

ARTICLE

Received 10 Dec 2015 | Accepted 9 Jan 2017 | Published 7 Feb 2017

DOI: 10.1038/ncomms14561

OPEN

# Molecular determinant of the effects of hydrostatic pressure on protein folding stability

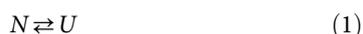
Calvin R. Chen<sup>1</sup> & George I. Makhatadze<sup>1</sup>

Hydrostatic pressure is an important environmental variable that plays an essential role in biological adaptation for many extremophilic organisms (for example, piezophiles). Increase in hydrostatic pressure, much like increase in temperature, perturbs the thermodynamic equilibrium between native and unfolded states of proteins. Experimentally, it has been observed that increase in hydrostatic pressure can both increase and decrease protein stability. These observations suggest that volume changes upon protein unfolding can be both positive and negative. The molecular details of this difference in sign of volume changes have been puzzling the field for the past 50 years. Here we present a comprehensive thermodynamic model that provides in-depth analysis of the contribution of various molecular determinants to the volume changes upon protein unfolding. Comparison with experimental data shows that the model allows quantitative predictions of volume changes upon protein unfolding, thus paving the way to proteome-wide computational comparison of proteins from different extremophilic organisms.

<sup>1</sup>Department of Biological Sciences and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180, USA. Correspondence and requests for materials should be addressed to G.I.M. (email: makhag@rpi.edu).

Many extremophilic organisms, the so-called barophiles, evolved to live under high hydrostatic pressure<sup>1,2</sup>. These organisms generally populate the deep ocean floor where hydrostatic pressure can reach 1,100 atm<sup>3,4</sup>. Evolving to live under high hydrostatic pressure is not exclusive to single cell organisms. For example, the segmented microscopic animal tardigrade ('water bear') can tolerate in their dormant state pressures up to 6,000 atm<sup>5</sup>. Pompeii worms (*Alvinella pompejana*) are species of polychaete worms that live at high pressures and temperatures near hydrothermal vents on the ocean floor<sup>6</sup>. Bacterial species have been isolated from 1,351 m into the Earth's crust where temperatures reach 102 °C and pressure is estimated to be in excess of 3,000 atm<sup>7,8</sup>. There are also reports of prokaryotic organisms at the bottom of oil well sediments and deep in the Arctic ice<sup>9</sup>.

What are the potential physicochemical implications for adaptation to high hydrostatic pressure? Biomacromolecules (proteins, DNA, RNA, lipid membrane) adopt unique three-dimensional structures (native or folded state) that are required for their biological function. The stability of these structures is very important for their function and thus biomacromolecules need to evolve to remain folded under the respective living conditions<sup>4,10–12</sup>. Increasing pressure, much like increasing temperature, perturbs the thermodynamic equilibrium between native folded state, *N*, and denatured unfolded state, *U*.



The response of the system to changes in pressure or temperature is governed by Le Chatelier's principle. Le Chatelier's principle is 'the tendency of a system to return to equilibrium by moving in the direction opposite to that caused by external perturbation'<sup>13</sup>. For a two-state equilibrium between *N* and *U* states, the temperature response is described by a well-known van't Hoff equation.

$$\ln K = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \quad (2)$$

where  $\Delta H = H_U - H_N$  is the enthalpy change upon unfolding, *K* is the equilibrium constant and *R* is the universal gas constant. For a two-state equilibrium between *N* and *U* states, the pressure response is defined by a sign of the volume change upon unfolding,  $\Delta V_{\text{Tot}}$

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = -\frac{\Delta V_{\text{Tot}}}{RT} = -\frac{V_{\text{Tot,U}} - V_{\text{Tot,N}}}{RT} \quad (3)$$

Combining equations (2) and (3) allows deriving the pressure-temperature phase diagram for protein stability<sup>14</sup>:

$$\begin{aligned} \Delta G(P, T) &= \Delta H - \Delta S \\ &\cdot T - \Delta C_p \left[ T \cdot \ln\left(\frac{T}{T_0} - 1\right) + T_0 \right] + \Delta \alpha \\ &\cdot (P - P_0) \cdot (T - T_0) + \Delta V_{\text{Tot}} \\ &\cdot (T - T_0) + \Delta \beta / 2 \cdot (P - P_0)^2 \end{aligned} \quad (4)$$

where  $\Delta H$ ,  $\Delta S$  and  $\Delta V_{\text{Tot}}$  are the unfolding enthalpy, entropy and volume changes at a reference temperature,  $T_0$ , and reference pressure,  $P_0$ ,  $\Delta C_p$  is the heat capacity change of unfolding. The reference temperature,  $T_0$ , and pressure,  $P_0$ , according to the biochemical convention are usually chosen to be 300 K and 0.1 MPa (= 1 atm). The change in compressibility upon unfolding,  $\Delta \beta$ , defined as  $\Delta \beta = (\partial \Delta V / \partial P)_T$  is very small and often ignored<sup>15</sup>. The change in thermal expansivity upon unfolding,  $\Delta \alpha$ , defined as  $\Delta \alpha = (\partial \Delta V / \partial T)_P$  is known to be positive<sup>16</sup>. Therefore,  $\Delta V_{\text{Tot}}$  is the key parameter in defining the effect of hydrostatic pressure on protein stability. According to Le Chatelier's principle, if  $\Delta V_{\text{Tot}} = V_{\text{Tot,U}} - V_{\text{Tot,N}} < 0$ , the

increase in pressure will shift the equilibrium from *N* to *U*, that is, will lead to a decrease in stability. Similarly, if  $\Delta V_{\text{Tot}} = V_{\text{Tot,U}} - V_{\text{Tot,N}} > 0$ , the increase in pressure will shift the equilibrium from *U* to *N*, that is, will lead to an increase in stability.

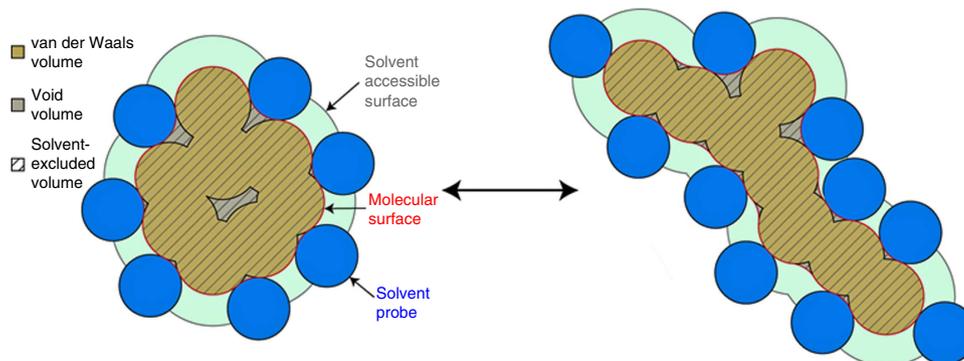
Here we will limit our discussion to only one type of biopolymers—proteins—but some of the issues can be pertinent to nucleic acids and to a lesser degree biological membranes. Experimental studies of the effects of hydrostatic pressure on protein stability have shown that most proteins unfold with an increase in pressure (negative  $\Delta V_{\text{Tot}}$ ), although some are actually stabilized (positive  $\Delta V_{\text{Tot}}$ ). The volume changes estimated from these experiments and also from direct measurements using pressure perturbation calorimetry are very small and range from  $-4\%$  to  $+0.5\%$ , relative to the total volume of the protein, that is,  $\Delta V_{\text{Tot}}/V_{\text{Tot,N}}$ <sup>16–19</sup>. Such small in amplitude and variable in sign changes in  $\Delta V_{\text{Tot}}$  have been puzzling the field for the past 50 years<sup>20–22</sup>. Since the first high-resolution structures of proteins have been solved, it has become evident that the native state of proteins contains a large number of voids (space inside protein that is not occupied by protein atoms) and cavities (voids large enough to accommodate a water molecule but may or may not be occupied by water)<sup>23</sup>. It has been shown that the volume fraction of the native protein that is occupied by voids ranges between 20 and 30%<sup>23–25</sup>. The presence of voids in the native state qualitatively explains the higher volume of the native state relative to the unfolded state. However, the magnitude of the negative volume changes due to the voids is much larger than the total negative volume changes observed experimentally and, moreover, cannot explain the experimentally observed positive volume changes. This led to a suggestion that possibly the volume changes due to the hydration of atoms inaccessible to the solvent in the native state and becoming exposed upon unfolding might lead to the positive contribution that offsets the negative void volume changes<sup>26,27</sup>. However, the transfer of model compounds from non-polar solvent into aqueous solution was also shown to be accompanied by negative volume changes<sup>27</sup>. This discrepancy between experimental volume changes and volume changes expected from protein structure and transfer studies has been termed the 'Protein Volume Paradox'<sup>27,28</sup>. We performed detailed analysis of the two major assumptions ((1) all voids in the native state contribute to the total volume change upon unfolding and (2) hydration leads to a decrease in volume) that led to the formulation of this apparent paradox and showed that they are both oversimplified. We show that the volume changes upon protein unfolding can be calculated by explicitly modelling the volumes of the unfolded state ensemble and using a proper thermodynamic transfer model to account for the volume changes upon hydration.

## Results

**Definitions.** The volume that a protein molecule occupies in solution is an additive quantity<sup>23</sup> and can be separated into the geometric volume of the protein ( $V_{\text{SE}}$ ), and the volume changes in solvent due to the interactions with the protein surface ( $V_{\text{Hyd}}$ ). The geometric volume is the volume that is encompassed by the molecular surface as first defined by Richards<sup>23</sup> and will be referred here as solvent-excluded volume ( $V_{\text{SE}}$ ). The  $V_{\text{SE}}$  of a protein comprises its van der Waals ( $V_{\text{vdW}}$ ) and void volume ( $V_{\text{Void}}$ ) (Fig. 1).

$$V_{\text{SE}} = V_{\text{vdW}} + V_{\text{Void}} \quad (5)$$

The van der Waals volume is the volume occupied by protein atoms, each given a specific van der Waals radius, while the void volume is the space inside the protein that is not occupied by



**Figure 1 | Pictorial definitions of volume changes upon protein unfolding.** The volume enclosed by the molecular surface (red line) is the geometric or solvent-excluded volume (shaded area,  $V_{SE}$ ). Molecular surface is calculated by using solvent probe of 1.4 Å (blue spheres). The solvent-excluded volume consists of van der Waals volume (dark yellow area,  $V_{vdW}$ ), that is, the volume occupied by protein atoms, and void volume (grey area,  $V_{void}$ ). Upon protein unfolding the molecular surface of the protein increases and some of the voids become solvent exposed.

protein atoms and is inaccessible to the solvent molecules modelled as spherical probes. The total volume of the protein in solution,  $V_{Tot}$ , can be defined as the sum of two components:

$$V_{Tot} = V_{SE} + V_{Hyd} \quad (6)$$

The volume changes upon protein unfolding are defined as the difference in total volumes of the unfolded,  $V_{Tot,U}$ , and native,  $V_{Tot,N}$ , states (Fig. 1).

