Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility

Alfonso De Simone[a], Anne Dhulesia[b], Gemma Soldi[c], Michele Vendruscolo[b], Shang-Te Danny Hsu[d,e], Fabrizio Chiti[c,1], and Christopher M. Dobson[b,1]

[a]Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, United Kingdom; [b]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom; [d]Institute of Bioinformatics and Structural Biology National Tsing Hua University, Hsinchu 30013, Taiwan; [e]Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan; and [c]Department of Biological Sciences, University of Florence, 50134 Florence, Italy

Edited by Peter G. Wolynes, University of California, San Diego, La Jolla, CA, and approved October 19, 2011 (received for review July 28, 2011)

The identification of the factors that enable normally folded proteins to remain in their soluble and functional states is crucial for a comprehensive understanding of any biological system. We have determined a series of energy landscapes of the acylphosphatase from *Drosophila melanogaster* under a variety of conditions by combining NMR measurements with restrained molecular dynamics simulations. We thus analyzed the differences in the structures, dynamics, and energy surfaces of the protein in its soluble state or in situations where it aggregates through conformational states that have native-like structure, folding stability, and enzymatic activity. The study identifies the nature of the energy barriers that under normal physiological conditions prevent the protein ensemble from populating dangerous aggregation-prone states. We found that such states, although similar to the native conformation, have altered surface charge distribution, alternative topologies of the β-sheet region, and modified solvent exposure of hydrophobic surfaces and aggregation-prone regions of the sequence. The identified barriers allow the protein to undergo functional dynamics while remaining soluble and without a significant risk of misfolding and aggregation into nonfunctional and potentially toxic species.

protein misfolding | free energy barriers | avoidance of protein aggregation | molecular simulations

The majority of proteins have evolved to adopt distinctive and well-defined functional states under physiological conditions, either as monomers or as complexes. The structures corresponding to these states are encoded in the sequence as is the crucial ability of the molecules to remain soluble within the crowded cellular environment (1–5). It is increasingly evident, however, that even under physiological conditions the aggregated states of proteins, such as the highly ordered amyloid form, can be thermodynamically more stable than native states (6–8), indicating that kinetic factors are of key importance in enabling protein homeostasis to be maintained (1, 9–11).

Proteins in vivo only rarely convert into aberrant aggregated states, such as those associated with pathological conditions such as Alzheimer's disease and type II diabetes, despite their inherent tendency to do so in vitro (5). It is thus clear that all living systems rely on a large variety of regulatory mechanisms, such as molecular chaperones (12, 13) and quality control processes, which act to inhibit aggregation (14, 15) and for triggering the degradation of partially unfolded proteins (16). But the primary mechanism of maintaining functional and soluble states of proteins is encoded in the sequence and involves the existence of intrinsic energy barriers that prevent the conversion into an aggregation-prone state (1, 5, 17). It is for these barriers that the majority of proteins are able to avoid aggregation under normal physiological conditions despite considerable evidence that show that amyloid formation in vitro is a common characteristic of normally folded proteins (10, 18).

In the present study, we used a combination of NMR experiments and molecular dynamics simulations to identify the characteristic features of the free energy landscapes that enable the majority of the proteins to avoid aggregation under physiological conditions. We chose for this scope the acylphosphatase from *Drosophila melanogaster* (AcPDro2) because this is a particularly well-suited system for investigating the molecular strategies used by living systems for the maintenance of protein solubility. AcPDro2 in its native state is a globular and monomeric protein with a structure consisting of five β-strands (S1–S5), which form a single β-sheet, and two α-helices (H1 and H2) that lie adjacent to this α-sheet. The importance that subtle intrinsic factors play in enabling this protein to remain soluble is clearly shown by the fact that a very low concentration (5% vol/vol) of trifluoroethanol (TFE) is sufficient to induce rapid formation of amyloid fibrils, although the protein still populates a highly native-like conformational ensemble before aggregation occurs (19). Indeed, under these conditions, the hydrodynamic radius, intrinsic fluorescence, secondary structure content, and enzymatic activity of AcPDro2 in its monomeric state are indistinguishable with those of the protein in the absence of TFE, where the propensity of AcPDro2 to aggregate is extremely low (19). Moreover, within experimental error, AcPDro2 has the same thermodynamic stability (i.e., the same free energy of unfolding, $\Delta G_{U\text{-}F}$) in the presence and in the absence of 5% (vol/vol) TFE as determined using urea-induced denaturation at equilibrium (19). The similar thermodynamic stabilities in 0% and 5% (vol/vol) TFE originate from the fact that the folding and unfolding rates are accelerated to similar extents following the addition of 5% (vol/vol) TFE (19). By contrast to AcPDro2, most folded proteins aggregate in the presence of much higher concentrations of TFE (15–30%) where a significant portion of the molecules are unfolded or strongly destabilized.

In the present study, we have explored the factors that enable AcPDro2 in its native state to remain soluble for long periods of time under some conditions but to aggregate rapidly under others. We have therefore determined the structural ensembles and free energy surfaces under a series of different conditions by combining NMR data from relaxation and H/D exchange experiments, with statistical mechanics techniques based on restrained molecular dynamics simulations. Analysis of these energy surfaces has enabled us to define a series of specific structural and dynamical factors along with the energy barriers that enable the protein to maintain its functional native state.

Author contributions: A.D.S., F.C., and C.M.D. designed research; A.D.S., A.D., G.S., and S.-T.D.H. performed research; A.D.S., A.D., M.V., S.-T.D.H., F.C., and C.M.D. analyzed data; and A.D.S., M.V., F.C., and C.M.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: fabrizio.chiti@unifi.it or cmd44@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112197108/-/DCSupplemental.

Results

Assignment and Analysis of NMR Spectra. All NMR measurements were carried out at 298 K and pH 4.0. This pH is close to that used in the determination of the AcPDro2 crystal structure (pH 4.5, ref. 20), and under these conditions all protein states of interest in this study, including those in 5% TFE, are stable in solution for a sufficient time to enable the relevant NMR experiments to be performed prior to aggregation. We checked by electrospray ionization–mass spectrometry that under these conditions the H/D exchange of AcPDro2 occurs in an EX2 regime, which ensures that the exchange data provide a measure of the local equilibrium constant between "open" (solvent-accessible) and "closed" (solvent-inaccessible) states of the protein amide groups (21–25). Experiments were carried out in either 30 mM acetate buffer or 30 mM phosphate buffer and in the presence or absence of 5% (vol/vol) TFE. These conditions are denoted *A* (acetate buffer and 0% TFE), *B* (acetate buffer and 5% TFE), *C* (phosphate buffer and 0% TFE), and *D* (phosphate buffer and 5% TFE). In the case of phosphate buffer solutions, a phosphate ion is bound to the enzyme active site (26).

Assignments of the NMR spectra were made in 30 mM phosphate buffer (condition *C*) by means of triple resonance experiments [HNCA, CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO, and ^{15}N-total correlation spectroscopy–heteronuclear single quantum coherence (TOCSY-HSQC) spectra]. The analysis of the spectra, following the procedure described in ref. 27, resulted in the identification of main-chain resonances ($^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, ^{1}HN) for 88 out of 101 residues (Fig. S1*A*). Most of the unassigned residues are in loop regions of the protein, although some are in helices H1 and H2 and strand S1 (Fig. S1*A*). Assignments in the ^{1}H-^{15}N-HSQC spectra in acetate buffer (condition *A*) were then derived by recording spectra at different concentrations of phosphate ions, such that the chemical shifts of the peaks could be correlated with those under condition *C*. Similarly, assignments of ^{1}H-^{15}N-HSQC spectra in the presence of 5% TFE were derived from the respective spectra in the absence of TFE. The ^{1}H-^{15}N-HSQC spectra of the ligand-free state (condition *A*) contain a smaller number (72) of resolved peaks than the spectra of the phosphate-bound state (condition *C*). The peaks that are not detected in the ligand-free state are mainly located in the ligand-binding pocket and in regions spatially close to it (Fig. S1*B*), and they are likely to be affected by exchange broadening resulting from the greater flexibility and disorder of these regions of the ligand-free state of the protein. It is worth noting that conditions *A* and *B* resulted in the same number of assigned main-chain resonances.

