SCIENTIFIC REPORTS

Towards a structural biology of the hydrophobic effect in protein folding

Carlo Camilloni[1], Daniela Bonetti[2], Angela Morrone[2], Rajanish Giri[2,†], Christopher M. Dobson[1], Maurizio Brunori[2], Stefano Gianni[2] & Michele Vendruscolo[1]

The hydrophobic effect is a major driving force in protein folding. A complete understanding of this effect requires the description of the conformational states of water and protein molecules at different temperatures. Towards this goal, we characterise the cold and hot denatured states of a protein by modelling NMR chemical shifts using restrained molecular dynamics simulations. A detailed analysis of the resulting structures reveals that water molecules in the bulk and at the protein interface form on average the same number of hydrogen bonds. Thus, even if proteins are 'large' particles (in terms of the hydrophobic effect, i.e. larger than 1 nm), because of the presence of complex surface patterns of polar and non-polar residues their behaviour can be compared to that of 'small' particles (i.e. smaller than 1 nm). We thus find that the hot denatured state is more compact and richer in secondary structure than the cold denatured state, since water at lower temperatures can form more hydrogen bonds than at high temperatures. Then, using Φ-value analysis we show that the structural differences between the hot and cold denatured states result in two alternative folding mechanisms. These findings thus illustrate how the analysis of water-protein hydrogen bonds can reveal the molecular origins of protein behaviours associated with the hydrophobic effect.

A detailed understanding of the molecular origins of the hydrophobic effect[1–10] in proteins and of its role as a driving force in protein folding and assembly[10–14] is still an open problem. This situation arises because protein sequences include a complex combination of hydrophobic (non-polar) and hydrophilic (polar) regions. The resulting patterns of non-polar regions have typical dimensions that correspond to a critical value for the hydrophobic effect[1–8]. This value, which is of about 1 nm, corresponds to the distinction between small and large non-polar solutes[1–9]. To understand the molecular origin of this characteristic length, one should start from the observation that water molecules tend to form networks of hydrogen bonds at the expense of their rotational entropy. Non-polar solutes smaller than 1 nm occupy a volume that is too small to perturb significantly the formation of such hydrogen bond networks. By contrast water molecules near the surfaces of non-polar solutes larger than 1 nm cannot form all the hydrogen bonds that they would do in bulk. These two situations correspond to a different scaling of the hydrophobic forces, with the volume or the surface, for small or large ideal hydrophobic solutes, respectively[1].

Here our aim is to clarify first whether proteins, whose surfaces as mentioned above are characterized by the presence of complex polar and non-polar patterns, behave effectively like small or large non-polar solutes, and then to investigate the consequences of this fact on their folding behaviour[11–13]. Addressing this question requires an accurate characterization of the conformational space of a protein in water to study the number of hydrogen bonds formed by water molecules in proximity of its surface with respect to the bulk.

To attack this problem, we exploited the opportunities offered by the case of yeast frataxin (Fig. 1a). Frataxin, which is a protein involved in the assembly of iron-sulfur clusters, is also related to Friedreich's ataxia, a fatal neurodegenerative condition[15]. We considered yeast frataxin because this protein represents one of the few examples for which both cold and hot denatured states have been observed at neutral pH and without addition of destabilizing agents (at 272 K and 323 K, respectively), and characterized by nuclear magnetic resonance (NMR)

[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK. [2]Istituto Pasteur – Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" Università di Roma "La Sapienza", 00185 Rome, Italy. [†]Present Address: School of Basic Sciences, Indian Institute of Technology Mandi, Mandi, HP, India, 175005. Correspondence and requests for materials should be addressed to S.G. (email: stefano.gianni@uniroma1.it) or M.V. (email: mv245@cam.ac.uk)
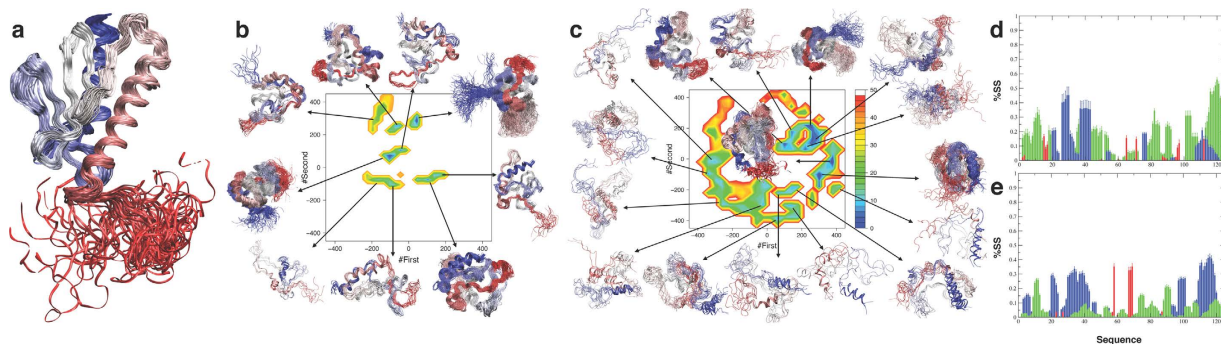
Figure 1. The cold and hot denatured states of frataxin are structurally different. (**a**) Structure of the native state of frataxin. (**b**) Free energy surface of the cold denatured state (CDS) determined at 272 K as a function of the sketch-map[63] collective variables that describe the conformational features of the ensembles (see Methods). Nine microstates are shown representing the local and global minima. These microstates comprise >90% of the total population. (**c**) Free energy surface of the hot denatured state (HDS) determined at 323 K as a function of the same two collective variables. Sixteen microstates are shown representing the local and global minima. These microstates comprise >90% of the total population (see Methods). (**d,e**) Secondary structure populations of the CDS (**d**) and HDS (**e**). α-Helices are represented in blue, β-strands in red and polyproline II in green; free energies are given in kJ/mol.