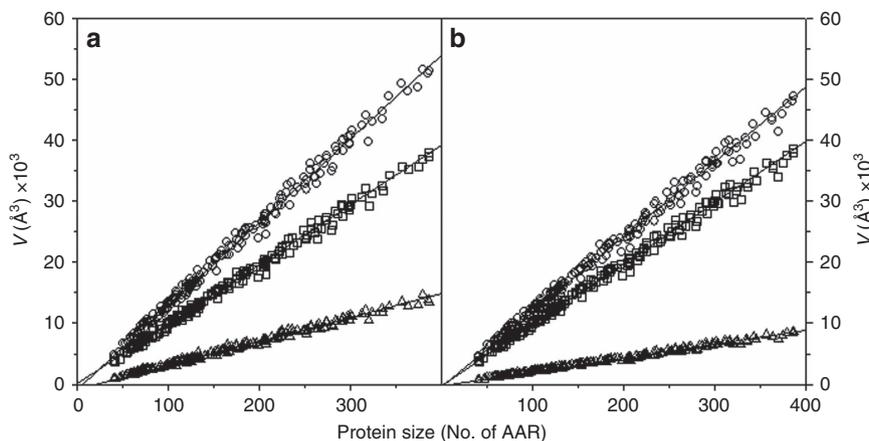
$$\Delta V_{Tot} = V_{Tot,U} - V_{Tot,N} = \Delta V_{SE} + \Delta V_{Hyd} \quad (7)$$

In the following section we will discuss the volumetric properties of the native and unfolded states and the contributions of the individual components,  $\Delta V_{SE}$ , and  $\Delta V_{Void}$ , to the total volume changes upon unfolding of proteins.

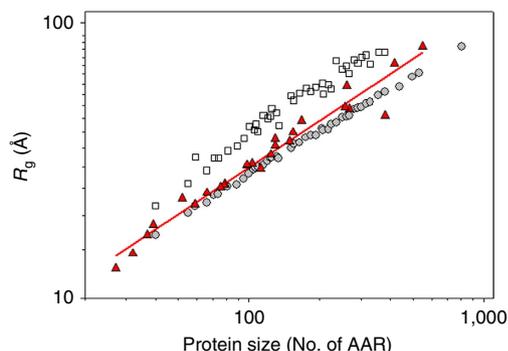
**Modelling the native state ensemble.** The properties of the native state are often modelled based on the X-ray and to a lesser degree NMR structures. However, the native state of a protein is not static, as portrayed by crystal structures, but dynamic. Thus, any property of the native state, such as volume, should be described by the properties of the ensemble. This allows the determination of not only the average property, but also the width of the distribution of volumes as reported by the s.d. Considering that molecular dynamics (MD) simulation is a well-established method to provide information about protein conformational fluctuations, we have used this method to model the fluctuations of native proteins<sup>29</sup>. As a test case for all calculations we have selected a non-redundant set of over 200 proteins of different size for which structures were solved with ultra-high (0.73–1.2 Å) resolution (see Supplementary Table 1 for the information on Protein Data Bank (PDB) structures used). For each protein, all-atom explicit solvent MD simulation was run starting from the X-ray coordinates as provided in the PDB file (see Methods for experimental details). The criterion for equilibration was the stable values of all-atom RMSD (root mean squared deviation) and stable values of solvent-excluded volume<sup>30</sup>. The average values of volume of structures from MD simulations were similar to the volumes obtained using the energy-minimized X-ray structure (Supplementary Fig. 1). The  $V_{SE,N}$  values for the native state ensembles scale linearly with protein size (Fig. 2a). The void volume also scales linearly with protein size (Fig. 2a) and the resulting packing density, defined as the ratio of van der Waals volume to solvent-excluded volume, varies between 0.7 and 0.8 (Supplementary Fig. 2) as was reported previously for other proteins<sup>23–25</sup>.

**Modelling of the unfolded state ensemble.** To draw conclusions about the volumetric properties of the unfolded state, it is necessary to explicitly model the unfolded state ensemble. Experimental characterization of the structural properties of the unfolded state remains very challenging. Available indirect data suggest that the unfolded state can contain residual helical structure, hydrophobic clusters and native-like contacts<sup>31–33</sup>. Other studies showed that the dimensions of the unfolded state ensemble have properties characteristic of a random coil polymer or even statistical coil<sup>34–36</sup>. The random and/or statistical coil representation of the unfolded state can, arguably, be considered the most extended conformational ensemble.

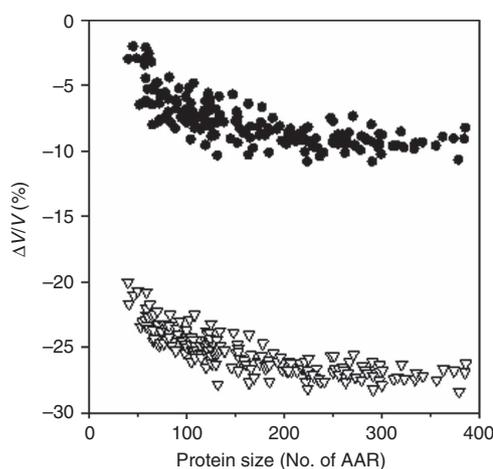
There are a number of ways to generate unfolded state ensembles<sup>35–39</sup>. It is well known that all-atom explicit solvent MD simulations produce overly compact or even highly helical structural ensembles that has been attributed to the artefacts of the current force-fields<sup>40</sup>. The empirical models or polymer-based models are free from such artefacts<sup>39</sup>. More importantly, the unfolded state ensembles, even those that differ by over 50% in the radius of gyration ( $R_g$ ), have very similar volumes (see Supplementary Fig. 3). This includes unfolded ensembles that incorporated significant fraction of helical structure. This underlines the fact noted by Fitzkee and Rose<sup>35</sup> that  $R_g$  is a very coarse property to describe the unfolded state ensemble and is not a sufficient criterion to assess the lack of or presence of, for example, secondary structure in the unfolded state ensemble. This also suggests that the use of any of these ensembles will produce qualitatively similar volumetric results. Based on these initial observations and on the fact that three computational models, statistical coil model of the unfolded state (SC)<sup>36</sup>, trajectory directed ensemble sampling (TraDES)<sup>37</sup> and flexible-meccano (FM)<sup>38</sup>, have been benchmarked against NMR-derived parameters such as chemical shifts and J-coupling constant<sup>36,41,42</sup>, we generated 1,000 unfolded conformations for each protein in our data set (see Supplementary Table 1 for the list of proteins). As expected, all ensembles show a lack of specific long-range contacts or secondary structure (see Supplementary Fig. 3). Furthermore,  $R_g$  compares well with the experimentally determined power-law dependence of  $R_g$  on protein size (that is, number of amino acid residues, see Fig. 3). The SC ensemble produces somewhat higher values of  $R_g$ <sup>36,39</sup>, while TraDES and FM produce smaller values of  $R_g$  and show remarkable agreement with the experimental values. There is ample evidence that pressure and chemical unfolding of proteins results in similar volume change<sup>43</sup>. Thus the unfolded state ensembles generated this way are consistent with the only



**Figure 2 | Breakdown of the contributions to the total geometric volume of protein.** Contributions of van der Waals  $V_{vdW}$ , squares) and void ( $V_{Void}$ , triangles) volumes to the total geometric volume ( $V_{SE}$ , circles) in the native (a) and unfolded (b) state ensembles as a function of protein size. Lines show linear regression fit. AAR, amino acid residue.



**Figure 3 | Radii of gyration ( $R_g$ ) of unfolded state ensembles generated using TraDES and SC compared with experimentally measured radii of gyration of unfolded proteins as a function of protein size.** TraDES-generated unfolded state ensemble shows a dependence of radii of gyration ( $R_g$ ) on protein size similar to experimentally measured values<sup>75</sup>. Red triangles show the experimentally measured (using SAXS) values of  $R_g$  of proteins of various sizes. Open squares show the  $R_g$  values calculated using SC-generated ensemble, while grey circles show the values calculated using TraDES ensemble. For clarity only every fifth data point is shown. See also Supplementary Fig. 4 that shows the results for the FM-generated ensemble.



**Figure 4 | Void volume changes upon unfolding as a function of protein size.** Comparison of the void volume changes expected by considering that all void volume of the native protein contributes to the  $\Delta V_{Void} = -V_{Void,N}$ , upon unfolding (triangles) with the volume changes that explicitly take into account the void volume of the unfolded state ensemble  $\Delta V_{Void} = V_{Void,U} - V_{Void,N}$  (circles).

currently available experimental data, that is, power law scaling of  $R_g$  for polymers in a good solvent. Although TraDES and FM show the closest agreement with experimental  $R_g$  values (Fig. 3 and Supplementary Fig. 4), due to limitation in the maximum sequence length in FM, TraDES was selected for all detailed calculations as a model of the unfolded state ensemble.

The  $V_{SE,U}$  values for the unfolded state ensemble also scale linearly with protein size but, as expected, the dependence is less steep than  $V_{SE,N}$  (see Fig. 2a,b). Importantly, the unfolded state ensemble also contains a significant amount of void volume that also scales linearly with the protein size (Fig. 2b). The packing density of the unfolded state ensemble (see Supplementary Fig. 2), defined as a ratio of van der Waals volume to solvent-excluded volume, is very uniform,  $0.824 \pm 0.005$  (max = 0.841; min = 0.815).

**Calculating  $\Delta V_{SE}$ .** The  $V_{SE}$  of a protein describes the geometric or structural contribution to protein volume. The changes in this volume upon protein unfolding then can be calculated by

subtracting the  $V_{SE}$  values of  $U$  and  $N$  state ensembles:

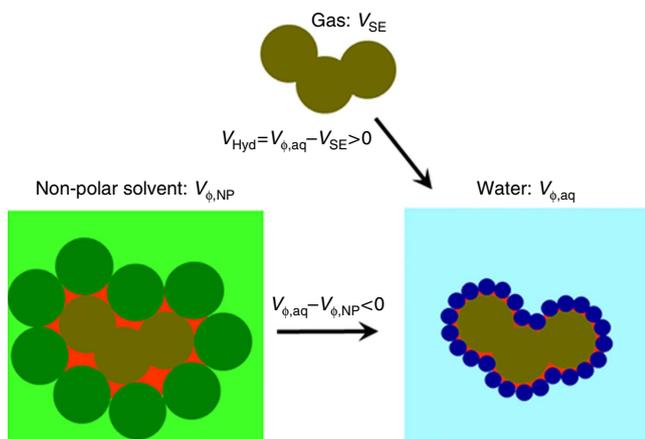
$$\begin{aligned} \Delta V_{SE} &= V_{SE,U} - V_{SE,N} = (V_{vdW,U} + V_{Void,U}) - \\ &(V_{vdW,N} + V_{Void,N}) = (V_{vdW,U} - V_{vdW,N}) + (V_{Void,U} - V_{Void,N}) \\ &= \Delta V_{vdW} + \Delta V_{Void} \end{aligned} \quad (8)$$

By this definition, the  $\Delta V_{SE}$  has two contributions, one from changes in the van der Waals volume,  $\Delta V_{vdW}$ , and another from the changes in void volume,  $\Delta V_{Void}$ . The van der Waals volume of the native state is always slightly lower than the van der Waals volume of the unfolded state (Fig. 4). This lower  $V_{vdW}$  of the native state is mainly associated with the extensive intramolecular hydrogen bonding in the native state<sup>30,44</sup>. Indeed, there is direct correlation between the difference in the average number of hydrogen bonds in the native and unfolded ensembles and the  $\Delta V_{vdW}$  (Supplementary Fig. 5). However, the  $\Delta V_{vdW}$  will not contribute to the total volume changes upon protein unfolding,  $\Delta V_{Tot}$ , because of the hydrogen bonding

with the solvent in the unfolded state that is of a similar magnitude and part of the hydration volume,  $\Delta V_{\text{Hyd}}$ .

