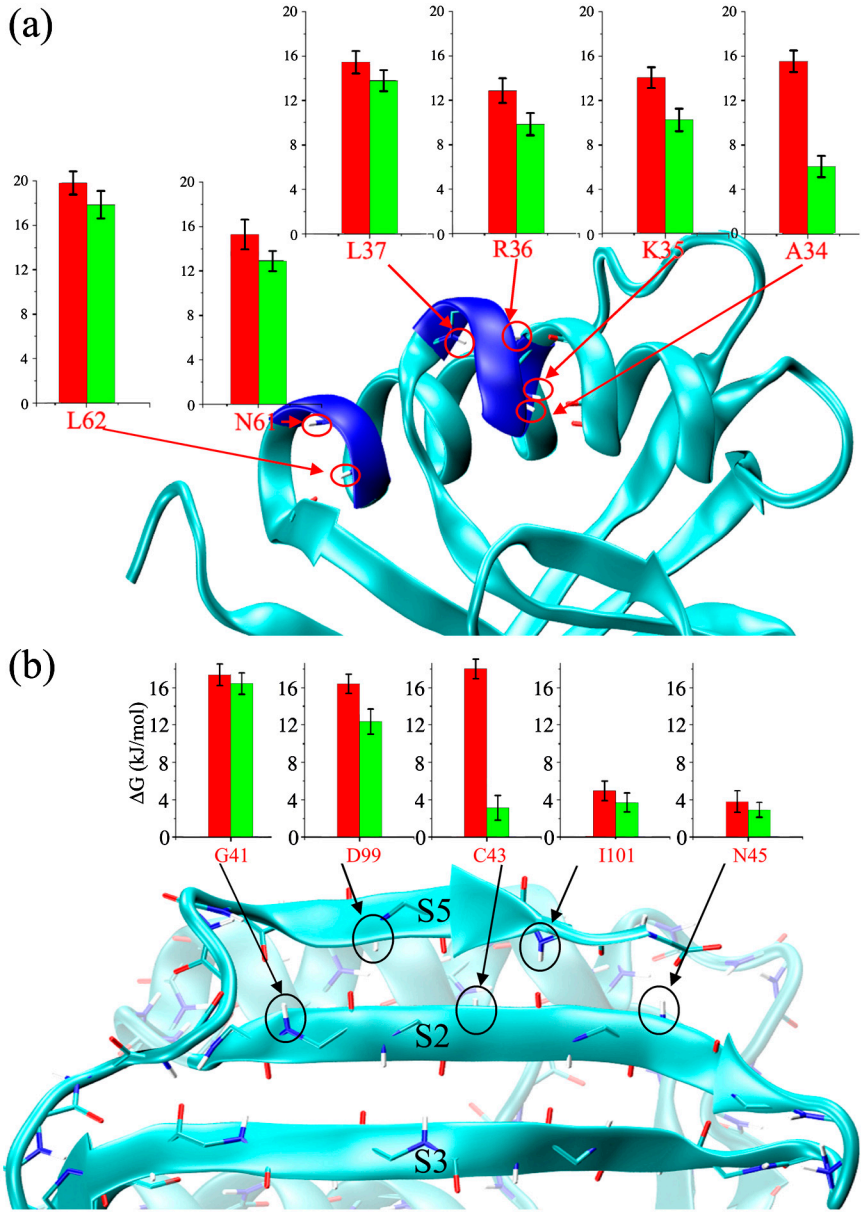
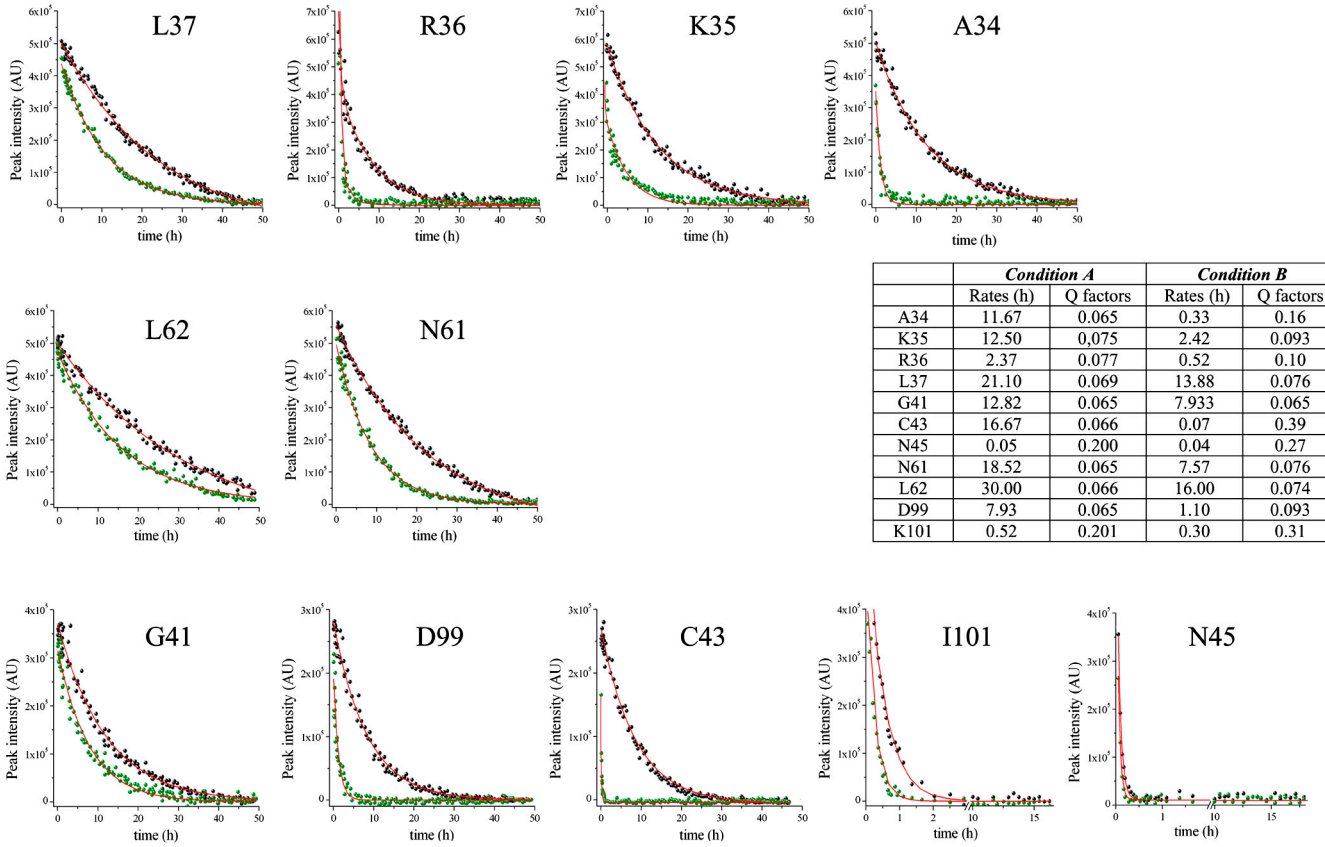