chemical shifts and circular dichroism (CD)[15–18]. As shown by the pioneering work of Privalov, the thermodynamic analysis of cold denaturation, compared to that of the thermal denaturation, offers unique advantages for understanding the molecular determinants of hydrophobicity[19,20]. Indeed, it is generally recognised that while thermal denaturation is the consequence of a temperature-induced increase in conformational fluctuations, cold denaturation is a consequence of an enthalpy gain of the solvent[19,20]. Not many atomistic details are known, however, about the structural consequences of this gain of entropy or enthalpy in the hot and cold denatured states, respectively. Thus, modelling the cold and hot denatured states of yeast frataxin in the absence of any additional agent should highlight the different roles played by the protein and solvent at different temperatures.

In the following we employ the chemical shifts measured in cold and hot denaturation conditions[15–17] together with replica-averaged metadynamics (RAM) simulations[21,22] (see Materials and Methods) to elucidate at atomic resolution the structure and the dynamics of the cold denatured state (CDS) and hot denatured state (HDS). In the RAM simulations the experimental information provided by NMR chemical shifts is incorporated in terms of structural restraints[22] and at the same time the sampling of the conformational space is enhanced by metadynamics[23]. This overall approach makes it possible to simultaneously change the force field used in the molecular dynamics simulations to improve their agreement with the experimental data in the spirit of the maximum entropy principle[24] and to significantly decrease the computational resources required to obtain convergence in the sampling. Then, by using Φ-values analysis[25–27] and Φ-values restrained molecular simulations[26], we determined their cold transition state (CTS) and hot transition state (HTS) to describe the differences in the cold and hot denaturation processes corresponding to the differences between the cold and hot denatured states.