As expected, the difference in void volumes between unfolded and native state  $\Delta V_{\text{Void}}$  is negative. When the polymer properties of the unfolded state ensemble are explicitly taken into account,  $\Delta V_{\text{Void}} = V_{\text{Void,U}} - V_{\text{Void,N}}$ , the change in void volume upon protein unfolding is much smaller ( $-7 \pm 2\%$ ) than the  $-20$  to  $-30\%$  previously defined by the elimination of all void volume of the native state ( $\Delta V_{\text{Void}} = -V_{\text{Void,N}}$ ) (Fig. 4). This suggests that the first postulate in the Protein Volume Paradox grossly overestimates the contribution of void volume to the total volume changes upon protein unfolding.

**Calculating volume of hydration  $\Delta V_{\text{Hyd}}$ .** When a protein unfolds, groups buried in the native state become exposed and will interact with water. These interactions can possibly lead to volume changes and thus must be accounted for when considering volume changes upon protein unfolding. Historically, this unfolding reaction was modelled as a transfer from non-polar solvent to water<sup>20,21,27</sup>. It has been shown that volume change upon such transfer, usually attributed to the volume changes upon hydration, is always negative, that is, volume of a solute in aqueous solution is always smaller than that in non-polar solvent<sup>20,21,27</sup>. However, transfer from non-polar solvent to aqueous solution has been proven to be inadequate to model thermodynamics of hydration (see for example, refs 45,46). In the case of volume changes, the larger size of non-polar solvent than that of water leads to an overestimation of the volume that a solute occupies in the non-polar phase<sup>47</sup> (see Fig. 5). It has been well established that transfer from gas phase into aqueous solution is a more appropriate way to model thermodynamics of hydration<sup>46,48–50</sup>. This is also the case for modelling the volume changes upon hydration (Fig. 5). To understand the volume changes upon hydration in more detail, we analysed the experimentally measured volume change upon transfer of over 150 different model solutes (aromatic and non-aromatic compounds, oligopeptides and *N*-acetyl amides of amino acids) from gas phase into aqueous solution at 25 °C<sup>51</sup>. The hydration volume upon transfer from gas phase into water,  $V_{\text{Hyd}}$ , is equal to the experimentally measured partial volume of a solute in water minus the geometric volume of this solute (that is,  $V_{\text{SE}}$ ) (Fig. 5).



**Figure 5 | Pictorial definition of hydration volume.** Volume of solute in non-polar phase ( $V_{\phi,\text{NP}}$ ) includes volume that a solute occupies in non-polar solvent (red area). The difference between the volume in the gas phase ( $V_{\text{SE}}$ ) and the partial volume of solute in water ( $V_{\phi,\text{aq}}$ ) accounts only for the volume changes due to the interactions with water. Thus, the hydration volume can be defined as  $V_{\text{Hyd}} = V_{\phi,\text{aq}} - V_{\text{SE}}$ .

This volume will, in a first approximation, depend on the number of water molecules that can directly interact with the solute and thus is proportional to the molecular surface area of the solute molecule. Figure 6a shows the dependence of the hydration volume,  $V_{\text{Hyd}}$ , on the total molecular surface area (MSA) of various solutes,  $MSA_{\text{Tot}}$ . Two major observations can be made from the plot shown in Fig. 6a. First, the hydration volume is always positive, that is, transfer of any (polar or non-polar) solute into aqueous solution leads to an increase in volume. Second, the increase in the molecular surface area leads to an increase in the corresponding hydration volume. However, there is a large spread in this trend (Fig. 6a). Interestingly, the correlation becomes much more pronounced ( $R^2 = 0.94$ ) if we plot only non-polar MSA ( $MSA_{\text{NP}}$ ) as a function of  $V_{\text{Hyd}}$  (see Fig. 6b). This suggests that the solvent interactions with non-polar surfaces make up a major contribution to the volume changes upon hydration. This observation is in line with some previous reports<sup>52</sup>; however, present analysis has been done on a much larger and more diverse data set. Importantly, not all variance in  $V_{\text{Hyd}}$  can be explained by  $MSA_{\text{NP}}$ . Therefore, we performed a fit to both non-polar,  $MSA_{\text{NP}}$ , and polar,  $MSA_{\text{Pol}}$ , surface areas

$$V_{\text{Hyd}} = a + k_{\text{NP}} \cdot MSA_{\text{NP}} + k_{\text{Pol}} \cdot MSA_{\text{Pol}} \quad (9)$$

where  $k_{\text{NP}}$  and  $k_{\text{Pol}}$  are the contributions of a given type of MSA to the hydration volume for non-polar and polar surfaces, respectively. The fit to equation (9) provides a much better description ( $R^2 = 0.97$ ) of the variance in  $V_{\text{Hyd}}$  (see Fig. 6b) than just a one parameter fit using only non-polar MSA. The  $k_{\text{NP}}$  coefficient is larger ( $0.38 \pm 0.2 \text{ \AA}$ ) than  $k_{\text{Pol}}$  ( $0.03 \pm 0.03 \text{ \AA}$ ). The small value for the polar coefficient can be easily rationalized as hydrogen bonding of water with polar groups, while the large value of the  $k_{\text{NP}}$  coefficient suggests that water molecules move away from non-polar groups due to the hydrophobic effect<sup>30,53–55</sup>.

Having empirically established the contribution of polar and non-polar groups to the volume of hydration and, assuming that such proportionality can be extrapolated to much larger surfaces, we can calculate the volume changes due to changes in hydration upon protein unfolding. Volume as a thermodynamic quantity is an additive state function that allows for the use of the thermodynamic cycle shown in Fig. 7. In this hypothetical thermodynamic transfer cycle, a native state ensemble is first transferred into the gas phase. This step essentially accounts for the volume change upon dehydration of the native state ensemble. The volume changes are calculated as:

$$\text{Step 1 } -V_{\text{Hyd,N}} = -(a + k_{\text{NP}} \cdot MSA_{\text{NP,N}} + k_{\text{Pol}} \cdot MSA_{\text{Pol,N}})$$

where the minus sign signifies that the transfer is from aqueous phase into gas phase (the opposite of hydration, that is, dehydration).

The second step is unfolding of a protein in the gas phase that corresponds to the changes in the geometric volume:

$$\text{Step 2 } \Delta V_{\text{SE}} = V_{\text{SE,U}} - V_{\text{SE,N}}$$

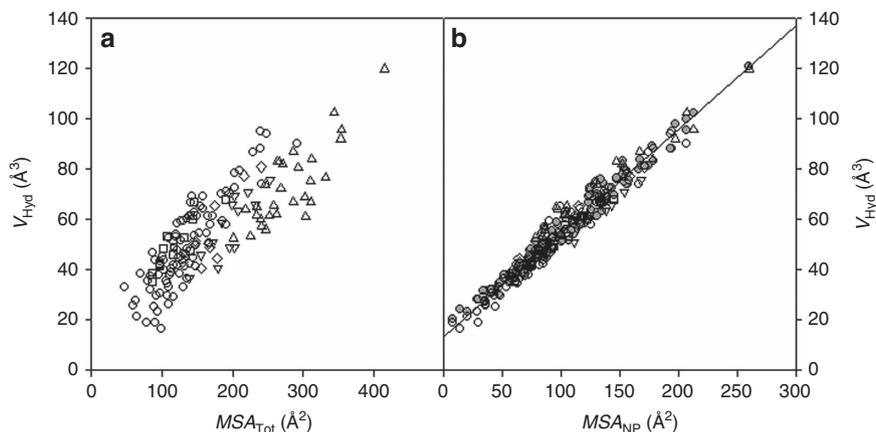
Finally, the unfolded state ensemble is transferred back to aqueous solution. This step corresponds to volume change upon hydration of the unfolded state ensemble and is calculated as

$$\text{Step 3 } V_{\text{Hyd,U}} = a + k_{\text{NP}} \cdot MSA_{\text{NP,U}} + k_{\text{Pol}} \cdot MSA_{\text{Pol,U}}$$

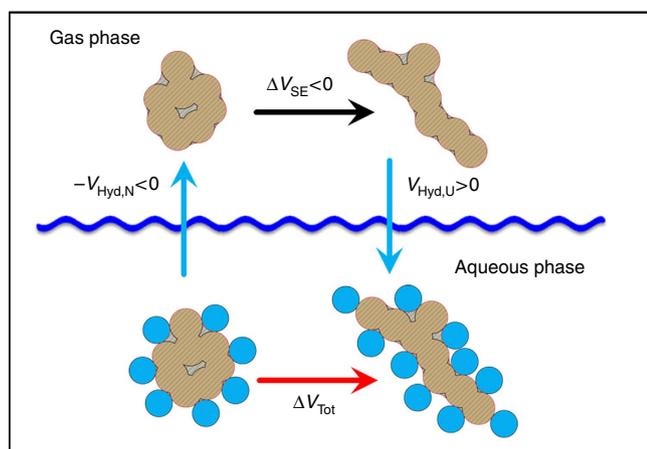
The sum of steps 1 and 3 reflects the change in hydration upon unfolding. Assuming that the proportionality coefficients  $k_{\text{NP}}$  and  $k_{\text{Pol}}$ , derived from model compounds (that is, equation (9)), can be extrapolated to much larger surfaces, the changes in hydration volume upon unfolding can be calculated as:

$$\Delta V_{\text{Hyd}} = k_{\text{NP}} \cdot \Delta MSA_{\text{NP}} + k_{\text{Pol}} \cdot \Delta MSA_{\text{Pol}} \quad (10)$$

where  $\Delta MSA_{\text{NP}} = MSA_{\text{NP,U}} - MSA_{\text{NP,N}}$  and  $\Delta MSA_{\text{Pol}} = MSA_{\text{Pol,U}} - MSA_{\text{Pol,N}}$  are the differences in the non-polar and polar surface areas of the unfolded and native states, respectively. These can be calculated using the native and



**Figure 6 | Hydration volume of model compounds.** The dependence of hydration volume,  $V_{\text{Hyd}}$ , of model compounds at 25 °C on the total ( $MSA_{\text{Tot}}$ , **(a)**) or non-polar ( $MSA_{\text{NP}}$ , **(b)**) surface area shows that non-polar groups make a major contribution. Aromatic model compounds ( $\square$ ); non-aromatic model compounds (circle); oligopeptides ( $n = 3-5$ ) (triangle); *N*-acetyl amides of amino acids (upside-down triangle); and *N*-acetyl amino acids (diamond). In **b**, the scatter in the grey circles calculated with equation (9) matches the scatter in  $V_{\text{Hyd}}$  data and the line shows linear regression of  $V_{\text{Hyd}}$  versus  $MSA_{\text{NP}}$ .



**Figure 7 | Thermodynamic cycle for separating the contributions of hydration of native ( $V_{\text{Hyd,N}}$ ) and unfolded ( $V_{\text{Hyd,U}}$ ) state ensembles from the contribution of geometric volume changes ( $\Delta V_{\text{SE}}$ ) to the total changes in volume ( $\Delta V_{\text{Tot}}$ ) upon protein unfolding in aqueous solution.**

The sum of all three steps is equal to the volume of unfolding of the protein in aqueous solution,  $\Delta V_{\text{Tot}}$ , as defined by equation (11). It must be emphasized that this process is valid because volume as a thermodynamic parameter is a state function and there are no conformational changes in the native or unfolded state ensembles upon transfer to and from the gas phase.

unfolded state ensembles modelled as described in previous sections. The resulting  $\Delta V_{\text{Hyd}}$  is always positive due to two factors: (1) the MSA of the unfolded state is larger than the MSA of the native state and (2) both  $k_{\text{NP}}$  and  $k_{\text{Pol}}$  coefficients are positive. This suggests that the second postulate in the Protein Volume Paradox misrepresents the contribution of hydration volume (that is actually positive and not negative) to the total volume changes upon protein unfolding.