NMR Measurements of AcPDro2 Under Conditions A to D. In order to begin to probe the differences between the solution conformations of AcPDro2 in the soluble or aggregation-prone states, we examined the chemical shift values in the spectra under conditions *A* and *B* (Fig. 1*A*). The analysis indicates small but significant changes for residues W42, N45, S85, I87, Q88, and I100 (from strands S2, S4, and S5), residues H32, R36, and L62 (from helices H1 and H2), and residues R47 and T96 (from loop regions).

In addition to measurements of chemical shifts, the exchange rate constants k_{obs} of the backbone amide hydrogens were determined for 59 residues under condition *A* (Fig. 1*B*) by direct observation of the decay in the intensities of peaks in the ^{1}H-^{15}N-HSQC spectra as a function of time after dilution into deuterated buffer. In the case of the remaining amides, the hydrogens of 13 residues exchanged effectively completely prior to acquisition of the first ^{1}H-^{15}N HSQC spectrum, preventing the determination of exchange rate constants k_{obs} from the fitting of the decay curves of the ^{1}H-^{15}N HSQC signals. For some of these residues, however, the rate constants k_{obs} could be estimated from phase-modulated clean chemical exchange (CLEANEX-PM, ref. 28) experiments (see *Materials and Methods* and Fig. S2). These values of k_{obs} allowed the amide hydrogen protection factors to be determined for eight residues located at the disordered N terminus of the protein (residues A2, G3, S4, G5, V6) and in the loop between strands S2 and S3 (residues T46, R47, and D48). Under condition *B*, 58 k_{obs} values could be measured by following the decays of ^{1}H-^{15}N-HSQC spectra, whereas 14 amide protons exchanged prior to acquisition of the first spectrum (Fig. 1*B*). Similar to condition *A*, k_{obs} rate constants of eight backbone amides (residues A2, G3, S4, G5, V6, T46, R47, and D48) could be obtained by the CLEANEX-PM method.

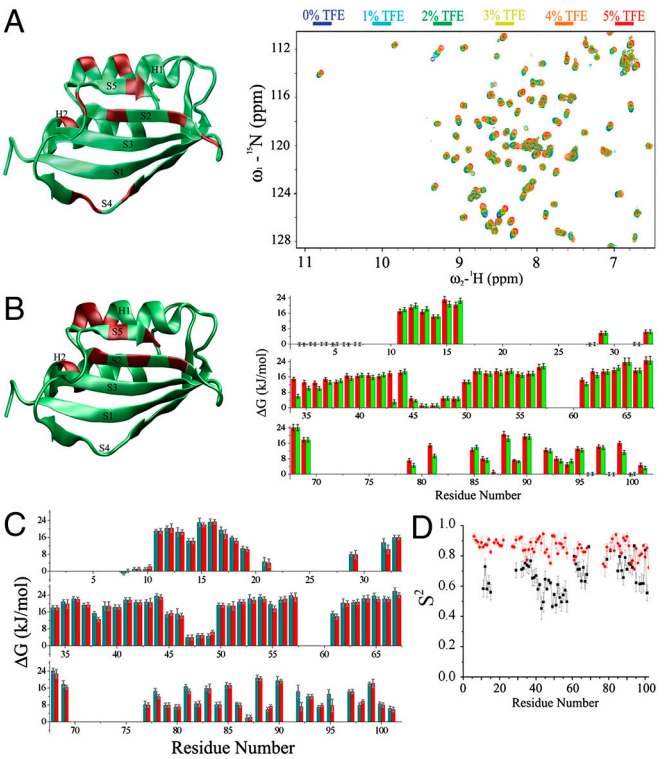


Fig. 1. NMR measurements of AcPDro2 under conditions where it is highly stable in the monomeric state (*A*, condition *A*) or readily self-assembles into amyloid fibrils (*B*, condition *B*) or bound to phosphate ions (*C*, conditions *C* and *D*). (*A*) Spectral changes upon addition of TFE (0% to 5%, condition *A* to *B*, respectively). The X-ray structure of AcPDro2 is shown on the left of the panel. Regions shown in red on the structure indicate residues with the largest (at least 1 ppm) changes in the ^{1}H-^{15}N-HSQC spectra. (*B*) Backbone amide protection factors reported as Δ*G* values. Red and green bars refer to 0% (condition *A*) and 5% (condition *B*) TFE, respectively. The structure of AcPDro2 is drawn on the left of the panel. Regions in red indicate residues showing the largest reductions of the protection factors (Fig. S2). (*C*) Phosphate-bound AcPDro2 solution states (condition *C* and *D*). Backbone amide protection factors are reported as Δ*G* values. Red and green bars refer to 0% (condition *C*) and 5% (condition *D*) TFE, respectively. (*D*) S^2 order parameters measured for the ligand-free (black) and phosphate-bound states (red).

Many of the protection factors are essentially unchanged in the aggregation-prone and resistant states but some clusters of residues show a higher degree of protection in the aggregation-resistant state (condition *A*) than in the aggregation-prone state (condition *B*, Fig. 1*B* and Figs. S3 and S4). These clusters include amide protons of residues G41, C43, N45, D99, and I101, with major reductions in the protections for residues C43 and D99 whose interactions represent the core of the network of H bonding at the interface between strands S2 and S5; this result indicates that the interface between these two strands is stabilized under conditions where the protein is aggregation resistant (Figs. S3*B*, and Fig. S4 for the decay profiles). Higher protection factors in the absence of 5% TFE are also found for residues clustered in helix H1 (residues 34–7) and in helix H2 (residues N61

and L62), albeit to a lesser extent (Figs. S3*A*, and Fig. S4 for the decay profiles). The lack of changes for the majority of the amide protection factors is consistent with previous measurements that revealed no significant differences in the folding stability in the presence or absence of 5% (vol/vol) TFE (19).

By following the decays of-HSQC spectra in the phosphate-bound state (conditions *C* and *D*), k_{obs} values for 72 residues could be determined; in addition, amides of 16 residues were found to exchange prior to acquisition of the first spectrum, none of which was detectable in CLEANEX-PM experiments at mixing times up to 25 ms (Fig. S2*B*). The protection factors calculated from the 72 k_{obs} values measured for the phosphate-bound condition *C* (Fig. 1*C*) are on average approximately 4 kJ/mol higher than those measured under the ligand-free condition *A* (Fig. 1*B*). Moreover, for the phosphate-bound state, protection factors measured in 0% and 5% TFE are virtually identical, a result that correlates with the finding that phosphate binding induces resistance to aggregation such that the protein is no longer prone to amyloid formation even in 5% (vol/vol) TFE.

Energy Landscapes of AcPDro2. The exchange protection factors discussed above were then used as the basis for generating ensembles of structures of AcPDro2 under the four sets of conditions studied in this work. This approach combines experimental NMR data and molecular dynamics simulations in order to obtain detailed conformational descriptions that incorporate both structural and dynamical information (29–32); this is achieved by performing restrained molecular dynamics simulations that enforce a pseudoenergy term in such a way as to maximize the agreement between experimental data and the NMR observables back-calculated from the ensembles of structures (33). Such a procedure assists conformational sampling and ensures that an enhanced representation of the solution state free energies is generated (29). Because the present restrained samplings enforce NMR observables (i.e., H/D exchange protection factors) that reflect fluctuations occurring in the millisecond timescale (and beyond) (21–24), the resulting structural ensembles are particularly valuable for describing backbone dynamics relevant to the processes of protein folding and misfolding, which are those that affect the maintenance of protein solubility and are governed by events that typically occur in these timescales (34, 35).

The ensembles generated from the NMR data recorded under the different conditions were analyzed by defining structural parameters that provide independent measures of the key conformational features of the various states of the protein. The choice of these parameters is crucial for describing appropriately both the molecular conformations and the Boltzmann populations of the species composing the ensembles. Accordingly, we selected for this purpose three specific "coordinates" [namely, the rmsd of C^{α} atoms from the crystal structure (c_1), the radius of gyration (c_2), and the fraction of native contacts (c_3)]. By calculating the distribution of states in these three dimensions, it has been possible to generate free energy surfaces of AcPDro2 from the Boltzmann populations of the ensembles as projected onto these three coordinates (Fig. 2). The free energy maps have been generated from 200,000 conformations in each sampling. A detailed error analysis has been performed in order to define the statistical significance of the energy barriers herein reported (see *SI Materials and Methods*).