Results and Discussion

Determination of the hot and cold denatured state ensembles. From the analysis of the RAM simulations at convergence (Fig. S1) we conclude that the two denatured states exhibit qualitatively different behaviours, as the regions of conformational space that they sample are essentially non-overlapping (Fig. 1b,c). None of the most populated microstates in the CDS is equally populated in the HDS and vice versa. Furthermore, the conformational space of the CDS is significantly smaller than that of HDS, with less local minima being populated in the CDS than in the HDS (Fig. 1b,c). In the CDS, 9 minima cover the 90% of the conformational space while 16 much broader minima are needed in the HDS to cover the same extent. Structurally the CDS is more expanded than the HDS. Furthermore, the α-helical content is 6% in the CDS and 10% in the HDS (Fig. 1d), and the β-sheet content is 0.7% in the CDS and 1.4% in the HTS (Fig. 1e). In the CDS there are three short partially populated α-helices in the region of residues 20–45 and a longer, less populated one, in the region of residues 107–120 in agreement with a previous observation about the presence of residual α-helical structure at the N-terminus[16]. In the HDS instead, there is a larger α-helical population distributed over a broader range of residues with two α-helices in the regions of residues 25–48 and 110–120. A remarkable difference between the two ensembles is also observed in the polyproline II content, which is 15% in the CDS and 5% in the HDS. The two ensembles are also different in terms of their radii of gyration (R_g), with 1.7 nm in the CDS and 1.6 nm in the HDS; for reference, this value is 1.5 nm in the native state (NS), even if this difference is may be underestimated because of known limitations of current force fields[28]. The higher secondary structure content and the lower R_g of the HDS correspond to a slightly higher resemblance of the NS to the HDS with respect to the CDS. The average fraction of native contacts[29], Q, is 0.18 and 0.22 for the CDS and the HDS, respectively (Fig. S2). Overall, the HDS is richer in secondary structure and more compact than the CDS. These observations agree and extend previous comparisons of the two denatured states based on single-molecule FRET, NMR and SAXS measurements[16,30].
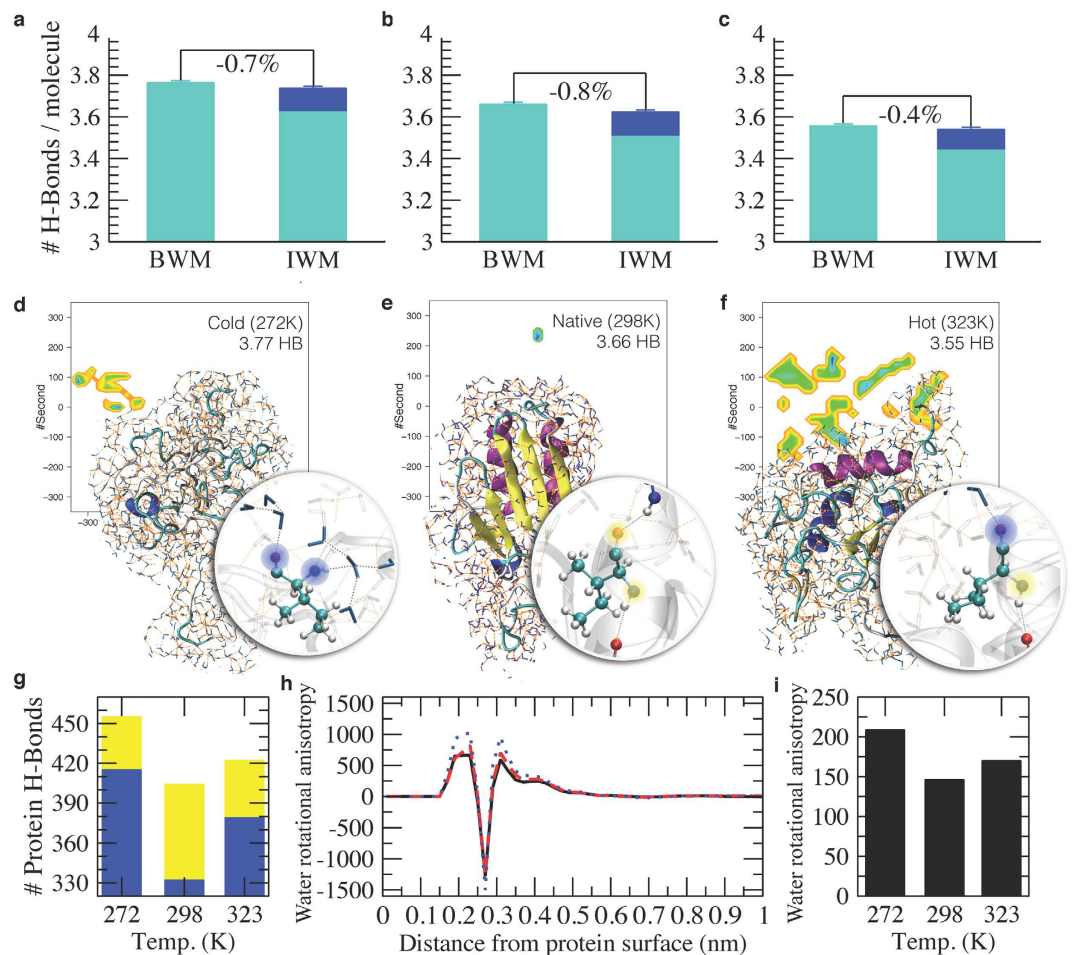
Figure 2. Structural characterization of the hydrophobic effect in protein folding. (**a–c**) Number of hydrogen bonds per water molecule for the CDS (272 K), NS (298 K) and HDS (323 K). The number of hydrogen bonds is calculated for bulk water molecules (BWM, farther than 0.4 nm), and for interface water molecules (IWM, water molecules in the first shell around the protein, within 0.4 nm) considering both the water-water hydrogen bonds (light blue) and the water-protein hydrogen bonds (blue). In our calculations, we found that bulk and interface water molecules form the same number of hydrogen bonds at the three temperatures. (**d–f**) At cold denaturation temperatures water molecules win the competition with the protein to form hydrogen bonds with the protein hydrogen bond donors and acceptors. Two hydrogen bonds formed by a hydrophobic residue are shown in particular, for which only the backbone amide and carbonyl groups can form hydrogen bonds; these two water-protein hydrogen bonds are highlighted in blue. By contrast, under folding conditions intra-protein hydrogen bonds are preferentially formed, and no particular preference is present at hot denaturation temperatures; the two protein-protein hydrogen bonds (within a α-helix) are highlighted in yellow. (**g**) Number of hydrogen bonds formed by frataxin in the CDS, NS and HDS ensembles. The number of protein-protein hydrogen bonds is shown in yellow, and that of protein-water hydrogen bonds in blue. (**h**) Water rotational anisotropy as a function of the distance from the protein surface for the CDS (dotted blue), NS (solid black) and HDS (dashed red), calculated as $<3\cos^2\vartheta(r) -1> \text{RDF}(r)$, where ϑ is the angle of the vector from a solvent oxygen to a protein atom with the normal of the solvent plane averaged over all the solvent molecules and all the protein atoms, and RDF(r) is the non-normalised radial distribution function of the water with respect to the protein surface. (**i**) Total water rotational anisotropy in presence of the protein calculated as the integral of the absolute value of the former curves. The degree of order of the water molecules near the protein correspond to the extent to which they form hydrogen bonds with the protein, which is large at cold denaturing conditions, small under folding conditions, and intermediate at hot denaturing conditions.

Structural characterization of the hydrophobic effect in protein folding.

The different structures of the CDS, NS and HDS correspond to differences in the overall properties of the protein-water system (Fig. 2). To investigate these differences, we first calculated the number of hydrogen bonds in the three states. In increasing the temperature, bulk water molecules (BWM) form decreasing numbers of hydrogen bonds[31–33], in the current case by going from 3.77 at 272 K, to 3.66 at 298 K, to 3.55 at 323 K (Figs 2a–c and 3 and Fig. S3). Remarkably, these numbers are rather well preserved also for water molecules in the first shell around the protein (interface water molecules, IWM, Fig. 2a–c). In this interface region the number of water-water hydrogen bonds per molecule

3

	Cold denatured state 272K	Native state 298K	Hot denatured state 323K
Water	**Energy gain** *3.77 H-bonds/molecule*	**Reference energy** *3.66 H-bonds/molecule*	**Energy loss** *3.55 H-bonds/molecule*
	Entropy loss *-4.54 kJ/mol K*	**Reference entropy**	**Entropy gain** *+3.67 kJ/mol K*
Protein-water interface	**Energy gain** *-2900 kJ/mol*	**Reference energy**	**Energy loss** *+137 kJ/mol*
	Rotational Entropy loss *1940 water molecules*	**Rotational Entropy loss** *1593 water molecules*	**Rotational Entropy loss** *1826 water molecules*
Protein	**Energy loss** *+919 kJ/mol*	**Reference energy**	**Energy loss** *+481 kJ/mol*
	Entropy gain *9 main microstates*	**Reference entropy** *1 main microstate*	**Entropy gain** *16 main microstates*

Figure 3. Summary of the entropic and energetic contributions to the cold and hot denatured states.

drops (Fig. 2a–c, light blue), but the total number of hydrogen bonds per molecule is recovered within less than 1% when one takes into account also the protein-water hydrogen bonds (Fig. 2a–c, blue). In essence, while the overall number of hydrogen bonds formed by a water molecule is relatively constant for water molecules in the bulk and at the interface[10], the protein itself respond to the change in this number with temperature by redistributing its protein-protein (Fig. 2d–f, yellow) and protein-water (Fig. 2d–f, blue) hydrogen bonds, for example by populating polyproline II conformations in the CDS in place of α-helical conformations in the HDS (Fig. 1, in particular in the C-terminal α-helix). In decreasing the temperature from 298 K to 272 K the interface water molecules in the CDS should ideally gain a total of about 94 hydrogen bonds to raise their average value per molecule from 3.66 to 3.77. Our results indicate that water forms 83 such hydrogen bonds with the protein, thus stabilising its CDS (Fig. 2g).