**Comparison of computed and experimental values for  $\Delta V_{\text{Tot}}$ .** Results presented in the previous sections allow us to calculate the expected total changes in volume upon protein unfolding as:

$$\Delta V_{\text{Tot}} = V_{\text{Tot,U}} - V_{\text{Tot,N}} = \Delta V_{\text{Void}} + \Delta V_{\text{Hyd}} \quad (11)$$

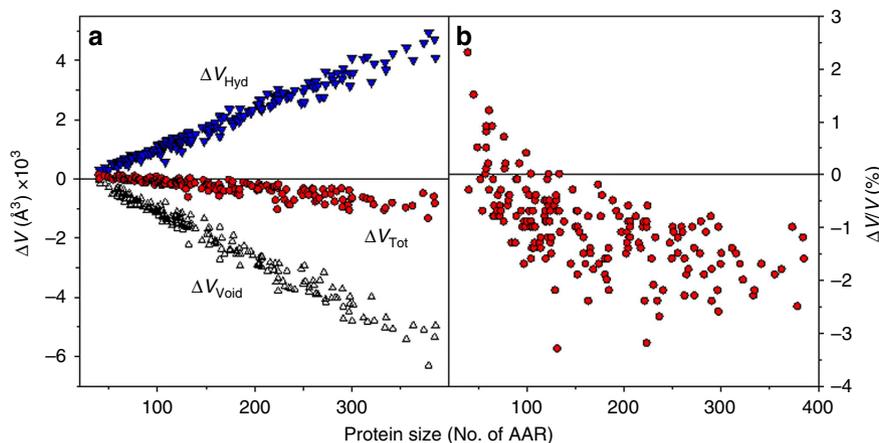
and analyse the relative contributions of each component. Figure 8a shows the protein size dependence of the calculated changes in the

$\Delta V_{\text{Tot}}$  and analyses the contributions from the changes in void,  $\Delta V_{\text{Void}}$ , and hydration,  $\Delta V_{\text{Hyd}}$ , volumes. The contribution of void volume to the total volume upon protein unfolding is relatively large and negative. It originates from the fact that there is larger void volume in the native state than in the unfolded state ensemble, represented here using the maximally unfolded polymer-based model (see also Figs 3 and 4). The volume changes due to hydration are also positive and comparable in absolute values to the corresponding absolute values of  $\Delta V_{\text{Void}}$ . This leads to rather small relative volume changes upon unfolding (Fig. 8b). Importantly, the balance between these two factors can produce both positive and negative changes in the total volume upon protein unfolding. This is in excellent qualitative agreement with the general experimental observations that proteins can be both stabilized (positive  $\Delta V_{\text{Tot}}$ ) and destabilized (negative  $\Delta V_{\text{Tot}}$ ) by hydrostatic pressure<sup>16-19,56,57</sup>.

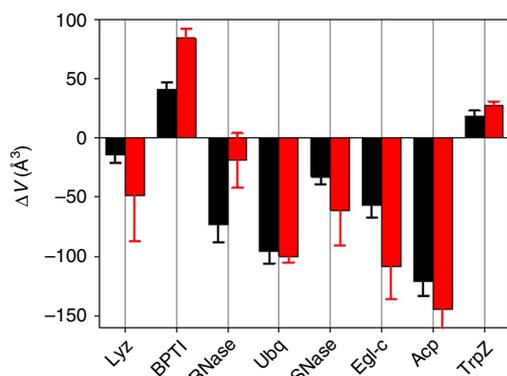
To test the reliability and accuracy of the presented formalism we have compared the calculated (using equation (11)) and experimentally measured values of volume change upon unfolding for eight proteins: lysozyme, pancreatic trypsin inhibitor, ribonuclease A, ubiquitin, acylphosphatase, eglin c, tryptophan zipper and staphylococcal nuclease  $\Delta + \text{PHS}$  variant<sup>16,57,58</sup>. Figure 9 shows that predicted values are in very good agreement with experimental values. In particular, the predicted values recapitulate not only the sign of the volume changes but also the magnitude. Thus, the formalism to calculate the volume changes upon protein unfolding presented here appears to provide near quantitative prediction of the expected volume changes. Furthermore, our results present a quantitative molecular picture of how hydrostatic pressure modulates the conformational equilibrium of proteins. Finally, the structure-based prediction of volume changes upon protein unfolding can now be applied to a proteome-wide comparison of proteins from organisms living under ambient pressure with those from organisms that evolved to live under extreme pressures. The expectation is that the proteins from barophilic organisms could have evolved to have volume changes upon unfolding that were either less negative or even positive relative to those of the mesophilic homologues, by decreasing the void volume in the native state and/or increasing the fraction of buried non-polar residues.

## Methods

**Sampling of native and unfolded ensembles.** A non-redundant set of proteins with X-ray structures solved to ultra-high resolution (0.73–1.2 Å resolution, see Supplementary Table 1) was selected for modelling.



**Figure 8 | The dependence of the total volume changes upon unfolding ( $\Delta V_{\text{Tot}}$  calculated using equation (11)) on protein size. (a) Contributions of void ( $\Delta V_{\text{Void}}$ , triangles) and hydration ( $\Delta V_{\text{Hyd}}$ , upside-down triangles) volume changes to the total volume ( $\Delta V_{\text{Tot}}$ , circles) changes upon unfolding as a function of protein size. (b) Fractional changes in the total volume  $\Delta V_{\text{Tot}}/V_{\text{Tot,N}}$  as a function of protein size.**



**Figure 9 | Direct comparison of experimentally measured (black bars) and calculated using equation (11) (red bars) volume changes upon unfolding of eight globular proteins.** Lyz, hen egg white lysozyme PDB:4LZT; BPTI, bovine pancreatic trypsin inhibitor PDB:6PTI; RNase, bovine pancreatic ribonuclease A PDB:7RSA; Ubq, human ubiquitin PDB:1UBQ; SNase,  $\Delta$ PHS variant of staphylococcal nuclease PDB:3BDC; Egl-c, leech eglin c PDB:1EGL; Acp, human acylphosphatase PDB:2ACY; TrpZ, Tryptophan Zipper PDB:1LE3. Experimental data (25 °C SNase; 40 °C RNase; 50 °C Acp, Egl and Lyz; 90 °C BPTI; and 80 °C TrpZ) are taken from refs 16,57,58. It is important to note that volume changes are temperature dependent<sup>19</sup> that can also contribute to the observed differences between experimental and calculated values. Error bars show s.d. of averaging the experimental data over measured temperature range or of the multiple ( $n = 3-8$ ) repeats of MD runs for the native state (see Methods section for details).

Native state all-atom explicit solvent MD simulations were carried out in GROMACS 4.6.3 (ref. 59) using the CHARMM27 force field and TIP3P water model<sup>60</sup>. The native crystal structure was solvated in a dodecahedron box, with dimensions such that all protein atoms are at least 10 Å deep in the box, and neutralized with 0.1 M excess NaCl, followed by energy minimization for 1,000 steps. All simulations underwent 200 ps of constant volume equilibration, 100 ps of constant pressure equilibration and 50 ns of production simulation at 300 K and 1 bar. We used the Parrinello–Rahman<sup>61</sup> pressure control with a 2 ps relaxation time and a compressibility of  $4.6 \times 10^{-5} \text{ atm}^{-1}$ , and v-scale temperature coupling<sup>59</sup>. LINCS<sup>62</sup> and SETTLE<sup>63</sup> algorithms were used to constraint high-frequency bond vibrations that allowed the use of a 2 ps integration step. The electrostatic interactions were modelled by the smooth particle mesh Ewald method<sup>64</sup>, using a  $75 \times 75 \times 75$  grid, with fourth-order charge interpolation and a real space cutoff of 1.0 nm. A native ensemble of 50 structures was extracted from the production trajectory (1 structure per ns). Equilibration criteria were stable all-atom r.m.s.d. with respect to the crystal structure ( $< 2 \text{ \AA}$ ) and stable values (drift 0.5%) of solvent-excluded volume throughout the production

trajectory. The low-frequency motions are expected to introduce only small corrections to the solvent-excluded volume, though further study would be warranted for proteins that undergo large-scale conformational fluctuations.

Sampling of the unfolded state ensembles was carried out using three generators: TraDES<sup>37</sup>, SC<sup>36</sup> and FM<sup>38</sup>. Each unfolded state ensemble consisted of 1,000 structures. TraDES ensembles were generated with the all-coil sampling flag (-c T) to remove all secondary structure propensity. Coil and FM ensembles were generated using default settings. SC and FM generate only the protein backbone, and hence Scwrl4 (ref. 65) was used to add side-chains. Before volume calculations, the structures generated by TraDES, SC or FM were energy minimized in implicit solvent, using the Generalized Born surface area (GBSA) model with protein and solvent dielectric constants of 80 in GROMACS 4.6.3. This step also explicitly incorporates all hydrogen atoms. The disulfide bridges for three individual proteins (BPTI, RNase and Lyz) modelled into the TraDES unfolded state ensemble by performing restrained all-atom MD simulations using the following protocol in GROMACS 4.6.3 with TIP3P water and CHARMM27 force field. First, the structures were energy minimized for 800 steps before enabling distance restraints. Distance restraints were gradually shortened in four 5 ps steps, where distance restraints were 3, 1, 0.5 and 0.2 nm, respectively. Upon reaching 0.2–0.3 nm distance between sulfur atoms, the structure was energy minimized in the CHARMM27 force field for 1,000 steps in implicit solvent, using the steepest descent algorithm.

**Model compounds.** The experimentally measured at 25 °C partial molar volumes in aqueous solution,  $V_{\phi,\text{aq}}$  of 150 model compounds relevant to proteins (alkanes, aromatic compounds, alcohols, diols, amines, amides, diamines, diamides, dicarboxylic acids, hydroxyamides, hydroxy acids, ketones, polyethylene glycols and ureas) were taken from refs 51,66. For each of these model compound PDB structures were generated using the CORINA webserver<sup>67</sup>. In addition, we included into our analysis more complex model compounds such as oligopeptides (3–5 residues), *N*-acetyl amino acid amides and *N*-acetyl amino acids that represent protein components, for which the experimentally measured partial molar volumes in aqueous solution,  $V_{\phi,\text{aq}}$  are also reported<sup>15,68,69</sup>. For these compounds, structures were generated using the same protocol as for the native proteins (see above). The hydration volume is calculated as a difference between partial molar volume of a compound in aqueous solution,  $V_{\phi,\text{aq}}$  and corresponding solvent-excluded volume,  $V_{\text{SE}}$  (see also Fig. 5). MSMS software package<sup>70</sup> was used to analytically calculate MSA per atom for all structures of model compounds. Each structure had its MSA broken down into carbon, nitrogen, oxygen and sulfur surface areas. Hydrogen surface areas were combined with the surface area of their parent heavy atom. Carbon and sulfur molecular surface areas were combined into non-polar MSA ( $MSA_{\text{NP}}$ ). Nitrogen and oxygen molecular surface areas were combined into polar MSA ( $MSA_{\text{POL}}$ ).

