Published in final edited form as:

Biophys Chem. 2017 December ; 231: 135–145. doi:10.1016/j.bpc.2017.04.006.

Simulated pressure denaturation thermodynamics of ubiquitin

Elizabeth A. Ploetz¹ and Paul E. Smith

Department of Chemistry, 213 CBC Building, 1212 Mid Campus Dr. North, Kansas State University, Manhattan, KS 66506-0401, United States

Abstract

Simulations of protein thermodynamics are generally difficult to perform and provide limited information. It is desirable to increase the degree of detail provided by simulation and thereby the potential insight into the thermodynamic properties of proteins. In this study, we outline how to analyze simulation trajectories to decompose conformation-specific, parameter free, thermodynamically defined protein volumes into residue-based contributions. The total volumes are obtained using established methods from Fluctuation Solution Theory, while the volume decomposition is new and is performed using a simple proximity method. Native and fully extended ubiquitin are used as the test conformations. Changes in the protein volumes are then followed as a function of pressure, allowing for conformation-specific protein compressibility values to also be obtained. Residue volume and compressibility values indicate significant contributions to protein denaturation thermodynamics from nonpolar and coil residues, together with a general negative compressibility exhibited by acidic residues.

Keywords

Kirkwood-Buff theory; Molecular dynamics simulations; Statistical mechanics; Solvation; Protein folding; Denatured state ensemble

1. Introduction

1.1. Background

Understanding protein folding is one of the great unsolved problems of modern science. Apart from being a topic that fascinates many, it is likely crucial that it be understood to address various devastating diseases, e.g. protein misfolding diseases like Alzheimer's disease, as well as to increase the success of protein engineering efforts. We are specifically working to develop new, rigorous tools to help understand conformation-specific protein thermodynamics using computer simulation approaches. Hopefully, these can be used to address a well-established complexity in understanding protein folding; namely, how the

Correspondence to: Paul E. Smith.

¹Present address: Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205.

Appendix A. Supplementary data: One table and five figures. Decomposition of the protein into physicochemical groups and secondary structure elements. Details of the Hill/Schnell approach used to obtain the residue volumes. Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.bpc.2017.04.006>.

denatured state ensemble (DSE) differs with the manner of denaturation, i.e. the use of temperature, pressure, or chemical denaturants [1].

Experimental data is bountiful and indispensable, but it typically lacks atomic detail. Computer simulation approaches are promising due to their exquisite detail, but the quality of the force fields and the sufficiency of sampling achieved is typically a concern. However, even assuming both these simulation issues were resolved (not addressed here), a third issue still remains: Are methodologies available to extract rigorous thermodynamic properties from simulation trajectories in a way that provides more insight than the equivalent experiment would have provided?

As we have recently reviewed [2], most previous approaches to obtain the thermodynamic properties of a protein are subjective. The enhanced sampling technique of replica exchange molecular dynamics (REMD) is an exception, because it does provide rigorous thermodynamic properties [3]. However, it is difficult to assign the properties obtained from REMD to specific conformations, and it is not possible to look at their distance-dependent properties or the contributions to the total protein properties from individual groups, e.g. individual residues or backbone vs. sidechain contributions. Therefore, the values obtained from REMD studies are useful when making comparisons with experimental properties, but are essentially providing little additional detail beyond the amount and type of information that the equivalent experiment would provide. The apparent molar approach that we have recently put forth is also rigorous, requires fewer simulations than would be necessary for the REMD method, and provides conformation-specific properties, but it also provides no distance-dependent or residue-specific information [2].

We have started to address this issue by establishing a way to extract the relevant, conformation-specific thermodynamic properties (infinitely dilute partial molar protein volume, isothermal compressibility, enthalpy, heat capacity, and thermal expansion) from single simulation trajectories, using Fluctuation Solution Theory (FST) [4–6]. The FST method for the volume had been in existence previously [4–5,7–13], using the traditional Kirkwood-Buff (KB) theory formalism [14]. It is our opinion that the KB/FST approaches hold the most promise as they represent exact theories of solution mixtures. Both formalisms provide expressions for thermodynamic properties of mixtures, but they are written either in terms of a series of integrals over molecular correlation functions (the KB approach), or in terms of the equivalent particle number fluctuations corresponding to local regions of the solution (the FST approach). The theories are described in detail elsewhere [4–6,13]. They have been used to successfully investigate thermodynamic properties of solution mixtures and protein denaturation by hydrostatic pressure and cosolvents, as well as to help clarify traditional osmotic stress analysis [4,6,13].