We obtained additional insight into the factors that determine the cold and hot denaturation processes by calculating the van der Waals and the Coulomb energies of the protein-water system (Fig. S4). While the protein-protein energy is minimized in the folded state, the protein-water energy is strengthened under cold denaturation conditions, and is slightly weakened in hot denaturation conditions with respect to the NS. These results suggest that the CDS is stabilized by interactions with the solvent and as a consequence the protein is more expanded. In the NS the balance of the interactions is moved towards the protein-protein interactions and indeed the protein is structured. Finally the HDS exhibits an intermediate behaviour where the protein conserves some residual structure and remains more compact.

A qualitative analysis of the role entropy is obtained by analysing the average orientations of water molecules with respect to the protein as a function of their distances from the protein surface in the CDS, the HDS and the NS (Fig. 2h). A value of zero indicates that on average a water molecule at a given distance does not have any preferred orientation with respect to the protein and so its rotational entropy is the same as in bulk water. Positive and negative values report about an anisotropic distribution of the solvent with respect to the protein and correspond to an entropy loss with respect to the bulk; these curves take also into account the number of water molecules at a given distance (Fig. 2). Up to 0.7 nm from the protein water molecules show an anisotropic distribution with respect to the protein, and a consequent entropy loss with respect to the bulk. This distance corresponds to the first two shells of water around the protein (Fig. S5). In the denatured states this behaviour is slightly more pronounced than in the NS, due to the more exposed surfaces and as a consequence the larger number of water molecules that are oriented. The integral of the absolute value of the curves of Fig. 2h reports on the overall entropy loss with respect to bulk water (Fig. 2i). The NS shows the least water entropy loss, indicating that in the native state the hydrophobic forces, here interpreted essentially in terms of loss of rotational freedom, are minimized.

The structural analysis described above characterizes the entropic and energetic contributions to the folding process of frataxin as a function of the temperature (Fig. 3). In decreasing the temperature from 298 K to 272 K, water gains in energy by increasing by about 3% the number hydrogen bonds, and correspondingly loses in entropy because of the increase in order. At the same time, the protein unfolds by gaining energy by interacting more strongly with the solvent at the expense of protein-protein interactions (Fig. 2b,c), and gaining some entropy by accessing a slightly larger conformational space (Fig. 1a). Water molecules in the first two shells around the protein lose in entropy with respect to the bulk (−12%). By contrast, in increasing the temperature from 298 K to 323 K water loses energy by decreasing by about 3% the number hydrogen bonds and gains entropy because of the decrease in order. In parallel, the protein unfolds, losing energy both due to loss of protein-protein and protein-water interactions and gaining entropy by accessing a larger conformational space (Fig. 1b,c). Water molecules in the first two shells around the protein lose in entropy with respect to the bulk (−11%).

It is also remarkable that both at low and at high temperatures the number of hydrogen bonds per surface water molecule does not depend on the free energy of the conformation in the ensemble (Fig. S6), although the number of hydrogen bonds depend on the identity of the amino acid and to the degree of solvent exposure (Figs S7–S9). This result is particularly important because it shows that the number of number of hydrogen bonds per surface water molecule is a property robust with respect to possible inaccuracies in the simulations, as for example a slightly overestimated compactness of the protein. While these results are consistent with previous observations about the relevance of protein-water interactions in driving cold denaturation[9,10,16,34,35], here we show explicitly that the number of hydrogen bonds is the same for bulk and interface water molecules at all the relevant temperatures. This phenomenon could determine the properties of the folded and unfolded states, as also recently suggested using a two-dimensional model[36].

Frataxin folds following different pathways from the cold and hot denatured states.

The differences observed for the residual structure of the two denatured state ensembles together with the differences observed in the protein-water interactions offer the opportunity to investigate the consequences of the denaturing conditions on the protein folding process. To this end, we carried out a Φ-value analysis[25,27] in combination with restrained molecular dynamics simulations[26] to determine the structures of the cold and hot transition states (CTS and the HTS, see Supporting Information). Here Φ values are structurally interpreted as the fraction of native contacts formed by a residue in the transition state and these fractions are thus used as restraints in a simulation to find an ensemble of conformations in agreement with all the Φ values[26]. The Φ values (Table S1 and Fig. S10) show that the hot and cold denaturation pathways are characterized by two structurally different transition states, which provide a structural perspective on the folding reaction.

A comparison of the contact maps of the CTS and HTS and those of the CDS, HDS and NS (Fig. 4a,b) reveals that in the CTS almost all the native secondary structures are already in place, whereas the HTS is more disordered. Also the tertiary contacts are more native-like in the CTS than in the HTS, in particular in the regions corresponding to the docking of the α-helices with the six-stranded β-sheet, as observed in urea[17]. Also, the fraction of native contacts Q is 0.65 and 0.49 for the CTS and HTS, respectively (see also Fig. S2).