**ProteinVolume calculations.** The ProteinVolume software package has been described by us previously<sup>1</sup>. Briefly, it uses a flood-fill algorithm to first calculate the molecular surface based on the coordinates, and then fill the enclosed volume with 0.02 Å probes. The sum of all volume probes is calculated and reported as the solvent-excluded protein volume ( $V_{\text{SE}}$ ). Van der Waals volume ( $V_{\text{vdW}}$ ) is also calculated during the same step as the solvent-excluded volume calculation procedure, but with an additional check of whether the volume probe is within the van der Waals radius of a protein atom. A probe that lies on top of a van der Waals boundary is stochastically accepted with the acceptance probability based on its magnitude of overlap with the atom. This increases the accuracy of the van der Waals volume calculation and reduces the volume underestimation of numerical integration methods. The sum of all van der Waals volume probes is

calculated and reported as van der Waals protein volume ( $V_{vdW}$ ). Void volume,  $V_{void}$ , is calculated as the difference between the solvent-excluded volume and the van der Waals volume. ProteinVolume uses all-atom Bondi radii set<sup>72</sup>. United atom radii are slightly larger to represent the additional radius of bonded hydrogen atoms averaged across the entire atom surface and this leads to an overestimation of van der Waals volume. Volume changes upon ionization (electrostriction) vary between different protein groups (see Supplementary Table 2). Importantly, most charged groups are on the surface of native proteins, and thus at most 1–2 protons will be released or absorbed upon protein unfolding. This might have an effect for smaller proteins <100 residues, but small proteins rarely have buried charged groups<sup>73</sup>. Continuum electrostatics calculations using h++ server<sup>74</sup> of net charge in the native and unfolded ensembles for eight proteins that were quantitatively compared with the experimental  $\Delta V$  values (for example, Fig. 9) show that there is no change in the net charge upon unfolding of these proteins (see Supplementary Fig. 6).

**Data availability.** The data that support the findings of this study are available from the corresponding author on reasonable request.

## References

- Horikoshi, K. Barophiles: deep-sea microorganisms adapted to an extreme environment. *Curr. Opin. Microbiol.* **1**, 291–295 (1998).
- Lauro, F. M. & Bartlett, D. H. Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **12**, 15–25 (2008).
- Yayanos, A. A. Microbiology to 10,500 meters in the deep sea. *Annu. Rev. Microbiol.* **49**, 777–805 (1995).
- Robb, F. T. & Clark, D. S. Adaptation of proteins from hyperthermophiles to high pressure and high temperature. *J. Mol. Microbiol. Biotechnol.* **1**, 101–105 (1999).
- Seki, K. & Toyoshima, M. Preserving tardigrades under pressure. *Nature* **395**, 853–854 (1998).
- Borgonie, G. *et al.* Nematoda from the terrestrial deep subsurface of South Africa. *Nature* **474**, 79–82 (2011).
- Mason, O. U. *et al.* First investigation of the microbiology of the deepest layer of ocean crust. *PLoS ONE* **5**, e15399 (2010).
- Amend, J. P. & Teske, A. Expanding frontiers in deep subsurface microbiology. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **219**, 131–155 (2005).
- Bartlett, D. H. Introduction to high-pressure bioscience and biotechnology. *Ann. NY Acad. Sci.* **1189**, 1–5 (2010).
- Fan, H. Y., Nazari, M., Chowdhury, S. & Heerklotz, H. Volume and expansivity changes of Micelle formation measured by pressure perturbation calorimetry. *Langmuir* **27**, 1693–1699 (2011).
- Royer, C. A. Application of pressure to biochemical equilibria: the other thermodynamic variable. *Methods Enzymol.* **259**, 357–377 (1995).
- Sun, M. M. & Clark, D. S. Pressure effects on activity and stability of hyperthermophilic enzymes. *Methods Enzymol.* **334**, 316–327 (2001).
- Dill, K. A. & Bromberg, S. *Molecular Driving Forces: statistical Thermodynamics in Chemistry and Biology* (Garland Science, 2003).
- Hawley, S. A. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **10**, 2436–2442 (1971).
- Lee, S., Tikhomirova, A., Shalvardjian, N. & Chalikian, T. V. Partial molar volumes and adiabatic compressibilities of unfolded protein states. *Biophys. Chem.* **134**, 185–199 (2008).
- Schweiker, K. L., Fitz, V. W. & Makhatadze, G. I. Universal convergence of the specific volume changes of globular proteins upon unfolding. *Biochemistry* **48**, 10846–10851 (2009).
- Chalikian, T. & Breslauer, K. On volume changes accompanying conformational transitions of biopolymers. *Biopolymers* **39**, 619–626 (1996).
- Lin, L. N., Brandts, J. F., Brandts, J. M. & Plotnikov, V. Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal. Biochem.* **302**, 144–160 (2002).
- Schweiker, K. L. & Makhatadze, G. I. Use of pressure perturbation calorimetry to characterize the volumetric properties of proteins. *Methods Enzymol.* **466**, 527–547 (2009).
- Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63 (1959).
- Brandts, J. *Conformational Transitions of Proteins in Water and in Aqueous Mixtures* (Marcel Dekker, 1969).
- Zipp, A. & Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry* **12**, 4217–4228 (1973).
- Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
- Liang, J. & Dill, K. A. Are proteins well-packed? *Biophys. J.* **81**, 751–766 (2001).
- Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **290**, 253–266 (1999).
- Hvidt, A. A discussion of pressure-volume effects in aqueous protein solutions. *J. Theor. Biol.* **50**, 245–252 (1975).
- Kauzmann, W. Protein stabilization - thermodynamics of unfolding. *Nature* **325**, 763–764 (1987).
- Chalikian, T. V. Volumetric properties of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 207–235 (2003).
- Daggett, V. & Levitt, M. Realistic simulations of native-protein dynamics in solution and beyond. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 353–380 (1993).
- Dadarlat, V. M. & Post, C. B. Adhesive-cohesive model for protein compressibility: an alternative perspective on stability. *Proc. Natl Acad. Sci. USA* **100**, 14778–14783 (2003).
- Cho, J. H., Sato, S., Hornig, J. C., Anil, B. & Raleigh, D. P. Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch. Biochem. Biophys.* **469**, 20–28 (2008).
- Bowler, B. E. Thermodynamics of protein denatured states. *Mol. Biosyst.* **3**, 88–99 (2007).
- Baldwin, R. L. A new perspective on unfolded proteins. *Adv. Protein Chem.* **62**, 361–367 (2002).
- Flory, P. J. *Principles of Polymer Chemistry* (Cornell University Press, 1953).
- Fitzkee, N. & Rose, G. Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA* **101**, 12497–12502 (2004).
- Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl Acad. Sci. USA* **102**, 13099–13104 (2005).
- Feldman, H. & Hogue, C. Probabilistic sampling of protein conformations: new hope for brute force? *Proteins* **23**, 8–23 (2002).
- Ozenne, V. *et al.* Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470 (2012).
- Tran, H. T., Wang, X. & Pappu, R. V. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* **44**, 11369–11380 (2005).
- Gnanakaran, S. & Garcia, A. E. Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. *Proteins* **59**, 773–782 (2005).
- Jensen, M. R., Zweckstetter, M., Huang, J. R. & Backledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).
- Salmon, L. & Blackledge, M. Investigating protein conformational energy landscapes and atomic resolution dynamics from NMR dipolar couplings: a review. *Rep. Prog. Phys.* **78**, 126601–126631 (2015).
- Rouget, J. B. *et al.* Unique features of the folding landscape of a repeat protein revealed by pressure perturbation. *Biophys. J.* **98**, 2712–2721 (2010).
- Schell, D., Tsai, J., Scholtz, J. M. & Pace, C. N. Hydrogen bonding increases packing density in the protein interior. *Proteins* **63**, 278–282 (2006).
- Baldwin, R. L. The new view of hydrophobic free energy. *FEBS Lett.* **587**, 1062–1066 (2013).
- Makhatadze, G. I. & Privalov, P. L. Energetics of protein structure. *Adv. Protein Chem.* **47**, 307–425 (1995).
- Lee, B. Partial molar volume from the hard-sphere mixture model. *J. Phys. Chem.* **87**, 112–118 (1983).
- Ben-Naim, A. *Solvation Thermodynamics* (Springer, 1987).
- Baldwin, R. L. Properties of hydrophobic free energy found by gas-liquid transfer. *Proc. Natl Acad. Sci. USA* **110**, 1670–1673 (2013).
- Graziano, G. On the mechanism of cold denaturation. *Phys. Chem. Chem. Phys.* **16**, 21755–21767 (2014).
- Cabani, S., Gianni, P., Mollica, V. & Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Solution Chem.* **10**, 563–595 (1981).
- Makhatadze, G. I., Medvedkin, V. N. & Privalov, P. L. Partial molar volumes of polypeptides and their constituent groups in aqueous-solution over a broad temperature-range. *Biopolymers* **30**, 1001–1010 (1990).
- Hummer, G., Garde, S., Garcia, A. E., Paulaitis, M. E. & Pratt, L. R. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl Acad. Sci. USA* **95**, 1552–1555 (1998).
- Marcus, Y. Electrostriction in electrolyte solutions. *Chem. Rev.* **111**, 2761–2783 (2011).
- Giovambattista, N., Lopez, C. F., Rossky, P. J. & Debenedetti, P. G. Hydrophobicity of protein surfaces: separating geometry from chemistry. *Proc. Natl Acad. Sci. USA* **105**, 2274–2279 (2008).
- Mitra, L., Rouget, J. B., Garcia-Moreno, B., Royer, C. A. & Winter, R. Towards a quantitative understanding of protein hydration and volumetric properties. *Chemphyschem* **9**, 2715–2721 (2008).
- Roche, J. *et al.* Cavities determine the pressure unfolding of proteins. *Proc. Natl Acad. Sci. USA* **109**, 6945–6950 (2012).
- Pandharipande, P. P. & Makhatadze, G. I. Applications of pressure perturbation calorimetry to study factors contributing to the volume changes upon protein unfolding. *Biochim. Biophys. Acta* **1860**, 1036–1042 2016.

59. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
60. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
61. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
62. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
63. Miyamoto, S. & Kollman, P. A. Settle - an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
64. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
65. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. Jr Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778–795 (2009).
66. Zamyatnin, A. A. Amino acid, peptide, and protein volume in solution. *Annu. Rev. Biophys. Bioeng.* **13**, 145–165 (1984).
67. Sadowski, J., Gasteiger, J. & Klebe, G. Comparison of automatic three-dimensional model builders using 639 X-ray structures. *J. Chem. Inf. Comp. Sci.* **34**, 1000–1008 (1994).
68. Hedwig, G. R. & Hinz, H.-J. Group additivity schemes for the calculation of the partial molar heat capacities and volumes of unfolded proteins in aqueous solution. *Biophys. Chem.* **100**, 239–260 (2003).
69. Mishra, A. & Ahluwalia, J. Apparent molal volumes of amino acids, N-acetyl amino acids, and peptides in aqueous solutions. *J. Phys. Chem.* **88**, 86–92 (1984).
70. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
71. Chen, C. R. & Makhatadze, G. I. ProteinVolume: calculating molecular van der Waals and void volumes in proteins. *BMC Bioinformatics* **16**, 101 (2015).
72. Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* **68**, 441–443 (1964).
73. Bush, J. & Makhatadze, G. I. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins* **79**, 2027–2032 (2011).
74. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537–W541 (2012).
75. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl Acad. Sci. USA* **102**, 12690–12693 (2004).