Ligand-free soluble state of AcPDro2 (condition A). The free energy surface of AcPDro2 under conditions in which the protein is stable in its monomeric state, condition *A*, is dominated by a single conformational well that covers a relatively broad region of free energy space and shows a substantial degree of structural variability indicative of the dynamics of the enzyme in the ligand-free state (Fig. 2*A*). The centroid of this energy well has a radius of gyration of 14.0 Å, an rmsd from the X-ray structure of 1.52 Å (calculated for all C^{α} atoms except those in loop regions), a hydrophobic accessible surface area of 5,470 $Å^2$, and a fraction of native contacts equal to 0.82 (all values reported in this section refer to the centroids of the wells/lobes). The free energy surface determined under condition *A* also contains a protruding lobe (lobe 1) consisting of conformations with a low C^{α} rmsd (1.30 Å) from the X-ray structure and a large fraction (0.92) of native contacts (additional parameters are reported in Table 1).

Aggregation-prone state (condition B). The energy surface under condition *B*, a state in which AcPDro2 readily aggregates, was again calculated by simulations restrained with NMR protection factors and reflects an ensemble of structures that is significantly more heterogeneous than that obtained from the equivalent restraints determined in the absence of TFE. The calculations

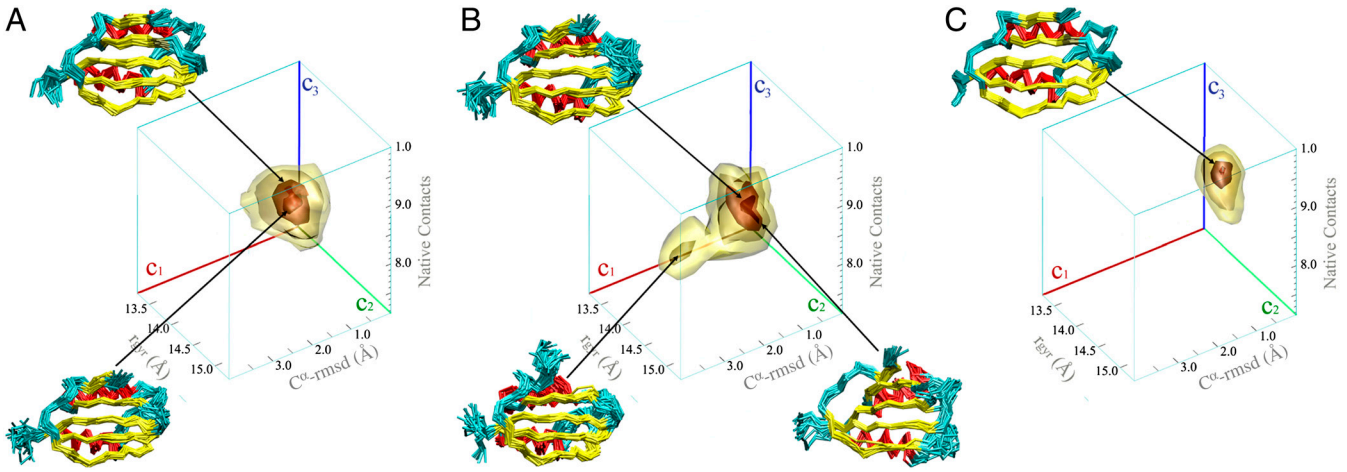


Fig. 2. AcPDro2 solution state ensembles under conditions where it is highly stable in the monomeric state (*A*, condition *A*) or readily self-assembles into amyloid fibrils (*B*, condition *B*) or bound to phosphate ions (*C*, conditions *C* and *D*). The solution ensembles were generated by means of restrained molecular dynamics simulations employing the amide exchange NMR protection factors as restraints. The conformations are projected on to three reaction coordinates to define a three-dimensional free energy landscape. The coordinates employed are as follows: C^{α} rmsd from the crystal structure (c1), radius of gyration (c2), and the fraction of native contacts (c3). The free energy surfaces are drawn by means of contour levels embedding isosurfaces of free energy at levels −7.5, −10, −18, and −24 kJ/mol. The statistical errors for these isosurfaces are 2.17, 1.1, 0.21, and 0.06 kJ/mol, respectively. Representative ensembles in the minima of conformational wells are drawn by C^{α} traces. Color codes are yellow for β-strands, cyan for loops, red for α-helices.

Table 1. Structural parameters of the free energy wells

	C^{α} rmsd,* Å	r_{gyr}, Å	Fraction of native contacts	Hydrophobic SAS, nm^2	Main-chain SAS , nm^2
Condition *A* main well	1.52	14.0	0.82	54.7	4.71
Condition *A* lobe 1	1.30	13.6	0.92	53.8	4.62
Condition *B* main well	1.57	14.2	0.83	55.0	4.73
Condition *B* lobe 1	1.62	14.5	0.76	55.3	4.72
Condition *B* lobe 2	2.75	15.0	0.72	58.6	5.12
Conditions *C* and *D*	1.08	13.7	0.96	53.2	4.56

The values refer to the centroids of the wells.
*Calculated by considering all C^{α} atoms except those located in loop regions of the protein.

did not explicitly include TFE; however, the effects of TFE are provided by the experimental data employed as restraints. The resulting surface is again characterized by a single dominant conformational well that is similar to the main well of the energy surface determined for condition *A*, although the well is generally broader and lacking structures with a very high degree of native-like content (i.e., as observed in the case of lobe 1 of condition *A*). In addition to the main well, however, the free energy surface under condition *B* shows two lobes, one of which (lobe 1) includes conformations with increased C^{α}-rmsd deviations from the crystal structure (1.62 vs. 1.57 Å) and a reduced fraction of native contacts (0.76 vs. 0.83) relative the main well (all values refer to the centroids of the wells).

The principal difference between the energy surfaces of AcPDro2 under the two conditions, however, is the existence of a second lobe under condition *B* that is not significantly accessible under condition *A* (Fig. 2). The conformations in this region have distinctive features compared to those in the other accessible regions of the surface, including a lower fraction of native contacts (0.72), a larger radius of gyration (15.0 Å), and a larger C^{α} rmsd from the crystal structure (2.75 Å). A particular distinctive trait is the more substantial exposure of regions of hydrophobic surface and main-chain atoms (512 Å^2 compared to 471 Å^2 of the main well of Ac0-TFE) as well as a lower content of secondary structure (defined by using the DSSP program; ref. 36 and Fig. S5).

Ligand-bound states (conditions C and D). The protection factors measured in phosphate buffer in the absence or presence of 5% TFE (conditions *C* and *D*, Fig. 1*C*) were employed for generating energy surfaces representative of the ligand-bound state of AcPDro2. Phosphate ions were not explicitly considered in the calculations but the effects of the phosphate binding are incorporated in the information carried in the experimental data. The resulting free energy surfaces indicate that in both cases the dynamics of the phosphate-bound state are significantly more limited than those of the corresponding ligand-free ensembles (Fig. 2). The free energy landscapes under conditions *C* and *D* (Fig. 2*C*) show a single deep conformational well with similar features to lobe 1 of the energy surface determined under condition *A*, indicating that the accessible conformations in the phosphate-bound state are very close to the crystal structure of the enzyme (C^{α} rmsd of 1.08 Å for the centroid of the basin). Moreover, in neither of the phosphate-bound structural ensembles, conformations similar to those found for lobe 2 of the energy surface of condition *B* are populated at any detectable level.

The rms fluctuation profiles indicate that the most significant reductions of the backbone dynamics associated with binding are located in the loop connecting strand S1 and helix H1, the region involved in binding the phosphate ions (Fig. S6). Relaxation experiments reveal significantly higher order parameters (S^2) for the phosphate-bound state compared to the ligand-free state (Fig. 1*D*), suggesting that the entire protein acquires a more rigid status upon binding and that such local effects in the binding site dynamics are eventually propagated throughout the protein. This result provides the detailed mechanism of the mode of actions of a small molecule that, upon binding, alters the energy landscape of the protein in a way to particularly disfavor dangerous aggregation-prone regions to be populated.