In this work, we focus exclusively on pressure denaturation using only two conformations of a single protein, as a proof of principle. We determine the protein volume, (additive) residue-based contributions to the volume, and pressure derivatives of these residue volumes to obtain residue-based isothermal compressibilities. While the decomposition of the total protein properties to residue properties is not unique, and is not thermodynamically defined, the sum of the contributions is the formally defined thermodynamic property [15]. Indeed,

the justification for such a decomposition approach can only be provided by the additional insight gained in the process. This study provides an initial step in this direction. The residue-decomposition portion of our approach is similar to that of Garcia and of Baynes and Trout, who used the simple “proximity criterion” of Beveridge [16] to obtain backbone vs. sidechain or residue-based contributions to the preferential interactions of proteins with cosolvents [17-19]. We also show how the FST protein volume expression avoids the complication of having to integrate the distribution around an irregularly shaped object, such as a protein. In the past, this problem has limited the application of the traditional KB expression for the infinitely dilute partial molar volume [10].

1.2. Thermodynamics

The thermodynamics of solute-solvent mixtures and protein denaturation is well established. However, in our opinion, the link between experiment and simulation requires some clarification. The aim of this section is to outline, in an exact manner, the relationships between the experimental quantities of interest and the equivalent statistical mechanical expressions that can then be used to provide an atomic level description using simulation data. To achieve this goal, we will adopt the pseudo chemical potential approach of Ben-Naim, the virtues of which have been presented elsewhere [5-6,20].

The chemical potential (μ_2) of a biomolecular solute (2) in a solvent (1) can be written,

$$\begin{aligned}\mu_2 &= \mu_{2,m}^\ominus + RT \ln(\gamma_{2,m} m_2 / m^\ominus) \\ \mu_2 &= \mu_{2,c}^\ominus + RT \ln(\gamma_{2,c} c_2 / c^\ominus) \\ \mu_2 &= \mu_2^* + RT \ln(\Lambda_2^3 \rho_2)\end{aligned}\quad (1)$$

where R is the gas constant and T the absolute temperature. The first two expressions in Eq. (1) correspond to two possible experimental approaches in terms of a standard chemical potential on the molal ($\mu_{2,m}^\ominus$) or molar ($\mu_{2,c}^\ominus$) concentration scales, using a molal ($\gamma_{2,m}$) or molar ($\gamma_{2,c}$) activity coefficient, the solute molality (m_2) or molarity (c_2), and the standard molality ($m^\ominus = 1$ mol/kg) or molarity ($c^\ominus = 1$ mol/L), respectively. The third expression can be considered as the defining statistical mechanical relationship for the pseudo chemical potential (μ_2^*) in terms of the thermal de Broglie wavelength of the solute (Λ_2), which can be treated as a constant here, and the solute number density ($\rho_2 = N_2/V$). This latter expression has been extensively used by Ben-Naim to probe a variety of solvation phenomena [20]. The pseudo chemical potential corresponds to the change in Gibbs free energy on introduction of a single solute molecule at a *fixed* point in space within the system. To understand pressure denaturation at a fixed temperature we need to follow changes in the chemical potentials with pressure. The pressure derivatives of Eq. (1) provide,

$$\begin{aligned}\bar{V}_2 &\equiv \left(\frac{\partial \mu_2}{\partial p}\right)_{m_2} = \left(\frac{\partial \mu_{2,m}^\Theta}{\partial p}\right)_{m_2} + RT \left(\frac{\partial \ln J_{2,m}}{\partial p}\right)_{m_2} \\ \bar{V}_2 &= \left(\frac{\partial \mu_{2,c}^\Theta}{\partial p}\right)_{m_2} + RT \left(\frac{\partial \ln J_{2,c}}{\partial p}\right)_{m_2} + RT \kappa_T \\ \bar{V}_2 &= \left(\frac{\partial \mu_2^*}{\partial p}\right)_{m_2} + RT \kappa_T\end{aligned}\quad (2)$$

where $\bar{V}_2$ is the partial molar volume of the solute and κ_T is the isothermal compressibility of the solution. For an infinitely dilute solute in the un-symmetric convention we require the activity coefficients to be unity for all pressures. Hence, the activity derivatives disappear, the compressibility becomes that of the pure solvent ($\kappa_{T,1}^0$), and we then have,

$$\bar{V}_2^\infty = V_{2,m}^\Theta = V_{2,c}^\Theta + RT \kappa_{T,1}^0 = V_2^{*,\infty} + RT \kappa_{T,1}^0 \quad (3)$$