A comparison between the structures of CDS, HDS, CTS and HTS allows extracting a structural depiction of the so-called anti-Hammond effect, which is sometimes observed in protein folding[37]. The anti-Hammond effect occurs when the transition state moves towards the product as the reactant becomes more stable. In case of frataxin, while the CDS is less structured than the HDS, the CTS and HTS display opposite behaviours, with the HTS being less structured than the CTS. To provide a structural rationalisation of this behaviour, it may be considered that, because the CDS is very heterogeneous, the free energy well associated with it is shallower than that of the HDS as a function of Q (Fig. S11). Thus, the intersection between the free energy wells of the native and denatured states, which identifies the transition state of the reaction, may shift towards the native conformation, resulting in an apparent anti-Hammond effect.

The solution structures of the CDS, HDS, CTS and HTS reveal important features of the cold and hot folding mechanisms. The increase of residual structure in the denatured state from the CDS to the HTS shifts the folding mechanisms from nucleation-condensation[38] (where the protein folds as a single cooperative unit) to diffusion-collision[39] (where the protein folds in contiguous blocks). A signature of these different behaviours lies in the so called Brønsted plot[37], which relates the activation free energy of a reaction with the corresponding equilibrium free energy and is linear in the nucleation-condensation scenario, while is scattered in diffusion collision[37]. The analysis of the Brønsted plot of the CTS and HTS (Fig. 4c) provides a direct test of the effect of the residual structure in the denatured states on the folding mechanisms. At low temperature, with a weakly structured CDS, there is a strong correlation (0.77) between $\Delta\Delta G_{D-N}$ and $\Delta\Delta G_{TS-N}$, while at high temperature the correlation is weaker (0.48) (Fig. 4c). These results provide a demonstration of the correlation between the fraction of native contacts in the transition state, Q, and the coefficient of correlation between $\Delta\Delta G_{D-N}$ and $\Delta\Delta G_{TS-N}$ (Fig. 4c, insets), and reveal its links with the hydrogen bonding behaviour of water at low and high temperatures and thus with the hydrophobic effect.

Conclusions

The results that we have presented provide an atomistic representation of the seminal observations by Privalov about the favourable hydration of protein upon cold denaturation[19,20]. We have shown how a major role in determining the state of yeast frataxin at different temperatures is played by the competition between water and protein molecules for forming hydrogen bonds with the protein hydrogen bonding donors and acceptors. Our results indicate that this protein interacts differently with water molecules at low and high temperatures, resulting in a more expanded and less structured denatured state at low temperatures, and in a more compact and structured denatured state at high temperatures.

These differences also determine the specific changes in the cold and hot folding pathways of this protein, and they suggest the possible origins of complex effects such as the anti-Hammond behaviour and the Brønsted plot in protein folding by linking them with the number of hydrogen bonds that water molecules tend to form. Upon increasing the temperature from cold to native conditions, as the number of protein-water hydrogen bonds decreases, the protein can form stable secondary and tertiary structures. Then, when the temperature is raised from native to hot denaturing conditions, thermal fluctuations destabilise the native fold, but at the same time a further decrease in the number of protein-water hydrogen bonds results in the conservation of many of secondary structure elements and in a relatively compact denatured state.