Energy Barriers for the Maintenance of Protein Solubility. The free energy profiles calculated along a single coordinate (rmsd from the crystal structure of the C^{α} atoms) provides a direct estimation of the energy barriers that have been selected during evolution to maintain protein solubility (Fig. 3*A*). Although the energy landscape determined under condition *A* shows a symmetric well around a C^{α}-rmsd value of 1.5 Å, the profile determined under condition *B* shows an additional accessible region with C^{α}-rmsd values ranging from 2.5 to 3.6 Å (Fig. 3*A*). At an rmsd value of 2.5 Å, a free energy difference between conditions *A* and *B*

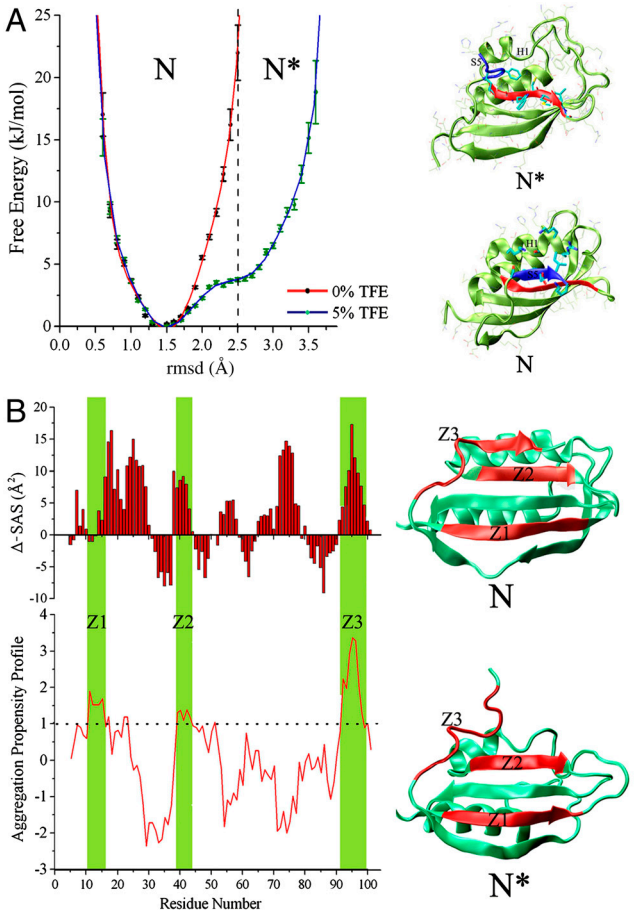


Fig. 3. Characteristics of the *N* and *N** states of AcPDro2. (*A*) Energy barriers for the aggregation-prone conformations of AcPDro2. Free energy profiles are computed as a function of the C^{α} rmsd from the crystal structure (loop residues are not considered in the calculation). Red and blue lines represent the free energy profiles determined under conditions *A* and *B*, respectively. Statistical errors in the energies are reported. The free energies have been shifted in order to present the minimum at 0 kJ/mol. *N** conformations range from rmsd of 2.5 to 3.6 Å. In the three-dimensional energy landscapes (Fig. 2), the *N* states populate the main wells and lobe 1 of the surfaces of both conditions *A* and *B*, whereas the *N** species correspond to the conformations clustered in lobe 2 of the free energy surface determined under condition *B*. (*B*) Differences in SAS area for *N* and *N** (upper histogram). The lower graph reports the Zaggregator profile for the AcPDro2 sequence. Aggregation-prone regions (Zaggregator values >1) are marked in green. These regions are designated as Z1 (residues 11–17), Z2 (residues 39–43), and Z3 (residues 92–100). The location of Z1, Z2, and Z3 are marked in red on the ribbon representations of the structures of *N* and *N** (right of the panel).

is 18.2 ± 2.4 kJ/mol (see Fig. 3*A*, dashed line), thus providing an estimate for the energy barrier that prevents the native state (condition *A*) to explore regions having higher values of rmsd, which are populated exclusively in aggregation-prone conditions (condition *B*). As regions at rmsd values higher than 2.5 Å are accessible only under condition *B*, we designate this part of the conformational ensemble as an activated state "N^*" for aggregation, whereas the conformations ranging between 1.0 and 2.5 Å of C^α-rmsd will be referred as the native state "*N*" and include the conformations with high solubility (Fig. 3*A*).

To assess the influence of the missing backbone N-H resonances in the spectra recorded under conditions *A* and *B*, we performed control calculations in which some missing protection factors were estimated by scaling the data obtained under condition *C* (see caption of Fig. S7). This analysis indicates that missing protection factors are not likely to affect the energy barrier separating the state N^* from *N* when the protein is in the native state (condition *A*). Moreover, a control simulation performed by using randomly decreased protection factors maintains energy barriers that are similar to those showed in condition *A* (Fig. S8). This finding indicates that the three-dimensional clustering of residues that become less protected in aggregation-prone conditions is a key factor for influencing the energy landscapes.

Comparison of the *N* and N^* ensembles reveals that the former possesses specific structural features that are likely to reduce the aggregation propensity. First, *N* fully preserves the structural features of a β-bulge extending from residues 85 to 87 that has been identified both generally (2) and specifically for AcP (37) as a negative designed element for reducing the aggregation propensity of edge strands, here S4. Because this element is partially disrupted in N^* (three times lower than in the *N* ensemble), strand S4 is expected to be more prone to engage in intermolecular β-sheet interactions with the edge strands of other molecules (2). Second, in N^* the interface between strands S2 and S5 is disrupted, giving rise to an increase in solvent accessible surface of 40 $Å^2$ from *N* to N^*. Hence, where *N* is able to maintain the integrity of the native edge strand S5, N^*, by disrupting the native interface of strand S5, exposes strand S2 which becomes a new, nonnative edge strand. This altered topology of the β-sheet is particularly crucial in terms of the average aggregation propensity of the protein, as strand S2 does not possess the elements of negative design for avoiding protein aggregation that are normally carried by edge strands (2). Third, although local surface charges from strand S5 and the N-terminal segment of helix H1 are exposed to the solvent in the *N* state (Fig. 3*A*), in N^*, the partial disruption of this helix and the displacement of strand S5 enhance the hydrophobicity of the local surface by increasing the exposure of residues W42, C43, M44, V52, and I100.

In order to quantify the changes in the solubility behavior of the two ensembles, we calculated the intrinsic aggregation propensity profile from the AcPDro2 sequence by using Zyggregator (17) and combined this information with the solvent accessible surface areas (SAS) of the main chains of the *N* and N^* states (Fig. 3*B*). This analysis highlights three regions of the AcPDro2 sequence with a high aggregation propensity (Zyggregator values larger than one; ref. 17); these are the fragment 11–17 (spanning strand S1), the fragment 39–43 (spanning strand S2), and the fragment 92–100 (spanning strand S5 and part of the preceding loop), and we designate these regions as Z1, Z2, and Z3, respectively. Although Z1 is completely buried in both *N* and N^* (see Fig. S9 for absolute SAS values), both Z2 and Z3 are significantly more buried in *N* than in N^*, as shown by the difference in SAS of the two states (Fig. 3*B*). This finding indicates that *N* shields the aggregation-prone regions of its sequence more effectively within its three-dimensional fold than does N^* and underlines how the energy landscapes of native states are designed to limit fluctuations leading to the exposure of such sequences by means of specific energy barriers.

Discussion

The internal motions associated with proteins are not only important for protein function (38–41) but also because they act as crucial determinants of the ability of a protein to avoid aggregation (18, 42). The present study has addressed this point in detail by showing the manner by which the native free energy landscape limits the accessibility of aggregation-prone regions of conformational space, thereby allowing the protein to maintain a soluble and functional state in solution. The analyses show how minor perturbations to the Boltzmann populations of specific conformational states in the native ensemble (here resulting from addition of 5% TFE or binding a phosphate ion) can change fundamentally the ability of the protein to remain soluble. In particular, under aggregation-prone conditions (condition *B*), the energy landscape lacks barriers for preventing an aggregation-prone state to be populated. This state is featured by the exposure of hot spots for aggregation as (nonnative) edge strands (2) or sequences predicted to have high aggregation profiles (17). We also determined at high resolution the effects of the binding of small molecules acting as inhibitors for protein aggregation, here the phosphate ion, on the local and global dynamics of the protein. We showed how such binding can significantly perturb the free energy surface and the conformational dynamics of the protein so as to make the population of aggregation-prone states energetically unfavorable, even under conditions where aggregation is otherwise favored.