## Acknowledgements

We thank Drs Angel Garcia (RPI), Cathy Royer (RPI) and Tigran Chalikian (U Toronto) for countless discussions of volume changes upon protein unfolding, Dr Tobin Sosnick (U Chicago) for help with Statistical Coil and Dr Nick Fitzkee (Mississippi State) for providing unfolded state ensembles generated according to Fitzkee and Rose<sup>35</sup>. This work was supported by grants CHE-1145407/1506468 from the US National Science Foundation. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575, and Center for Computational Innovations at Rensselaer Polytechnic Institute.

## Author contributions

C.R.C. and G.I.M. conceived the project, analysed the data and wrote the manuscript. Both authors discussed the results and contributed to the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

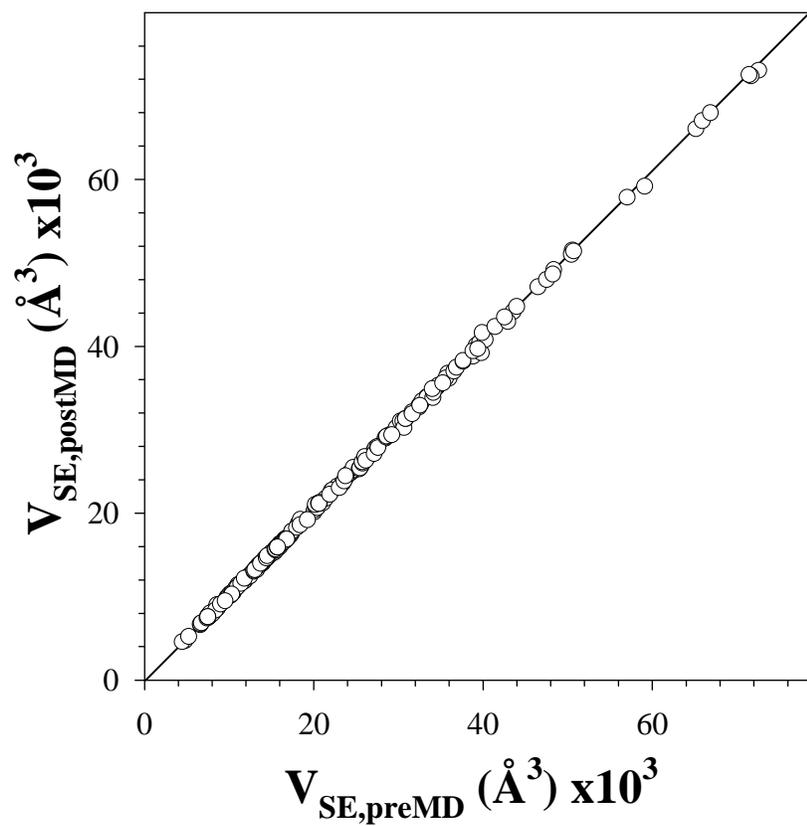
**How to cite this article:** Chen, C. R. & Makhatadze, G. I. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat. Commun.* **8**, 14561 doi: 10.1038/ncomms14561 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

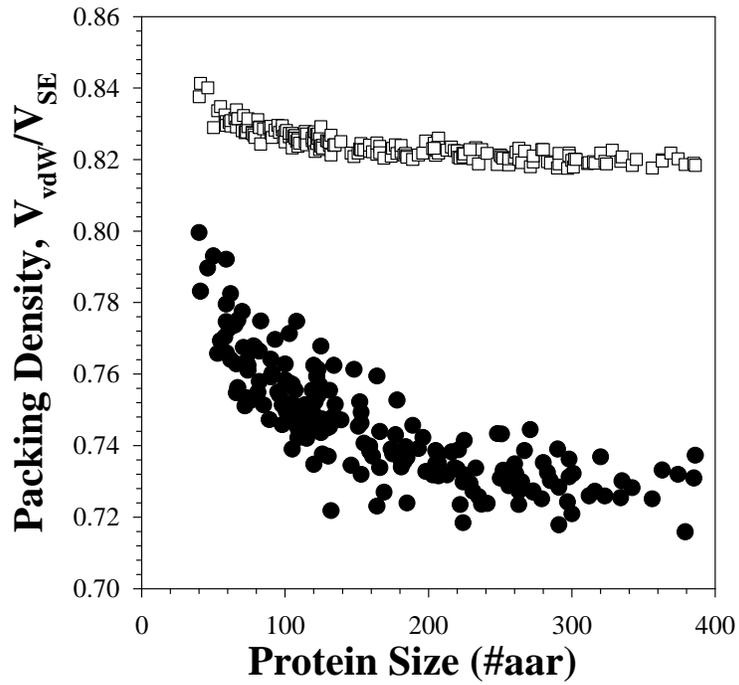


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

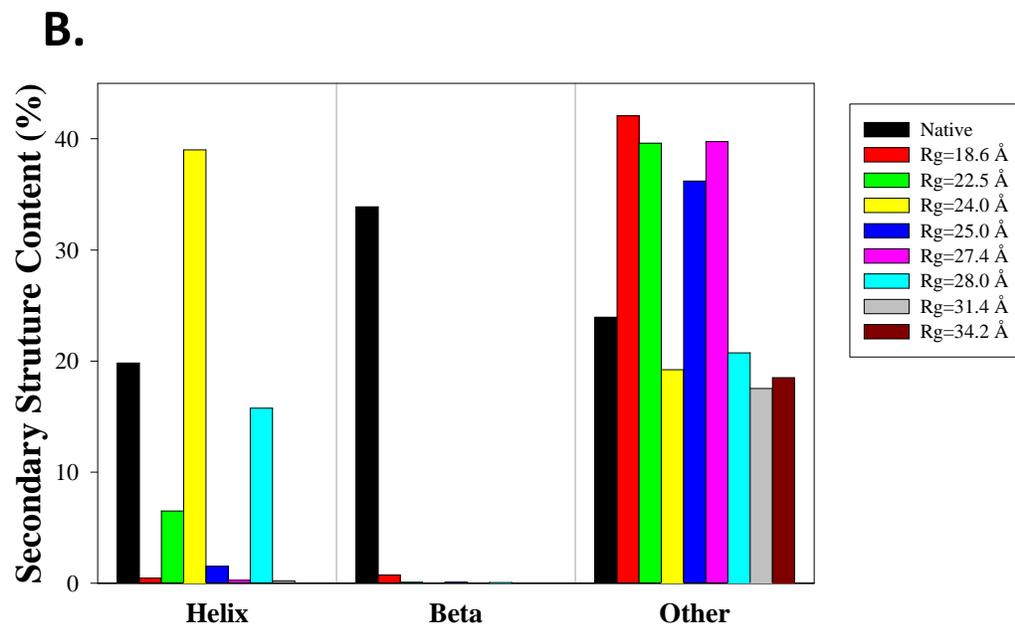
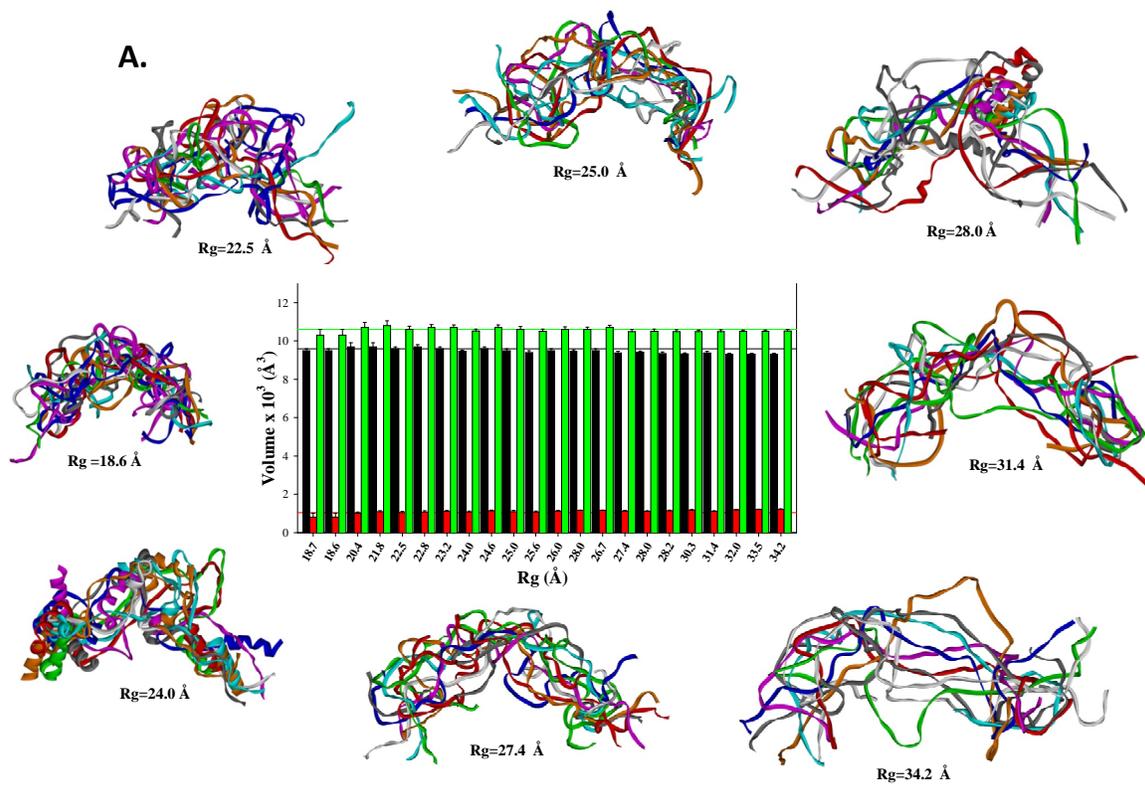
© The Author(s) 2017