These observations may also explain the finding of an increase in compactness and secondary structure content for denatured and for intrinsically disordered proteins upon increasing temperatures[30,40,41]. Since intrinsically disordered proteins are richer in polar groups than globular proteins, by increasing the temperature, the loss of
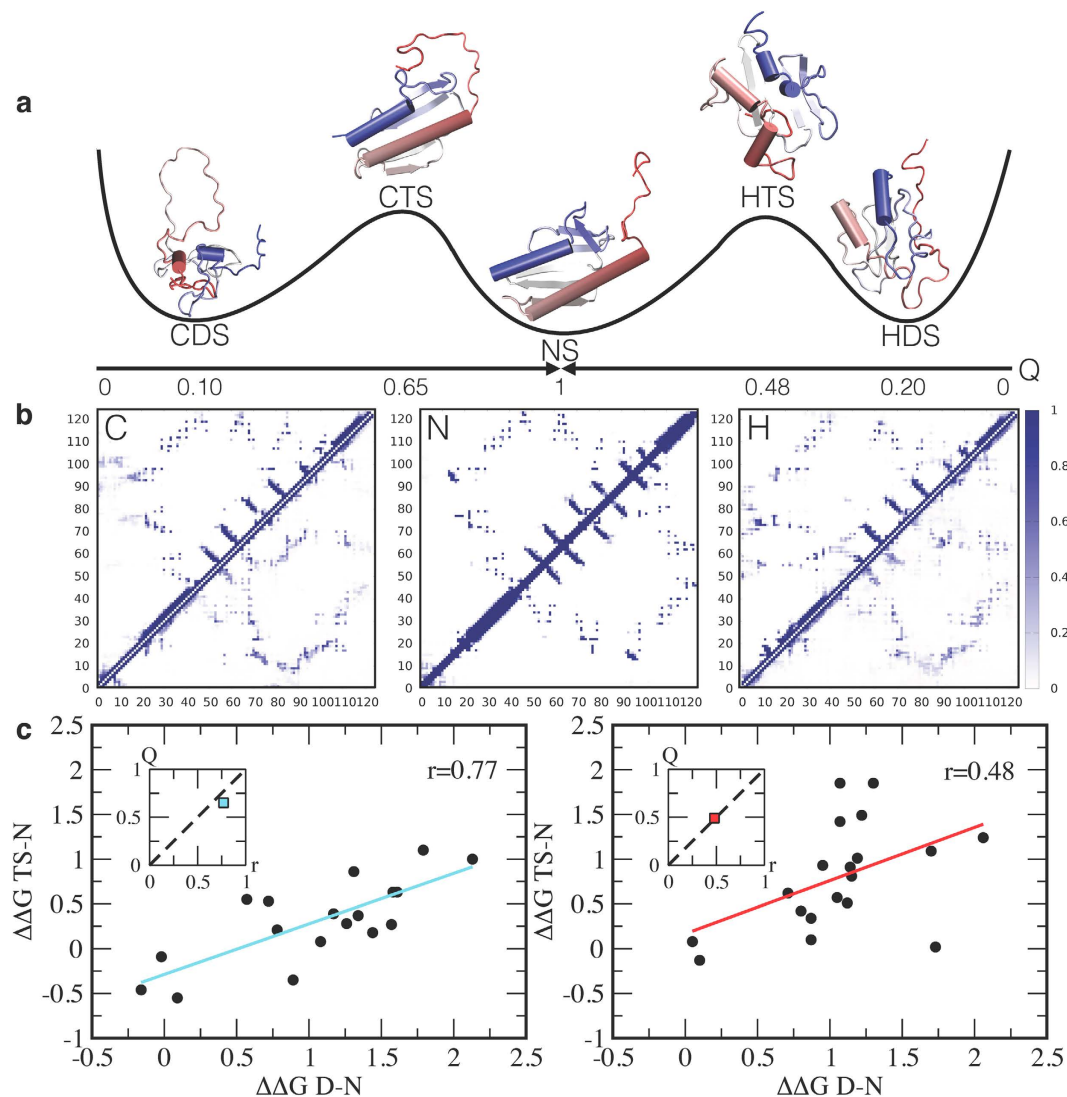
Figure 4. Frataxin folds following different pathways from the cold and hot denatured states. (a) Schematic representation of the cold and hot folding pathways along the fraction of native contacts Q. The structural ensembles shown correspond the most populated microstates in the two denatured states, the two transition states and the native state. **(b)** Contact maps for the cold denatured, native and hot denatured states; the color code corresponds to the fraction of contact formation. The contact map of the cold transition state (upper half) and the cold denatured states (lower half) is shown on the left, the contact map for the native state in the middle, and the contact map of the hot transition state (upper half) and the hot denatured states (lower half) on the right. **(c)** Brønsted plots determined at 287 K and 323 K from mutagenesis (see Methods). At low temperature the data are more linearly correlated than at high temperature, corresponding to a higher degree of structure in the cold transition state than in the hot transition state, as reported by the corresponding values of Q (insets).

protein-water hydrogen bonds is rescued by protein-protein hydrogen bonds resulting in more compact conformations than those observed by standard proteins of similar length[30,40,41].

In summary, we have shown that the analysis of the hydrogen bonds between water and protein molecules can provide a detailed characterisation of the hydrophobic effect in protein folding. Our results suggest that proteins behave rather closely according to the classic description of small hydrophobic particles[1,42–45], as we show that interface water molecules do not lose hydrogen bonds with respect to bulk water (Fig. 5), but only rotational entropy.

Materials and Methods

Molecular dynamics simulations. Molecular dynamics simulations of yeast frataxin were performed using the Amber03W force field[46] with the TIP4P05 water model[47]. All the simulations were run in GROMACS[48] using PLUMED2[49] and Almost[50]. A time step of 2 fs was used together with LINCS constraints[51]. Van der Waals and Coulomb interactions were implemented with a cutoff at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method[52] on a grid with a mesh of 0.1 nm. All simulations were carried out
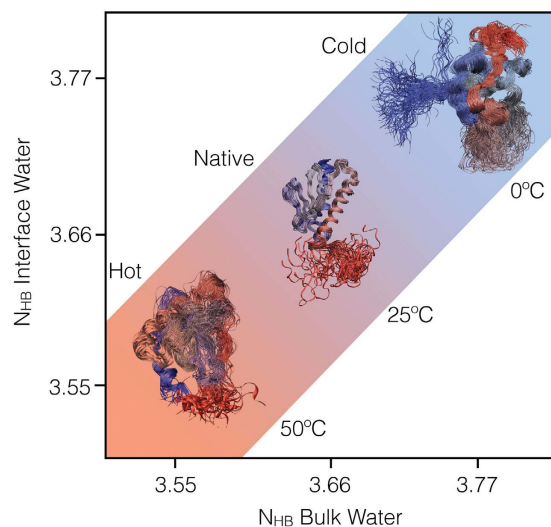
Figure 5. Our calculations indicate that bulk and surface water molecules form the same number of hydrogen bonds. As a consequence of this result, as the temperature is varied the conformation of the protein should adapt to this requirement which causes it to unfold both a low and at high temperatures.

in the canonical ensemble at constant volume and by thermosetting the system using a stochastic velocity rescaling[53]. The starting conformation was taken from an available NMR structure (PDB code 2GA5) and solvated with 19,000 water molecules and 15 sodium ions. A standard 100 ns molecular dynamics simulation at 298 K was performed as a reference for the native state ensemble.