The sensitivity of the energy surfaces of proteins to minor perturbations supports the view that there is a delicate balance between functionality, stability, and solubility, which is encapsulated by the concept of "life on the edge" (3, 42). Access to regions of conformational space that can readily be modulated by external factors (for example, the binding of small molecules or partner proteins) is crucial for enabling proteins to carry out their functions and also to allow complex processes such as molecular signalling or degradation to occur under cellular conditions. Such conformational access must, however, be moderated to avoid the population of any states of the protein that enable aberrant and nonfunctional interactions to occur. The manifestation of this balance is exemplified by the present finding that modifications of the correct positioning of a small fragment (strand S5) can enable the protein to maintain a fully soluble state and suppress the population of strongly aggregation-prone species (e.g., N^*).

In conclusion, the present investigation describes and illustrates at high resolution, by a method to define energy surfaces, the nature of the energy barriers that evolution has designed to allow proteins to maintain a functional dynamics without having access to dangerous aggregation-prone conformational states. The principles emerging here add to our previous understanding of how proteins avoid aggregation (2, 17, 43) and emphasize previously unidentified strategies adopted by proteins. The approach described in this study should be generally applicable to other systems, and illustrates how exploration of such surfaces can identify specific factors that serve to maintain proteins in their soluble and functional states or to avoid protein aggregation and its consequences.

Materials and Methods

NMR Assignments. Standard triple resonance experiments (HNCA, CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO spectra) as well as HNHA and ^{15}N-TOCSY-HSQC were recorded at 25 °C on a Bruker Avance 700 MHz spectrometer equipped with a cryogenic triple resonance probe (Bruker BioSpin). The chemical shifts of individual spin systems ($^1H^\alpha$, $^{13}C^\alpha$, $^{13}C^\beta$, $^{13}C'$, ^{15}N, ^1HN) were determined by using a computer-aided procedure as described previously (27). The triple resonance spectra needed for the full backbone assignment were recorded in phosphate buffer, condition *C*, because the protein is in a particularly well-defined conformational state as a result of the binding of phosphate ions and is highly resistant to aggregation to permit the use of concentrated samples. Assignment of the ^{1}H-^{15}N-HSQC spectrum under

condition *A* was achieved by following the peaks in spectra measured at different concentrations of phosphate ions, which link the ^{1}H-^{15}N-HSQC spectra under condition *A* to that under condition *C*. Assignments of spectra in 5% (vol/vol) TFE solutions were derived directly from the assignments of ^{1}H-^{15}N-HSQC spectra recorded in the absence of TFE.

Measurements of Protein Dynamics. The conformational dynamics of AcPDro2 were assessed by means of several NMR techniques covering the subnanoseconds to microseconds timescales and beyond. All the NMR experiments were conducted at 298 K and pH 4.0. Backbone dynamics were probed by means of H/D exchange and relaxation experiments.

H/D Exchange. H/D exchange was monitored by following the intensities of ^{1}H-^{15}N-HSQC signals upon exposure of the protein to D_2O. The settings employed enabled a full 2D acquisition to be recorded in approximately 2 min. The k_{obs} values were extracted by fitting the signal decay curves to single exponential functions. Data obtained under conditions promoting AcPDro2 aggregation (i.e., 5% TFE) were processed with a baseline correction from the decay curves recorded in 90% H_2O–10% D_2O. Rapidly exchanging amide protons were characterized by CLEANEX-PM (28). This technique allows an estimation of the exchange rates from the slope of the linear interpolation of the intensities of amide peaks from spectra recorded at different mixing times τ_m (5, 10, 15, 20, and 25 ms). In particular, the volumes V_i of the peaks were normalized relative to those of the corresponding ^{1}H-^{15}N-HSQC peaks V_0. By plotting V_i/V_0 as a function of τ_m, k_{obs} can be defined from the slopes of the linear interpolation.

Relaxation Experiments. Longitudinal (T_1) and transverse (T_2) ^{15}N-spin relaxation times were measured at ^{1}H frequencies of 500 and 700 MHz (^{15}N frequencies of 50.6 and 70.8 MHz, respectively) at 298 K. Further details are provided in *SI Materials and Methods*.

Structural Ensemble Refinement by H/D Exchange Data. The protection factors determined from measurements of amide exchange rates were used to define ensembles of protein structures for the various conditions studied in this work by using the data as restraints in molecular dynamics samplings (29). Further details are reported in *SI Materials and Methods*.

ACKNOWLEDGMENTS. This work was supported by grants from the Medical Research Council (C.M.D.), Engineering and Physical Sciences Research Council (A.D.S. and C.M.D.), European Molecular Biology Organization (A.D.S.), Marie Curie (A.D.S.), Human Science Frontier Program (S.T.H.), Italian Ministero dell'Istruzione dell'Università e della Ricerca (F.C.), Boehringer Ingelheim Foundation and Murray Edwards College, Cambridge (A.D.), National Science Council of the Republic of China–Taiwan (S.T.H.), Biotechnology and Biological Sciences Research Council (C.M.D. and M.V.), and the Royal Society (M.V.).

1. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424:805–808.
2. Richardson JS, Richardson DC (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 99:2754–2759.
3. Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M (2007) Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* 32:204–206.
4. Lopez de la Paz M, Serrano L (2004) Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci USA* 101:87–92.
5. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
6. Chatani E, Goto Y (2005) Structural stability of amyloid fibrils of beta(2)-microglobulin in comparison with its native fold. *Biochim Biophys Acta* 1753:64–75.
7. Knowles TP, et al. (2007) Kinetics and thermodynamics of amyloid formation from direct measurements of fluctuations in fibril mass. *Proc Natl Acad Sci USA* 104:10016–10021.
8. Baldwin AJ, et al. (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* 133:14160–14163.
9. Gianni S, et al. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* 17:1431–1437.
10. Dobson CM (2003) Protein folding and misfolding. *Nature* 426:884–890.
11. Douglas PM, Cyr DM (2010) Interplay between protein homeostasis networks in protein aggregation and proteotoxicity. *Biopolymers* 93:229–236.
12. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475:324–332.
13. Liberek K, Lewandowska A, Zietkiewicz S (2008) Chaperones in control of protein disaggregation. *EMBO J* 27:328–335.
14. Behrends C, et al. (2006) Chaperonin TRiC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Mol Cell* 23:887–897.
15. Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. *Cell* 125:443–451.
16. Molinari M (2007) N-glycan structure dictates extension of protein folding or onset of disposal. *Nat Chem Biol* 3:313–320.
17. Tartaglia GG, et al. (2008) Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* 380:425–436.
18. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5:15–22.
19. Soldi G, et al. (2005) Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state. *Biophys J* 89:4234–4244.
20. Zuccotti S, et al. (2004) Three-dimensional structural characterization of a novel Drosophila melanogaster acylphosphatase. *Acta Crystallogr D Biol Crystallogr* 60:1177–1179.
21. 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.
22. Clarke J, Itzhaki LS (1998) Hydrogen exchange and protein folding. *Curr Opin Struct Biol* 8:112–118.
23. Englander SW (2000) Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* 29:213–238.
24. Veglia G, Zeri AC, MA C, Opella SJ (2002) Deuterium/hydrogen exchange factors measured by solution nuclear magnetic resonance spectroscopy as indicators of the structure and topology of membrane proteins. *Biophys J* 82:2176–2183.
25. Hernandez G, Anderson JS, LeMaster DM (2010) Assessing the native state conformational distribution of ubiquitin by peptide acidity. *Biophys Chem* 153:70–82.
26. Soldi G, Plakoutsi G, Taddei N, Chiti F (2006) Stabilization of a native protein mediated by ligand binding inhibits amyloid formation independently of the aggregation pathway. *J Med Chem* 49:6057–6064.
27. Fusco G, et al. (2011) ^{1}H, ^{13}C and ^{15}N resonance assignments of human muscle acylphosphatase. *Biomol NMR Assign*, 10.1007/s12104-011-9318-1.
28. Hwang TL, van Zijl PC, Mori S (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J Biomol NMR* 11:221–226.
29. 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.
30. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14.
31. 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.
32. Best RB, Vendruscolo M (2004) Determination of protein structures consistent with NMR order parameters. *J Am Chem Soc* 126:8090–8091.
33. Vendruscolo M, Paci E, Dobson CM, Karplus M (2003) Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J Am Chem Soc* 125:15686–15687.
34. Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci USA* 104:11257–11262.
35. Kern G, Handel T, Marqusee S (1998) Characterization of a folding intermediate from HIV-1 ribonuclease H. *Protein Sci* 7:2164–2174.
36. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
37. Soldi G, Bemporad F, Chiti F (2008) The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. *J Am Chem Soc* 130:4295–4302.
38. Mittermaier A, Kay L (2009) Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* 34:601–611.
39. Tzeng SR, Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. *Nature* 462:368–372.
40. Gao YQ, Yang W, Karplus M (2005) A structure-based model for the synthesis and hydrolysis of ATP by F1-ATPase. *Cell* 123:195–205.
41. Freiburger LA, et al. (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. *Nat Struct Mol Biol* 18:288–294.
42. Dobson CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* 24:329–332.
43. Monsellier E, Chiti F (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8:737–742.