Fig. S3. Structural clusters showing the reduction of the protection factors in condition *A* (red bars) and condition *B* (green bars). (*A*) Helices H1 and H2. The protein is represented by cyan ribbons, whereas blue marks highlight regions showing major drops in protection factors. (*B*) Interface between strands S2 and S5. The amides involved in the β-sheet hydrogen bonding network are highlighted.



	Condition A		*Condition B*	
	Rates (h)	Q factors	Rates (h)	Q factors
A34	11.67	0.065	0.33	0.16
K35	12.50	0,075	2.42	0.093
R36	2.37	0.077	0.52	0.10
L37	21.10	0.069	13.88	0.076
G41	12.82	0.065	7.933	0.065
C43	16.67	0.066	0.07	0.39
N45	0.05	0.200	0.04	0.27
N61	18.52	0.065	7.57	0.076
L62	30.00	0.066	16.00	0.074
D99	7.93	0.065	1.10	0.093
K101	0.52	0.201	0.30	0.31

Fig. S4. Fitting curves of the experimental decays of the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) intensities in experiments of H/D exchange. Data order reflect the disposition of Fig. S3. Black and green points refer to conditions *A* and *B*, respectively. Curve fittings are performed by using a single exponential decay. Fitted decay rates are reported in the inset table. Q factors of the fittings are used as estimates of the errors of the protection factors.