Replica averaged metadynamics (RAM) simulations of the cold and hot denatured state ensembles.
A high-temperature (450 K) 30 ns preliminary unfolding simulation was used to select four starting conformations. Each conformation was then subsequently relaxed at the two target temperatures (272 and 323 K) for 10 ns at constant pressure. The final box size was adjusted to a common average value at each temperature and the systems have been additionally relaxed for 10 ns at constant volume. RAM simulations[21,22] were performed using chemical shifts (BMRB depositions 17068 and 17641) as replica-averaged restraints[21,22] and bias-exchange metadynamics[54] following a protocol recently introduced to sample the denatured state of ACBP[22]. Briefly, four replicas of the system were simulated in parallel at the target temperature with a restraint applied on the average value of the CamShift[55,56] back-calculated NMR chemical shifts[56] with a force constant set to 24 kJ/(mol ppm²). In this way the system evolves with a force field that is perturbed in such a way to increase the agreement with the experimental chemical shifts as resulting by the application of the maximum entropy principle[24,57–59]. In principle the number of replicas can be increased at expense of an increasing computational cost, but, as previously shown[21], four replicas are sufficient to recover with very good accuracy the dynamics from chemical shifts. Four collective variables (CVs) have been employed to enhance the search in the conformational space with bias-exchange metadynamics: the total α-helical content[48], the total β-sheet content[48], the radius of gyration and the number of contacts between the regions containing the native α-helices and the native β-sheet. The choice of the secondary structures and radius of gyration as CVs was guided by the hypotheses that transient secondary structures formation and the volume fluctuations of the polypeptide chain capture to a good extent the relevant dynamics of a disordered protein. The additional choice of the number of contacts between the regions containing the native helices and the native β-sheet was suggested by the fact that in a recent characterization of the transition state ensembles frataxin folding from an urea-denatured state most of the lost contact in the transition state are in this region. Gaussians deposition was performed with an initial rate of 0.125 kJ/mol/ps, where the σ values were set to 0.16, 0.1, 0.02 and 0.17, for the four CVs, respectively. In order to keep under control the convergence of the simulations we rescaled the height of the Gaussians using the well-tempered scheme[60] with a bias-factor of 10. Furthermore, in order to limit the extent of accessible space along each collective variable and correctly treat the problem of the borders, we set the bias as constant outside a defined interval for each CV[61], as it has been shown that this approach lead to a correct reconstruction of a one-dimensional free energy landscape inside the chosen range; intervals were set to 0–14, 0–10, 1.2–4.0 and 1–18 for the four CVs, respectively. Each replica have been evolved for 800 ns, with exchange trials every 50 ps. Protein conformations have been saved every 20 ps and full systems conformations (protein and solvent) have been saved every 100 ps for further analysis.

Cold and hot transition states ensembles.
Cold and hot transition state ensembles were determined following a standard procedure based on the interpretation of Φ-value analysis in terms of fraction of native contacts[26]. Briefly, given a set of experimental Φ−values, a pseudo energy term has been added to the force field as the squared difference between experimental and simulated Φ−values in order to maximize the agreement with the experimental value while keeping the simulation stable. The Φ−value for a residue i is calculated from the fraction of native contacts that it makes in a conformation. Given two residues that are not nearest neighbours,

the native contacts between them are defined as the number of heavy side-chain atoms within 0.65 nm in the native structure. With this approach only Φ−values between 0 and 1 can be incorporated as structural restraints.

The different transition state ensembles were generated using simulated annealing. Each ensemble is the results of 300 annealing cycles, 150 ps long, in which the temperature is varied between 272 K or 323 K and 500 K. Only the structures sampled at the reference temperatures are retained for further analysis, resulting in TSE of ~1000 structures each.

Bulk water simulations.   A parallel tempering simulation[49] of a box of 714 TIP4P05 water molecules was performed to analyse the property of water using seven replicas at 272, 274, 283, 292, 302, 312 and 323K, respectively. Each replica has been evolved for 20 ns in the NPT ensemble using the Parrinello-Rahman barostat[62] and try exchanges every 0.5 ps.