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.

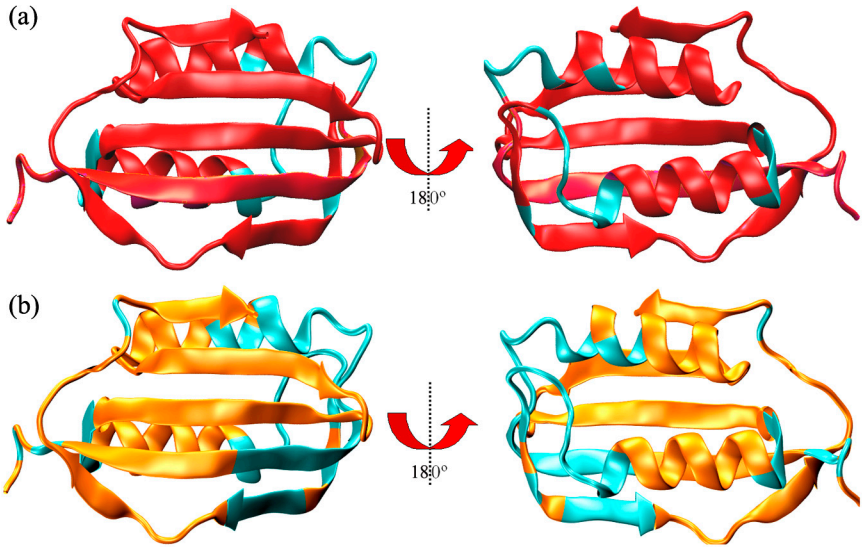


Fig. S1. AcPDro2 assignment. (*A*) Phosphate-bound state (condition *C*). Residues with assigned resonances are shown as red ribbons. Residues when resonances are missing in the spectra or overlapping in the heteronuclear single quantum coherence (HSQC) spectra are marked by cyan ribbons. (*B*) Ligand-free 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).

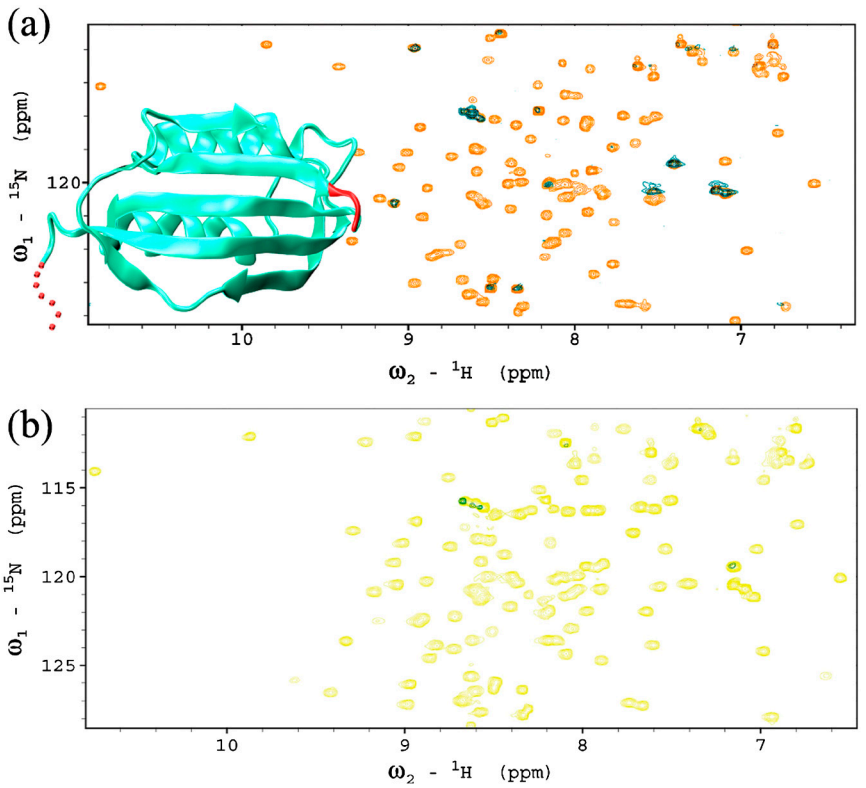


Fig. S2. Phase-modulated clean chemical exchange (CLEANEX-PM) spectra of AcPDro2. (*A*) Protein in acetate buffer, ligand-free state (condition *A*). ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) (orange) and CLEANEX spectra (green peaks). Backbone amides showing CLEANEX peaks are drawn 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.