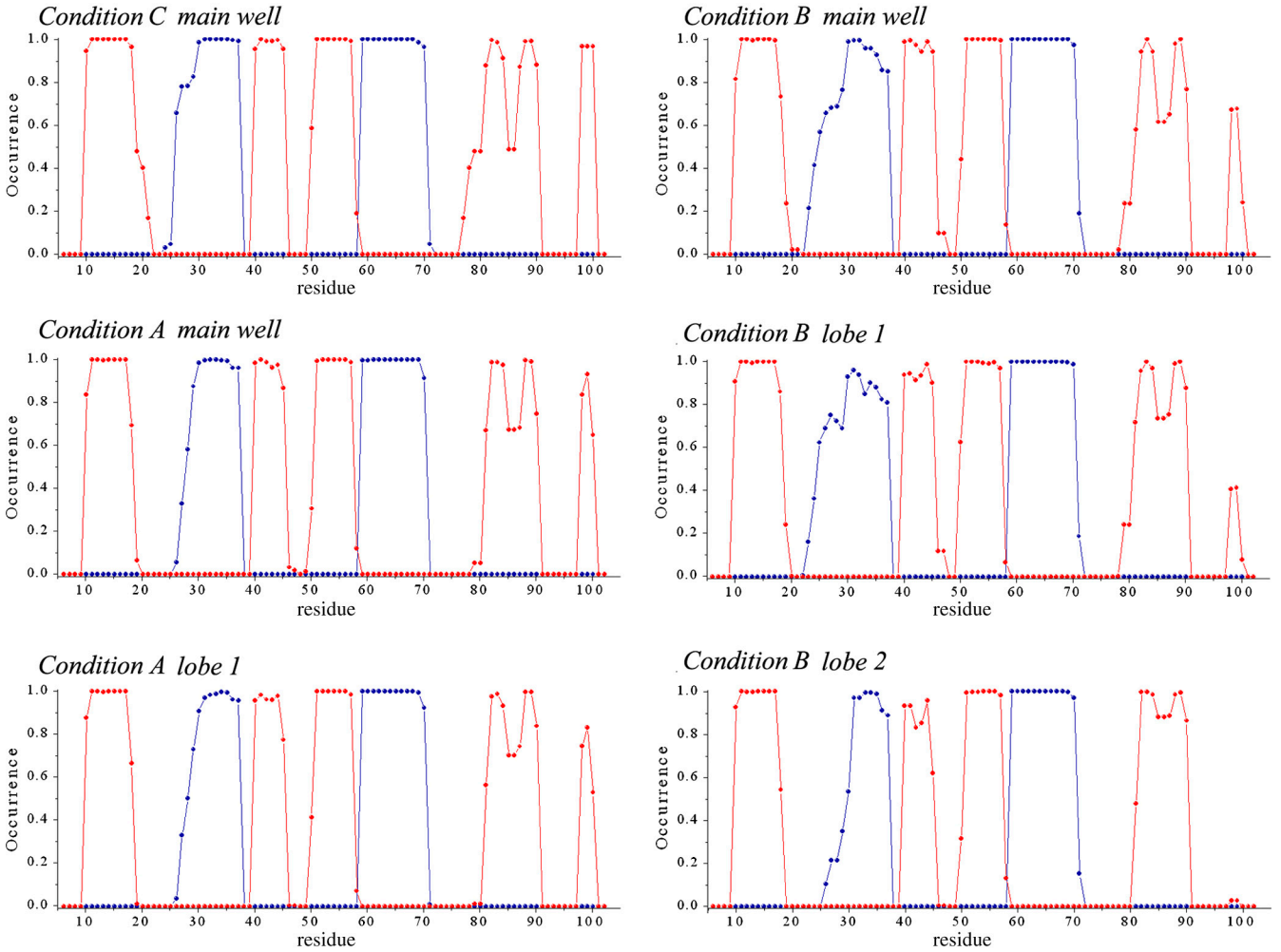


Fig. S5. Occurrence of secondary structure in the conformational wells of the different solution states of the protein. The populations of β-strands and α-helix are shown by red and blue lines, respectively.