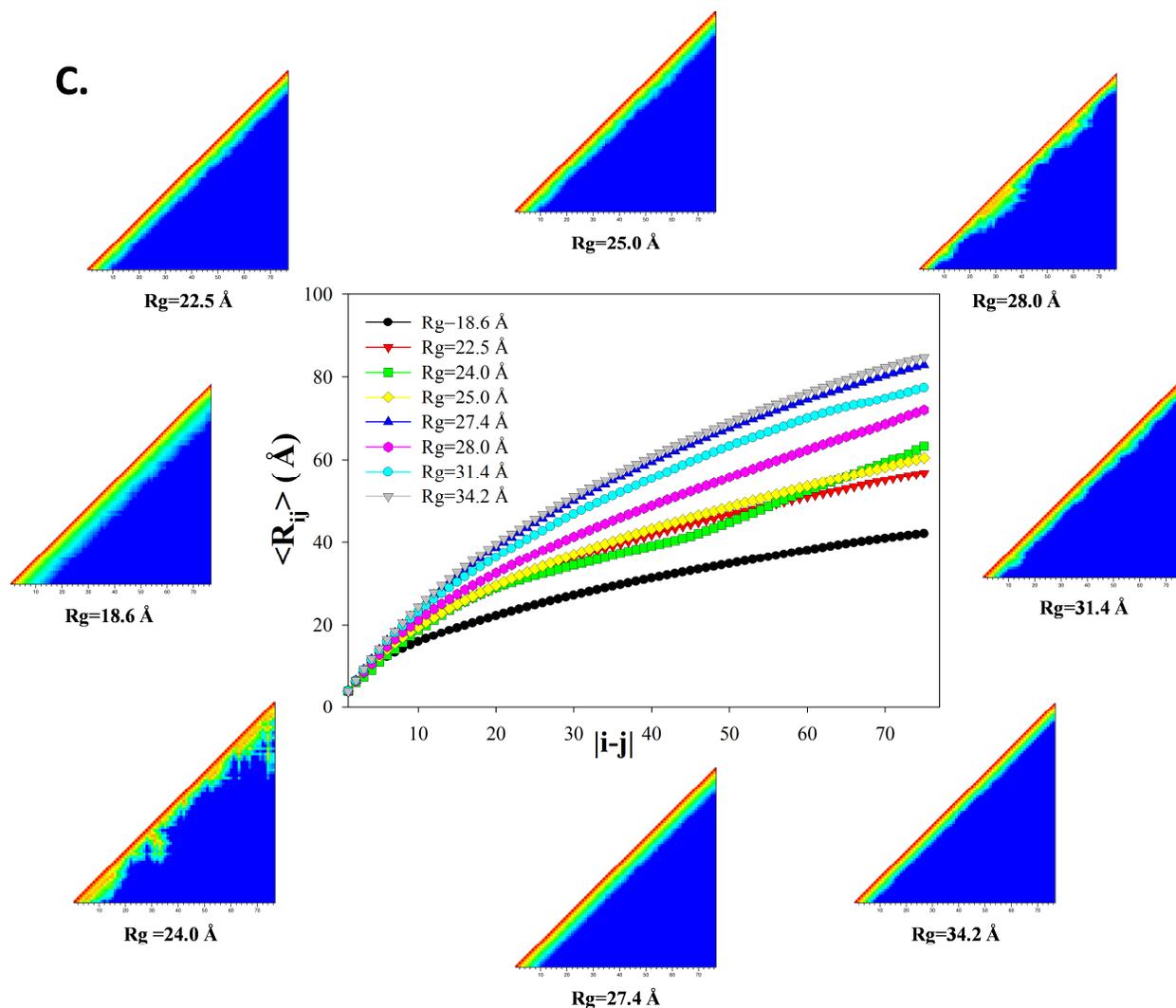


**Supplementary Figure 1.** Comparison of the solvent-excluded volumes of proteins in the native state based on the x-ray structure ( $V_{SE,preMD}$ ) with the solvent-excluded volumes calculated using native state ensemble obtained using MD simulations ( $V_{SE,postMD}$ ). The standard deviations of  $V_{SE,postMD}$  values are smaller than the symbol size.



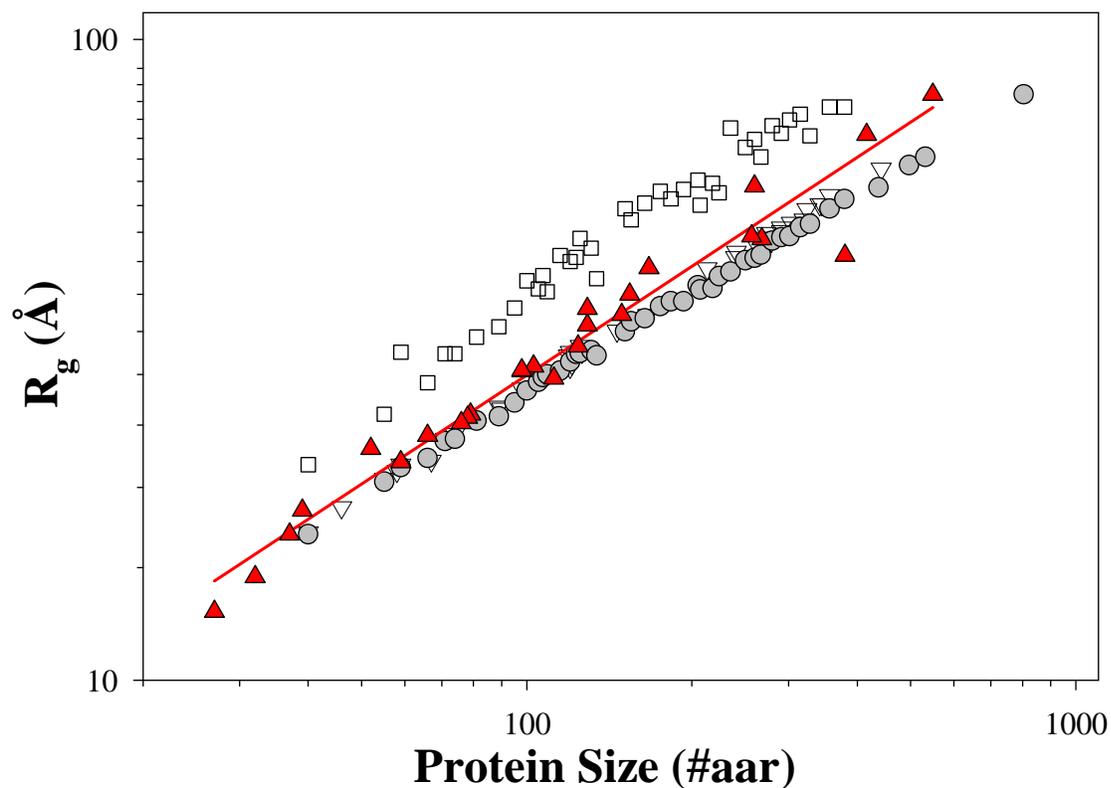
**Supplementary Figure 2.** Comparison of packing densities of the native (circles) and unfolded (squares) state ensembles. Packing density is defined as the ratio of the van der Waals volume to solvent-excluded volume.



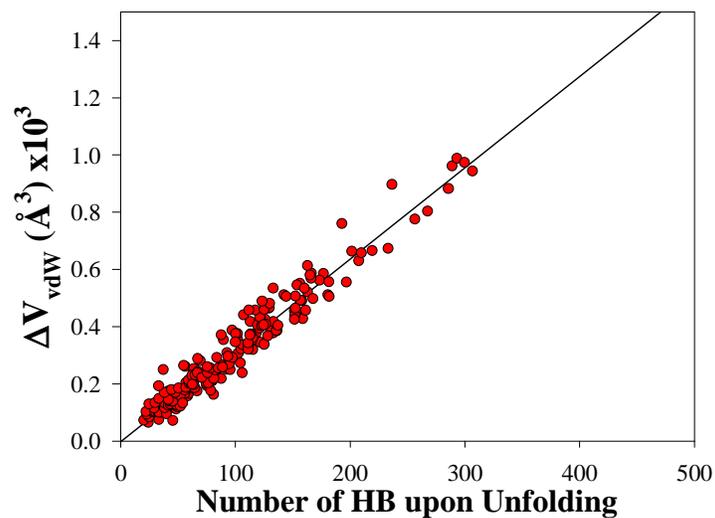


**Supplementary Figure 3.** Characterization of various unfolded state ensembles generated using TraDES<sup>1</sup> ( $R_g=22.5$  Å), flexible-meccano<sup>2</sup> (FM,  $R_g=25.0$  Å), Statistical Coil<sup>3</sup> (SC,  $R_g=31.4$  Å), Fitzkee & Rose<sup>4</sup> approach ( $R_g=24.0$ ), Fitzkee & Rose-like approach ( $R_g=28.0$  Å), or CAMPARI<sup>5,6</sup> ( $R_g=18.6$ , 27.4 and 34.2 Å). The listed CAMPARI simulations at  $R_g=18.6$  Å and 34.2 Å nm were generated by sampling the Theta solvent and EV (excluded volume) ensembles, respectively. The Theta solvent ensemble was generated by turning off attractive and repulsive LJ forces and randomly sampling backbone dihedral angles from grid files for each respective amino acid type. The EV ensemble was generated by setting attractive LJ force to 0 and repulsive LJ force to 1. The simulation at  $R_g=27.4$  Å was generated by setting the attractive LJ force to 0 and repulsive LJ force to 0.001. The Fitzkee/Rose-like simulations were carried out by generating an all-atom structure-based model from the crystal structure of using SMOGv2<sup>7,8</sup>. All pairwise interactions were removed so the Hamiltonian has only parameters for bonds, angles, and dihedral angles remaining. Residues with secondary structure were identified in the native

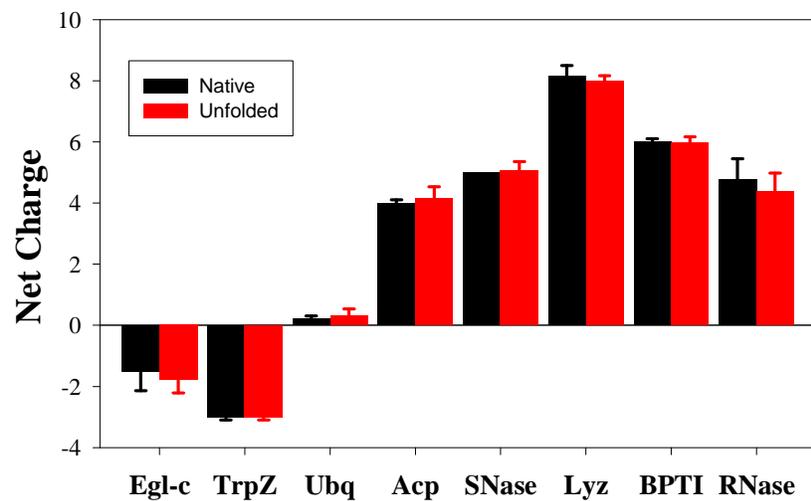
crystal structure and had their respective dihedral angles fixed in the structure-based model. The simulation was then run at 120K to generate an unfolded state that still maintained its native secondary structure. Data, shown for illustrative purposes, was computed from the conformational ensembles of ubiquitin (PDB:1UBQ). **Panel A** shows ribbon traces of 10 representative structures from each structural ensemble with the corresponding ensemble average Rg. Central insert shows a comparison of the volumes for these ensembles.  $V_{SE}$  (black bars)  $V_{Hyd}$  (red bars) and  $V_{Tot}$  (green bars). In addition, volumes of other ensembles, generated using CAMPARI, with intermittent Rg are also shown. **Panel B**. Comparison of secondary structure content in the unfolded state ensembles. Secondary structure content was calculated using DSSP 2.2.1<sup>9</sup>. Helical structure is the sum of 3-, 4-, and 5-turn helices, sheet structure is the sum of  $\beta$ -bridges and  $\beta$ -sheets, and other includes turn and bend. **Panel C**. Comparison of contact maps of the unfolded ensembles calculated as averaged distance matrices consisting of the smallest distance between residue pairs generated using the GROMACS utility g\_mdmat. Central insert shown fractal dimensions of the unfolded state ensembles calculated as the average distance  $\langle R_{ij} \rangle$  between C $\alpha$  atoms of residues  $i$  and  $j$  as a function of sequence separation  $|i-j|$ <sup>10</sup>.



**Supplementary Figure 4.** TraDES and FM generated unfolded state ensembles show a dependence of radius of gyration ( $R_g$ ) on protein size similar to experimentally measured values<sup>11</sup>. Red triangles show the experimentally measured (using SAXS) values of  $R_g$  for proteins of various sizes. Open squares show the  $R_g$  values calculated for the SC-generated ensemble, while gray circles show the values calculated for the TraDES ensemble, and open triangles show the results for the FM-generated ensemble. For clarity only every fifth data point is shown.