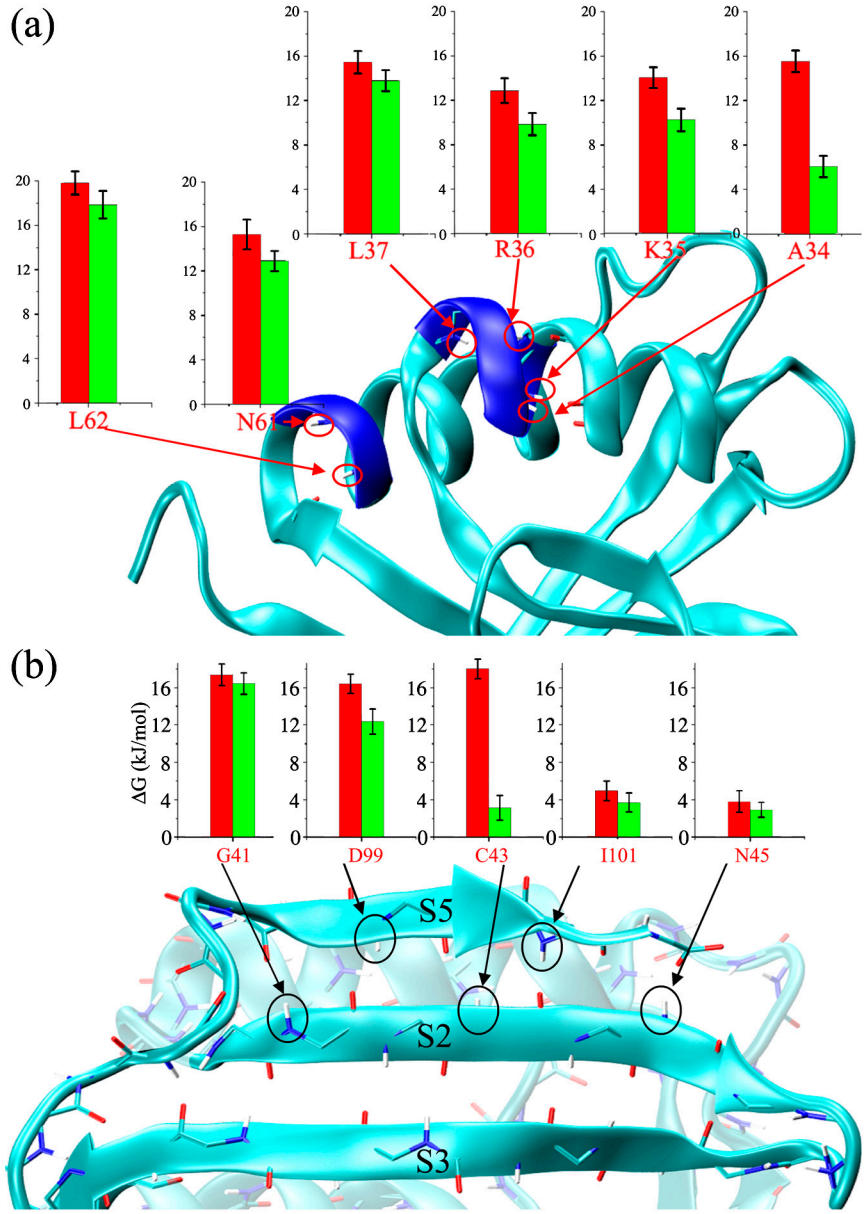
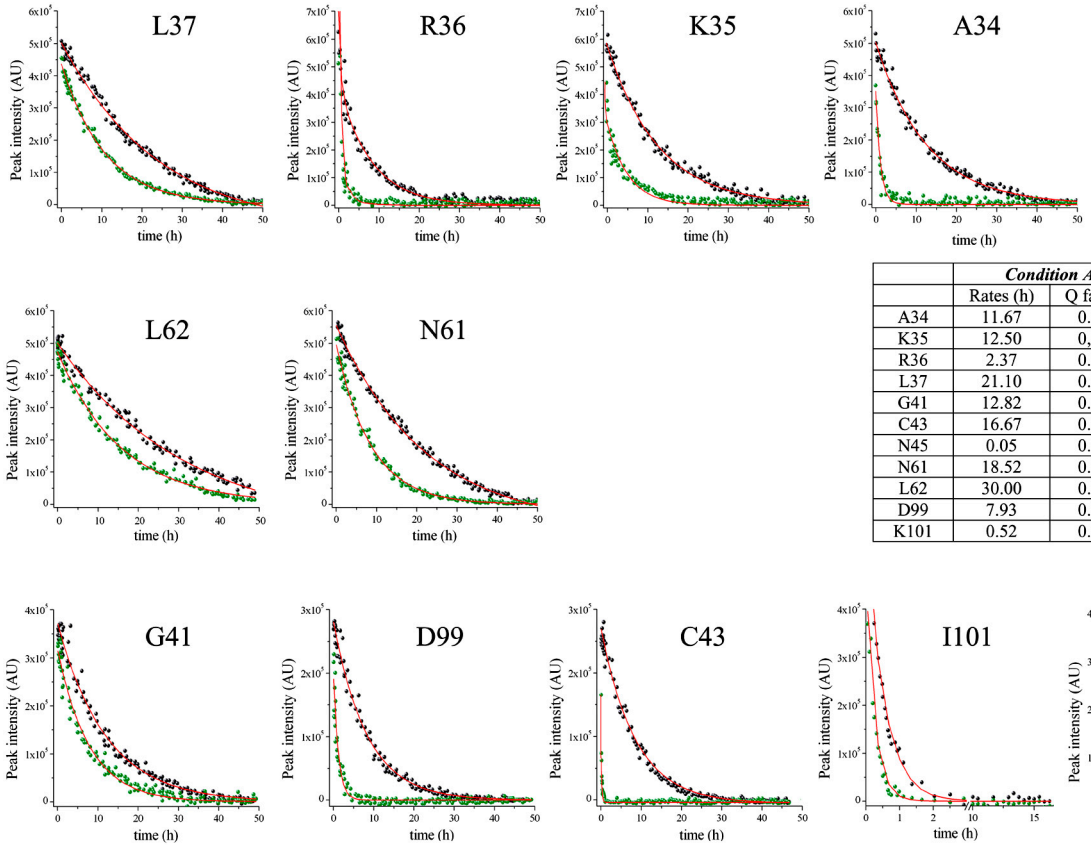


Fig. S3. Structural clusters showing the reduction of the protection factors in condition *A* (red bars) and condition *B* (green bars). (*A*) Helices H1 and H2. The protein is represented by cyan ribbons, whereas blue marks highlight regions showing major drops in protection factors. (*B*) Interface 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.

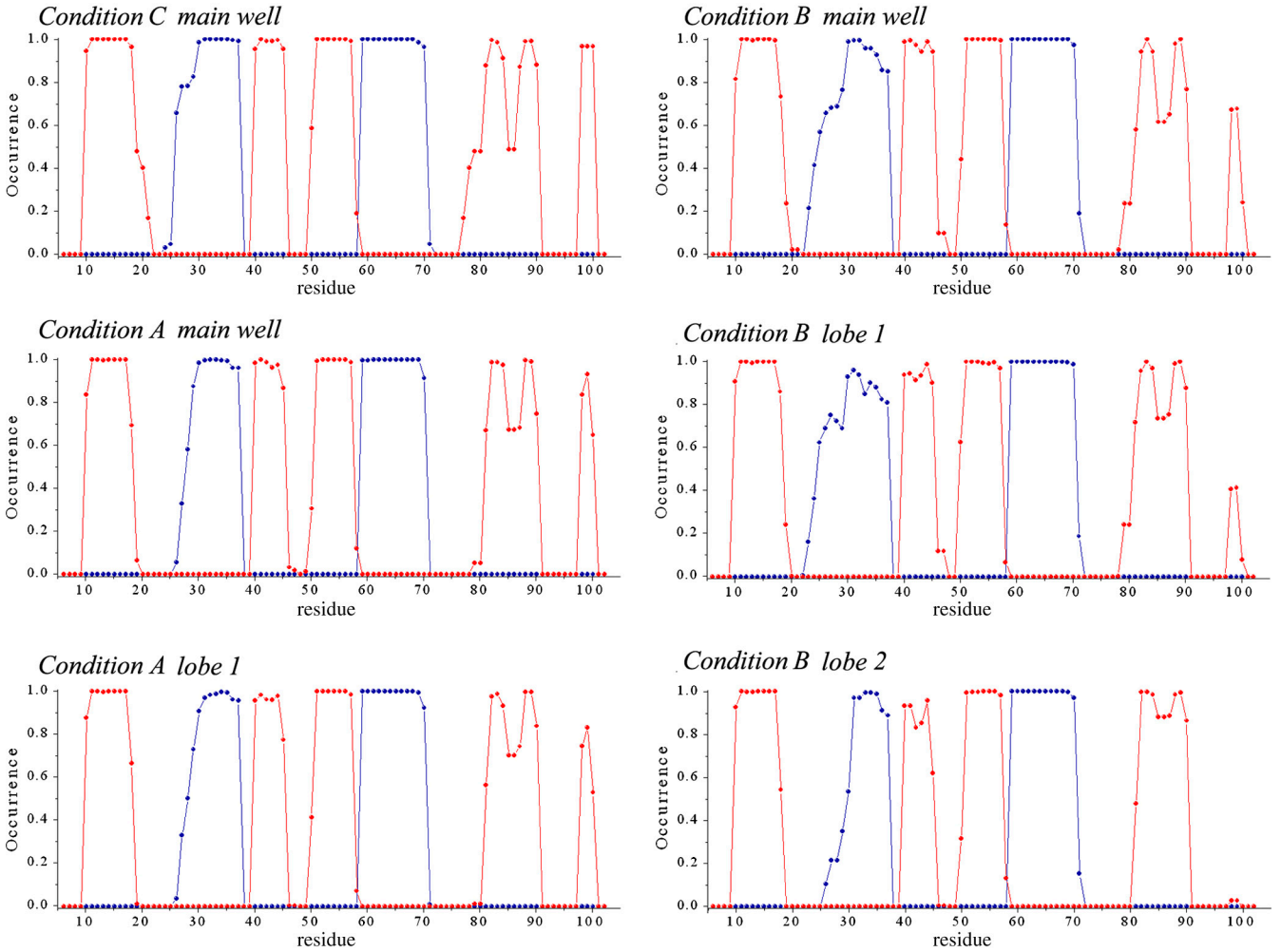


Fig. S5. Occurrence of secondary structure in the conformational wells of the different solution states of the protein. The populations of β-strands and α-helix are shown by red and blue lines, respectively.