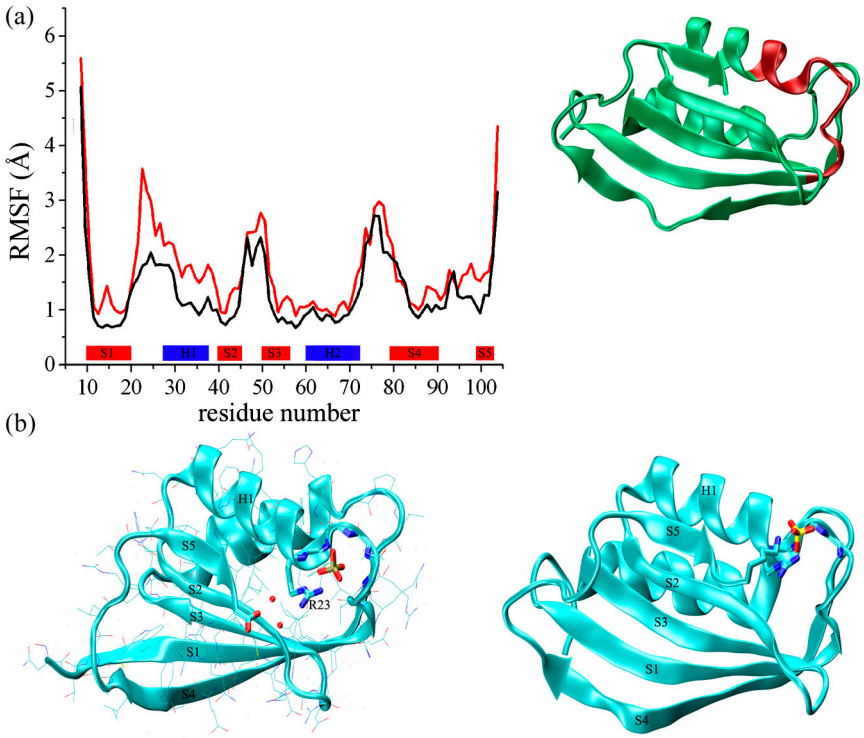


Fig. S6. (*A*) Root-mean-square fluctuations (RMSF) calculated for N and H atoms for the ligand-free and phosphate-bound states at 0% trifluoroethanol (conditions *A* and *C*). The region showing the largest differences in the fluctuations of the two states is highlighted in red on the structure. (*B*) Crystal structures of ligand bound states in the acylphosphatease family. Protein Data Bank ID codes: 1GXU (*Left*) and 2ACY (*Right*).