**Supplementary Figure 5.** Difference in the van der Waals volumes of native and unfolded state ensembles ( $\Delta V_{\text{vdW}}$ ) is due to the larger number of hydrogen bonds in the native relative to the unfolded state ensemble. Dependence of  $\Delta V_{\text{vdW}}$  on the average change in the number of hydrogen bonds upon unfolding is linear with a Pearson correlation coefficient of 0.96. The average number of hydrogen bonds for each ensemble was calculated using DSSP 2.2.1<sup>9</sup>.



**Supplementary Figure 6.** Comparison of the net charge of the native and unfolded state ensembles calculated using h++ server<sup>12, 13</sup>. Error bars are the standard deviations of the mean.

## References

1. Feldman H, Hogue C. Probabilistic sampling of protein conformations: New hope for brute force? *Proteins* **23**, 8-23 (2002).
2. Ozenne V, *et al.* Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463-1470 (2012).
3. Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. *P Natl Acad Sci USA* **102**, 13099-13104 (2005).
4. Fitzkee N, Rose G. Reassessing random-coil statistics in unfolded proteins. *P Natl Acad Sci USA* **101**, 12497-12502 (2004).
5. Vitalis A, Pappu RV. Methods for Monte Carlo Simulations of Biomacromolecules. *Annu Rep Comput Chem* **5**, 49-76 (2009).
6. Vitalis A, Pappu RV. ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J Comput Chem* **30**, 673-699 (2009).
7. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75**, 430-441 (2009).
8. Noel JK, *et al.* SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput Biol* **12**, e1004794 (2016).
9. Joosten RP, *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res* **39**, D411-419 (2011).
10. Tran HT, Wang X, Pappu RV. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* **44**, 11369-11380 (2005).
11. Kohn JE, *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *P Natl Acad Sci USA* **102**, 12690-12693 (2004).

12. Anandakrishnan R, Aguilar B, Onufriev AV. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* **40**, W537-541 (2012).
13. Chan CH, Wilbanks CC, Makhatadze GI, Wong KB. Electrostatic contribution of surface charge residues to the stability of a thermophilic protein: benchmarking experimental and predicted pKa values. *PLoS one* **7**, e30296 (2012).
14. Lee S, Tikhomirova A, Shalvardjian N, Chalikian TV. Partial molar volumes and adiabatic compressibilities of unfolded protein states. *Biophys Chem* **134**, 185-199 (2008).

## Supplementary Table 1

**Ultra-High Resolution Protein Set (0.73 - 1.20 Å): First 4 letters are PDB code, fifth letter is chain id. In parenthesis - number of amino acid residues and crystallographic resolution in Å.**

**2ERL\_ (40, 1.00); 1P9GA (41, 0.84); 1CNR\_ (46, 1.05); 2A26A (50, 1.20); 1BRFA (53, 0.95); 2CS7C (55, 1.20); 1G6XA (58, 0.86); 1OAIA (59, 1.00); 2FMAA (59, 0.85); 2G6FX (59, 0.92); 1NKD\_ (59, 1.07); 2IGD\_ (61, 1.10); 1G2BA (62, 1.12); 1V6PA (62, 0.87); 2SN3\_ (65, 1.20); 1C9OA (66, 1.17); 1HG7A (66, 1.15); 1TUKA (67, 1.12); 1VFYA (67, 1.15); 2DLBA (70, 1.20); 2B97A (71, 0.75); 1WM3A (72, 1.20); 1WXCBA (72, 1.20); 1CC8A (73, 1.02); 1I27A (73, 1.02); 1L9LA (74, 0.92); 1OK0A (74, 0.93); 2BWFB (77, 1.15); 1USMA (77, 1.20); 1UCRB (78, 1.20); 1XMKA (79, 0.97); 1IQZA (81, 0.92); 1R6JA (82, 0.73); 1ZZKA (82, 0.95); 2D8DB (83, 1.15); 1B0YA (85, 0.93); 1CTJ\_ (89, 1.10); 1U07A (90, 1.13); 2BT9A (90, 0.94); 1X6IB (91, 1.20); 2FHZB (93, 1.15); 1C5EA (95, 1.10); 1LNIB (96, 1.00); 1CZPA (98, 1.17); 2AIBA (98, 1.10); 1NQJA (98, 1.00); 1KZKB (99, 1.09); 1MN8D (100, 1.00); 1PSRB (100, 1.05); 1M2DA (101, 1.05); 2DKOB (103, 1.06); 2H3LA (103, 1.00); 1LKKA (105, 1.00); 1TQGA (105, 0.98); 2GBAA (105, 0.92); 1M9ZA (105, 1.05); 2FRGP (106, 1.19); 1V8HA (107, 1.20); 1GMXA (108, 1.10); 1J0PA (108, 0.91); 2AGYD (108, 1.10); 1BKRA (109, 1.10); 1RWYA (109, 1.05); 2FHZA (109, 1.15); 1H4XA (110, 1.16); 1I8OA (114, 1.15); 2CHHA (114, 1.00); 1F86A (115, 1.10); 1SAUA (115, 1.12); 1O7IA (119, 1.20); 2ICCA (119, 1.20); 2F01B (120, 0.85); 1W0NA (120, 0.80); 1VR7A (120, 1.20); 1WN2A (121, 1.20); 2GUDB (121, 0.94); 1LWBA (122, 1.05); 2FWGA (122, 1.10); 1VL9A (123, 0.97); 1DY5A (123, 0.87); 1GU2A (124, 1.19); 1UNQA (124, 0.98); 1NWZA (125, 0.82); 2FJ8A (125, 1.19); 1VZIA (126, 1.15); 1JBEA (126, 1.08); 4LZT\_ (129, 0.95); 1KNLA (130, 1.20); 1JF8A (131, 1.12); 1OH0B (131, 1.10); 1C7KA (132, 1.00); 1IFC\_ (132, 1.19); 1TU9A (134, 1.20); 2AXWA (134, 1.05); 1NKIA (135, 0.95); 1CZ9A (139, 1.20); 1RG8A (146, 1.10); 1EXRA (148, 1.00); 1A6M\_ (151, 1.00); 1QTNA (152, 1.20); 1GWMA (153, 1.15); 2C9VA (153, 1.07); 1J98A (153, 1.20); 2FLHB (155, 1.20); 1UOWA (159, 1.04); 1Y93A (159, 1.03); 1P6OB (161, 1.14); 1L3KA (163, 1.10); 1TT8A (164, 1.00); 1WKQA (164, 1.17); 1N62A (166, 1.09); 2CE2X (166, 1.00); 1OBOA (169, 1.20); 1AMM\_ (174, 1.20); 2AU7A (175, 1.05); 1EB6A (177, 1.00); 2C2UA (178, 1.10); 1I4UA (181, 1.15); 1WC2A (181, 1.20); 1KT6A (183, 1.10); 2AT7X (184, 0.98); 1PMHX (185, 1.06); 1QV0A (185, 1.10); 2BBRA (189, 1.20); 2PTH\_ (193, 1.20); 2CARA (196, 1.09); 1QQ4A (198, 1.20); 1Z0WA (203, 1.20); 1JM1A (204, 1.11); 1IXBA (205, 0.90); 2AB0A (205, 1.10); 2C71A (205, 1.05); 1HDOA (206, 1.15); 1G66A (207, 0.90); 1H4GB (207, 1.10); 1KWNA (207, 1.20); 1SFSA (213, 1.07); 1ME3A (215, 1.20); 1K4IA (216, 0.98); 1W66A (218, 1.08); 1FYEA (220, 1.20); 1O08A (221, 1.20); 2A6ZA (222, 1.00); 1OLRA (224, 1.20); 1UAIA (224, 1.20); 2AWKA (224, 1.15); 1KG2A (225, 1.20); 1FSGC (233, 1.05); 1K7CA (233, 1.12); 1YMTA (235, 1.20); 1JBC\_ (237, 1.20); 1GVKB (240, 0.94); 1QL0A (241, 1.10); 1HBNC (248, 1.16); 2J27A (250, 1.15); 1ZJYA (251, 1.05); 1QXYA (252, 1.04); 1XQOA (256, 1.03); 1MOOA (256, 1.05); 1P1XA (260, 0.99); 1UWCA (261, 1.08); 1NYMA (263, 1.20); 1ARB\_ (263, 1.20); 1XDNA (265, 1.20); 1WUIS (267, 1.04); 1KQPA (271, 1.03); 2CI1A (273, 1.08); 1WXCA (273, 1.20); 1IC6A (279, 0.98); 2BOGX (280, 1.04); 1E9GB (283, 1.15); 1QTWA (285, 1.02); 1LC0A (290, 1.20); 2EUTA (291, 1.12); 1RTQA (291, 0.95); 2J45B (297, 1.14); 2CIWA (298, 1.15); 2BLNA (298, 1.20); 8A3HA (300, 0.97); 2CNQA (301, 1.00); 1V0LA (302, 0.98); 1ZL0B (303, 1.10); 1Z2NX (311, 1.20); 2IAVA (312, 1.07); 1T2DA (315, 1.10); 1PWMA (316, 0.92); 1YS1X (320, 1.10); 1DS1A (323, 1.08); 1RYOA (324, 1.20); 1OEWA (328, 0.90); 2BW4A (334, 0.90); 2C1VA (335, 1.20); 1YFQA (342, 1.10); 1YQSA (345, 1.05); 1M15A (356, 1.20); 1C0PA (363, 1.20); 1VYRA (363, 0.90); 1GA6A (369, 1.00); 1N8KA (374, 1.13); 3SIL\_ (379, 1.05); 1KJQB (385,**

1.05); **1MUWA** (386, 0.86); **1RA0A** (423, 1.12); **1UG6A** (426, 0.99); **2BMOA** (437, 1.20); **1HBNB** (442, 1.16); **2BF6A** (449, 0.97); **1M1NA** (477, 1.16); **2FBAA** (492, 1.10); **1QW9A** (497, 1.20); **1GWEA** (498, 0.88); **1JETA** (517, 1.20); **1M1NB** (522, 1.16); **1Q6ZA** (524, 1.00); **1WUIL** (532, 1.04); **1HBNA** (543, 1.16); **1UWKB** (553, 1.19); **2BHUA** (580, 1.10); **1SU8A** (633, 1.10); **1N62B** (804, 1.09); **1QWNA** (1014, 1.20).

## Supplementary Table 2

### Volumes of Ionization for Various protein Groups

Protein Group	Volume ( $\text{\AA}^3$ ) <sup>1</sup>
N terminus	-8.0
Arg	-9.1
Lys	-9.8
His	-2.3
Asp	-20.1
Glu	-19.6
C terminus	-17.4

<sup>1</sup> Volume changes are for the reactions  $AH \rightarrow A^- + H^+$  and  $B + H^+ \rightarrow BH^+$  where A is the acid and B is the base. The volumes of ionization can be used to predict the volumes of a protein in the native or unfolded state by adding the following term to equation 5 in the main text:

$$V_{el} = f_{Nter} \cdot V_{Nter} + f_{Cter} \cdot V_{Cter} + \sum_{i=D,E,R,K,H} f_i \cdot N_i \cdot V_i$$

where  $N_i$  is the number of a given type of ionizable group,  $f_i$  is fraction exposed, and  $V_i$  is the ionization volume<sup>14</sup>.