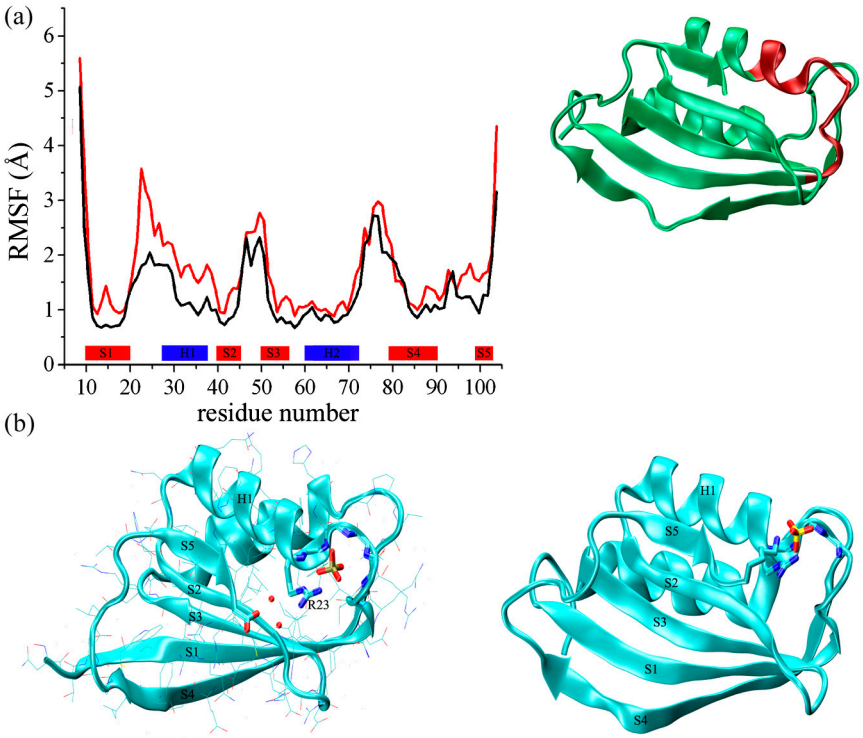


Fig. S6. (*A*) Root-mean-square fluctuations (RMSF) calculated for N and H atoms for the ligand-free and phosphate-bound states at 0% trifluoroethanol (conditions *A* and *C*). The region showing the largest differences in the fluctuations of the two states is highlighted in red on the structure. (*B*)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

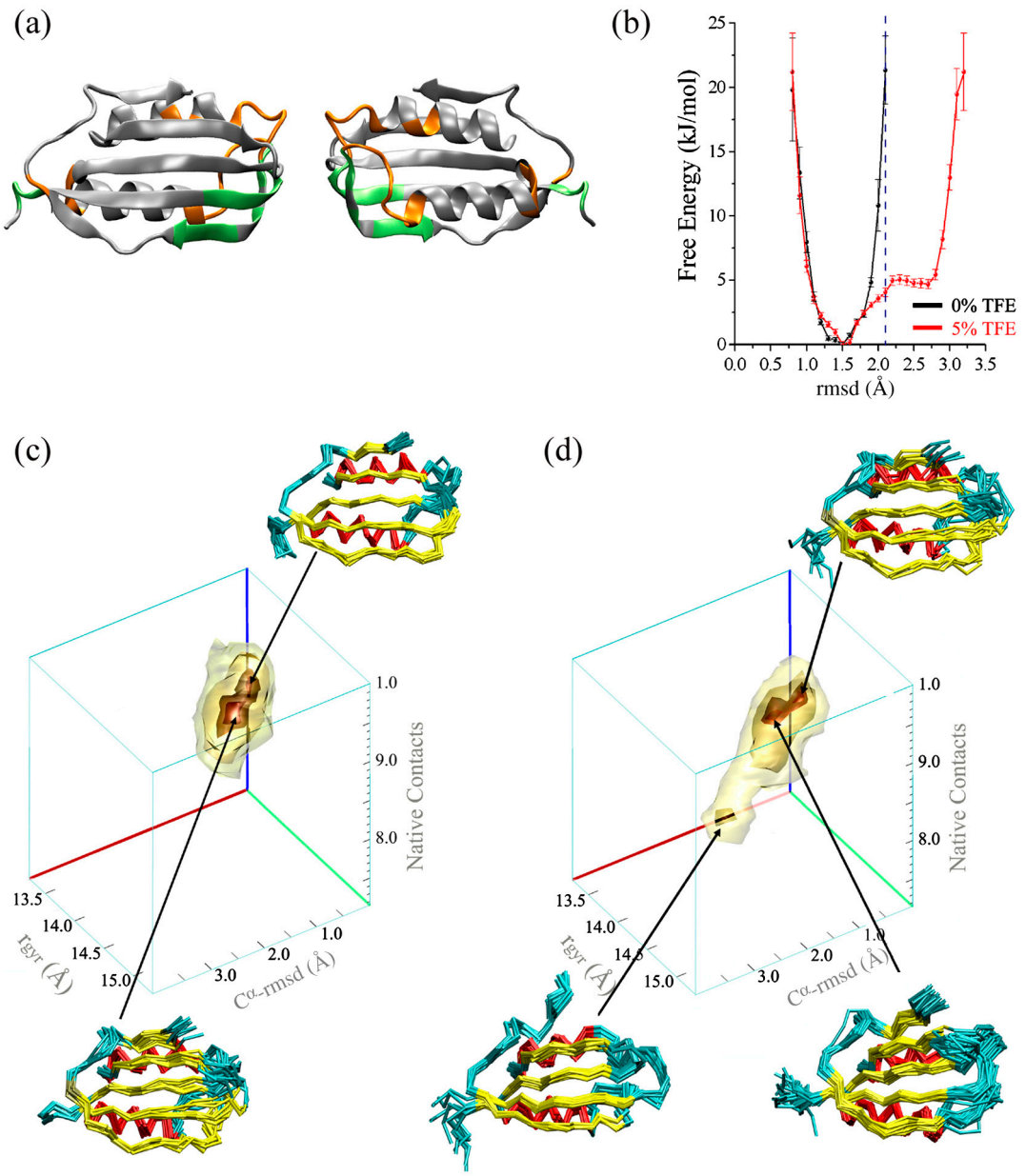


Fig. S7. Control simulation of conditions *A* and *B*. (*A*) The AcPDro2 structure. Residues for which experimental protection factors were measured in both 0% and 5% trifluoroethanol (TFE; vol/vol) are shown in gray. Residues for which the protection factors have been estimated from the data measured 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.

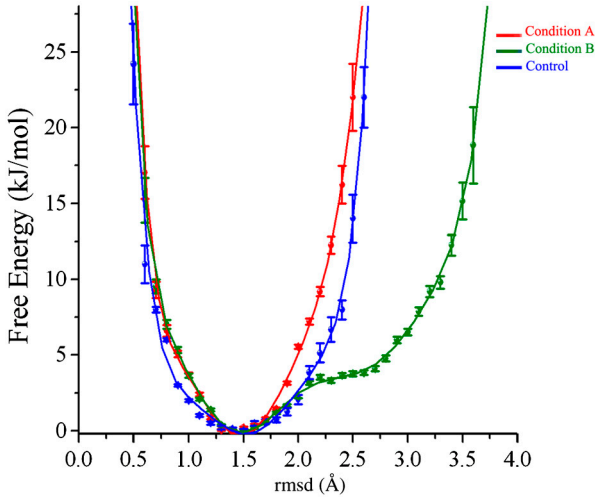


Fig. S8. One-dimensional energy landscape on a control simulation. In this control, the protection factors for condition *A* (0% trifluoroethanol, TFE) are randomly altered to produce a similar amount of deprotection as measured for condition *B* (5% TFE). Red and green lines report the unperturbed free 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*.

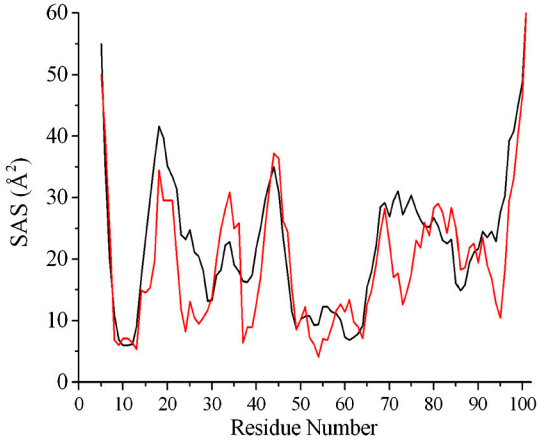


Fig. S9. Main-chain solvent accessible surface (SAS) area of the *N* (red) and *N** (black) states of AcPDro2.