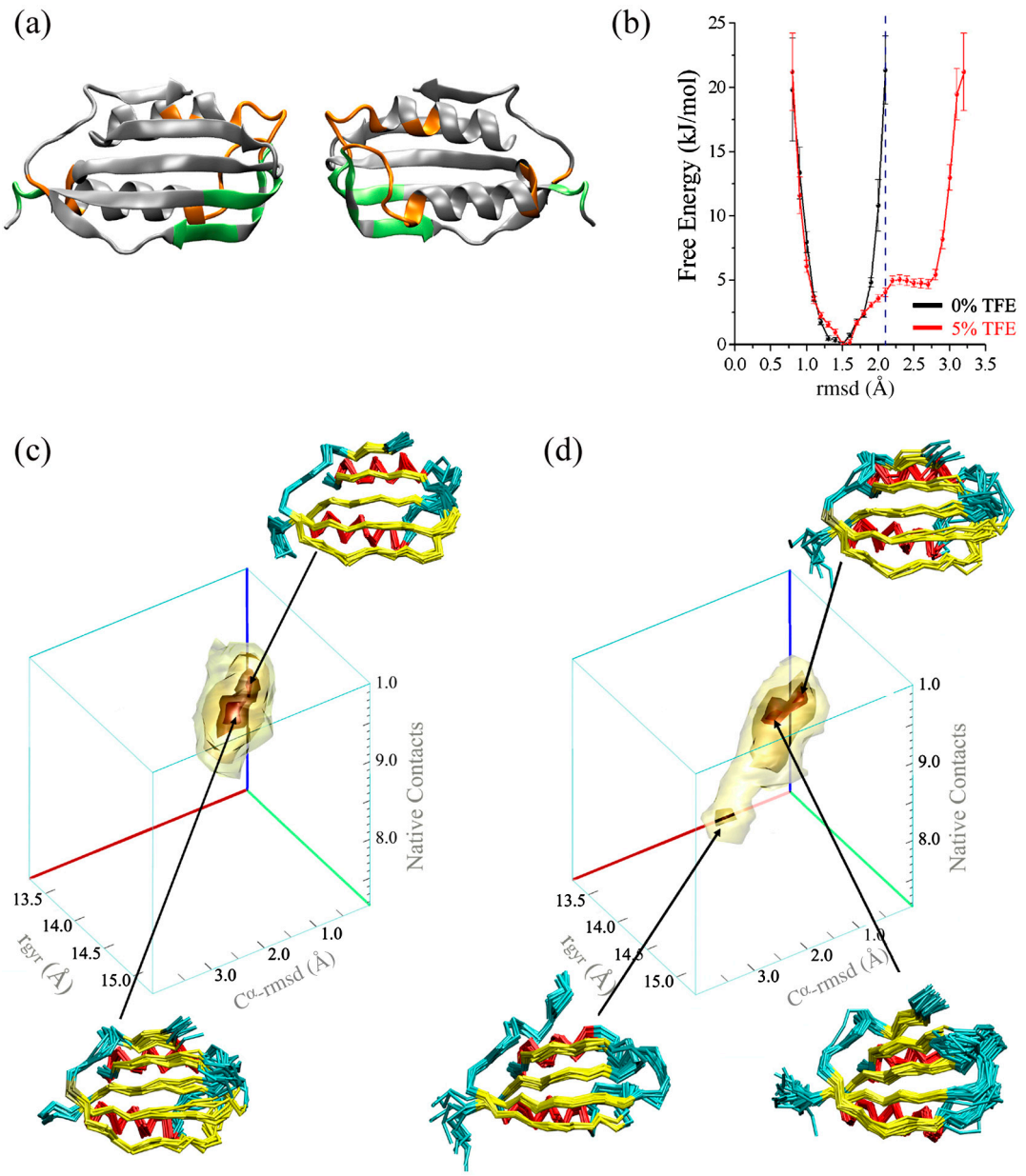


Fig. S7. Control simulation of conditions *A* and *B*. (*A*) The AcPDro2 structure. Residues for which experimental protection factors were measured in both 0% and 5% trifluoroethanol (TFE; vol/vol) are shown in gray. Residues for which the protection factors have been estimated from the data measured in the presence of phosphate (condition *C*) are shown in green. In particular, missing data in conditions *A* and *B* were recovered from condition *C* by applying a scaling factor arising from the average ratio between protection factors measured in conditions *A* and *C*. (*C* and *D*) The free energy surfaces of the control simulations of conditions *A* and *C*. Color codes as in Fig. 2. The results largely match with the original findings except for a higher content of native contacts in both 0% and 5% TFE, which is likely to arise from the contribution of the additional protection factors. This control simulation evidences that the energy barriers are not affected by the lack of experimental data of some resonances from the backbone under such conditions.

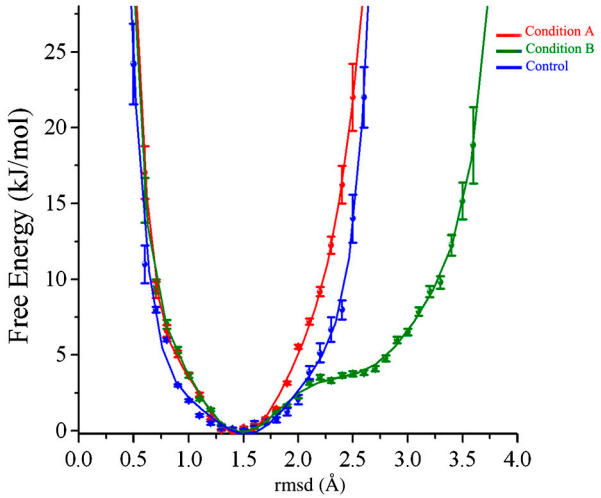


Fig. S8. One-dimensional energy landscape on a control simulation. In this control, the protection factors for condition *A* (0% trifluoroethanol, TFE) are randomly altered to produce a similar amount of deprotection as measured for condition *B* (5% TFE). Red and green lines report the unperturbed free energy profiles for condition *A* and *B*. The blue line reports the control simulation with randomly decreased protection factors. This plot shows a wider conformational space of the condition *A* ensemble but lacks the features evidenced in the ensemble representing condition *B*.

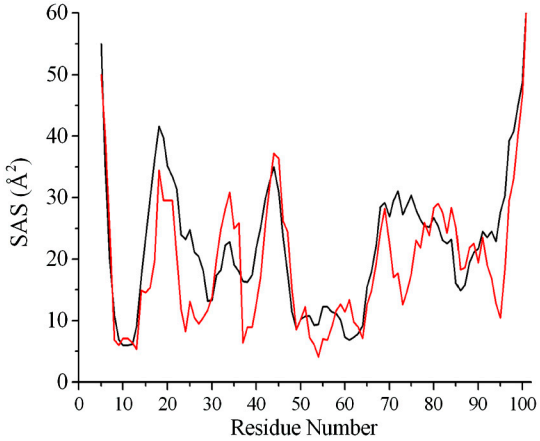


Fig. S9. Main-chain solvent accessible surface (SAS) area of the *N* (red) and *N** (black) states of AcPDro2.