References

1. Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437,** 640–647 (2005).
2. Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **108,** 74–108 (2008).
3. Levy, Y. & Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 389–415 (2006).
4. Berne, B. J., Weeks, J. D. & Zhou, R. Dewetting and hydrophobic interaction in physical and biological systems. *Annu. Rev. Phys. Chem.* **60,** 85 (2009).
5. Dill, K. A., Truskett, T. M., Vlachy, V. & Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Biomol. Struct.* **34,** 173–199 (2005).
6. Rasaiah, J. C., Garde, S. & Hummer, G. Water in nonpolar confinement: From nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **59,** 713–740 (2008).
7. Godawat, R., Jamadagni, S. N. & Garde, S. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proc. Natl. Acad. Sci. USA* **106,** 15119–15124 (2009).
8. Garde, S. 150 water near proteins and interfaces: A new molecular perspective. *J. Biomol. Struct. Dyn.* **33,** 97–97 (2015).
9. Sirovetz, B. J., Schafer, N. P. & Wolynes, P. G. Water mediated interactions and the protein folding phase diagram in the temperature–pressure plane. *J. Phys. Chem. B* **119,** 11416–11427 (2015).
10. Cheng, Y.-K. & Rossky, P. J. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392,** 696–699 (1998).
11. Dobson, C. M., Šali, A. & Karplus, M. Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.* **37,** 868–893 (1998).
12. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4,** 10–19 (1997).
13. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21,** 167–195 (1995).
14. Ben-Naim, A. Theory of cold denaturation of proteins. *Adv. Biol. Chem.* **3,** 29–39 (2013).
15. Pastore, A. *et al.* Unbiased cold denaturation: Low-and high-temperature unfolding of yeast frataxin under physiological conditions. *J. Am. Chem. Soc.* **129,** 5374–5375 (2007).
16. Adrover, M. *et al.* The role of hydration in protein stability: Comparison of the cold and heat unfolded states of Yfh1. *J. Mol. Biol.* **417,** 413–424 (2012).
17. Bonetti, D. *et al.* The kinetics of folding of frataxin. *Phys. Chem. Chem. Phys.* **16,** 6391–6397 (2014).
18. Sanfelice, D., Morandi, E., Pastore, A., Niccolai, N. & Temussi, P. A. Cold denaturation unveiled: The molecular mechanism of asymmetric unfolding of yeast frataxin. *ChemPhysChem* **16,** 3599–3602 (2015).
19. Makhatadze, G. I. & Privalov, P. L. Energetics of protein structure. *Adv. Protein Chem.* **47,** 307–425 (1995).
20. Privalov, P. L. Cold denaturation of protein. *Crit. Rev. Biochem. Mol. Biol.* **25,** 281–306 (1990).
21. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-averaged metadynamics. *J. Chem. Theory Comput.* **9,** 5610–5617 (2013).
22. Camilloni, C. & Vendruscolo, M. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J. Am. Chem. Soc.* **136,** 8982–8991 (2014).
23. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99,** 12562–12566 (2002).
24. Cavalli, A., Camilloni, C. & Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **138,** 094112 (2013).
25. Fersht, A. R. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding.* (Freeman, W. H., 1999).
26. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **409,** 641–645 (2001).
27. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224,** 771–782 (1992).
28. Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Current opinion in structural biology* **24,** 98–105 (2014).
29. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA* **92,** 3626–3630 (1995).
30. Aznauryan, M., Nettels, D., Holla, A., Hofmann, H. & Schuler, B. Single-molecule spectroscopy of cold denaturation and the temperature-induced collapse of unfolded proteins. *J. Am. Chem. Soc.* **135,** 14040–14043 (2013).
31. Stillinger, F. H. Water revisited. *Science* **209,** 451–457 (1980).
32. Wernet, P. *et al.* The structure of the first coordination shell in liquid water. *Science* **304,** 995–999 (2004).
33. Luzar, A. & Chandler, D. Effect of environment on hydrogen bond dynamics in liquid water. *Phys. Rev. Lett.* **76,** 928 (1996).
34. Lopez, C. F., Darst, R. K. & Rossky, P. J. Mechanistic elements of protein cold denaturation. *J. Phys. Chem. B* **112,** 5961–5967 (2008).
35. Ben-Naim, A. Theory of cold denaturation of proteins. *Adv. Biol. Chem.* **3,** 29–39 (2013).
36. Bianco, V. & Franzese, G. Contribution of water to pressure and cold denaturation of proteins. *Phys. Rev. Lett.* **115,** 108101 (2015).
37. Fersht, A. R. Relationship of Leffler (Brønsted) α values and protein folding Φ values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. USA* **101,** 14338–14342 (2004).
38. Fersht, A. R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA* **92,** 10869–10873 (1995).
39. Karplus, M. & Weaver, D. L. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **3,** 650–668 (1994).
40. Wuttke, R. *et al.* Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc. Natl. Acad. Sci. USA* **111,** 5213–5218 (2014).
41. Granata, D. *et al.* The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **5,** 15449 (2015).
42. Stillinger, F. H. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Solution Chem.* **2,** 141–158 (1973).

43. Lum, K., Chandler, D. & Weeks, J. D. Hydrophobicity at small and large length scales. *J. Phys. Chem. B* **103,** 4570–4577 (1999).
44. Raschke, T. M., Tsai, J. & Levitt, M. Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water. *Proc. Natl. Acad. Sci. USA* **98,** 5965–5969 (2001).
45. Southall, N. T. & Dill, K. A. The mechanism of hydrophobic solvation depends on solute radius. *J. Phys. Chem. B* **104,** 1326–1331 (2000).
46. Best, R. B. & Mittal, J. Protein simulations with an optimized water model: Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* **114,** 14916–14923 (2010).
47. Abascal, J. L. & Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **123,** 234505 (2005).
48. Pietrucci, F. & Laio, A. A collective variable for the efficient exploration of protein beta-sheet structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **5,** 2197–2201 (2009).
49. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. Plumed 2: New feathers for an old bird. *Comp. Phys. Comm.* **185,** 604–613 (2014).
50. Fu, B. *et al.* Almost: An all atom molecular simulation toolkit for protein structure determination. *J. Comp. Chem.* **35,** 1101–1105 (2014).
51. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4,** 116–122 (2008).
52. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An NlogN method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).
53. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126,** 014101 (2007).
54. Piana, S. & Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **111,** 4553–4559 (2007).
55. Kohlhoff, K. J., Robustelli, P., Cavalli, A., Salvatella, X. & Vendruscolo, M. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* **131,** 13894–13895 (2009).
56. Robustelli, P., Kohlhoff, K., Cavalli, A. & Vendruscolo, M. Using NMR chemical shifts as structural restraints in molecular dynamics simulations of proteins. *Structure* **18,** 923–933 (2010).
57. Pitera, J. W. & Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **8,** 3445–3451 (2012).
58. Roux, B. & Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **138,** 084107 (2013).
59. Boomsma, W., Ferkinghoff-Borg, J. & Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comp. Biol.* **10,** e1003406 (2014).
60. Barducci, A., Bussi, G. & Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100,** 020603 (2008).
61. Baftizadeh, F., Cossio, P., Pietrucci, F. & Laio, A. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Curr Phys Chem* **2,** 79–91 (2012).
62. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. App. Phys.* **52,** 7182–7190 (1981).
63. Ceriotti, M., Tribello, G. A. & Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. USA* **108,** 13023–13028 (2011).

Author Contributions

C.C., C.M.D., M.B., S.G. and M.V. designed the study. C.C., C.M.D., M.B., S.G. and M.V. analysed the data. C.C. performed the calculations. D.B., A.M. and R.G. performed the experiments. C.C., C.M.D., M.B., S.G. and M.V. wrote the manuscript. All authors reviewed the manuscript.

Additional Information