Consequently, derivatives of the standard and pseudo chemical potentials provide standard and pseudo volumes that are equal to the partial molar volume to within a constant. Taking another pressure derivative provides,

$$\left(\frac{\partial \bar{V}_2^\infty}{\partial p}\right)_{m_2} = \left(\frac{\partial V_{2,m}^\Theta}{\partial p}\right)_{m_2} = \left(\frac{\partial V_{2,c}^\Theta}{\partial p}\right)_{m_2} + RT \left(\frac{\partial \kappa_{T,1}^0}{\partial p}\right)_{m_2} = \left(\frac{\partial V_2^\infty}{\partial p}\right)_{m_2} + RT \left(\frac{\partial \kappa_{T,1}^0}{\partial p}\right)_{m_2} \quad (4)$$

and relates the pressure derivatives of the standard and pseudo volumes to the partial molar compressibility at infinite dilution. It should be noted that the difference between the partial, standard, and pseudo volumes is typically small, since $RT \kappa_{T,1}^0 = 1\text{-}2 \text{ cm}^3/\text{mol}$ for water under ambient conditions.

If we now turn to the common situation of a two state protein denaturation problem involving a transition from a native (N) state to a denatured (D) state ($N \rightarrow D$), and consider the equilibrium constant $K = N_D/N_N$ for an infinitely dilute solute ($N_2 = N_D + N_N$), then from Eq. (1) and the equilibrium condition ($\mu_D = \mu_N = \mu_2$) we find,

$$-RT \ln K = \mu_{D,m}^\Theta - \mu_{N,m}^\Theta = \mu_{D,c}^\Theta - \mu_{N,c}^\Theta = \mu_D^{*,\infty} - \mu_D^{*,\infty} = \Delta G^\Theta \quad (5)$$

The effect of pressure on this equilibrium is typically expressed in terms of a Taylor expansion around a reference pressure (p_0), usually 1 bar, where the equilibrium constant is K_0 ,

$$RT \ln \left(\frac{K}{K_0} \right) = RT \left(\frac{\partial \ln K}{\partial p} \right)_{m_2} \Delta p + \frac{RT}{2} \left(\frac{\partial^2 \ln K}{\partial p^2} \right)_{m_2} (\Delta p)^2 \quad (6)$$

and $p = p - p_0$. Hence, the above pressure derivatives characterize the effects of hydrostatic pressure on the equilibrium in terms of properties of the native and denatured state evaluated at the reference pressure. Using Eq. (5) the above derivatives can be expressed as,

$$\begin{aligned} -RT \left(\frac{\partial \ln K}{\partial p} \right)_{m_2} &= \left(\frac{\partial \mu_D^{*,\infty}}{\partial p} \right)_{m_2} - \left(\frac{\partial \mu_N^{*,\infty}}{\partial p} \right)_{m_2} = \Delta V_2^{*,\infty} \\ -RT \left(\frac{\partial^2 \ln K}{\partial p^2} \right)_{m_2} &= \left(\frac{\partial^2 \mu_D^{*,\infty}}{\partial p^2} \right)_{m_2} - \left(\frac{\partial^2 \mu_N^{*,\infty}}{\partial p^2} \right)_{m_2} = -\Delta K_2^{*,\infty} \end{aligned} \quad (7)$$

Similar expressions can be written using the standard chemical potentials. At this stage, most approaches simply use Eqs. (2)–(4) to express the central derivatives in Eq. (7) in terms of the difference in volumes, $V_2^{*,\infty}$, and compressibility factors (sometimes referred to as partial molar compressibilities), $K_2^{*,\infty}$, of the two forms. Technically, this is incorrect as the two forms are not thermodynamically independent and the results from Eqs. (3) and (4) cannot be directly applied to the above equilibrium. For instance, the equilibrium condition provides,

$$\left(\frac{\partial \mu_2}{\partial p} \right)_{m_2} = \left(\frac{\partial \mu_D}{\partial p} \right)_{m_2} = \left(\frac{\partial \mu_N}{\partial p} \right)_{m_2} \quad (8)$$

and a similar logical use of Eqs. (2)–(4) in this expression then suggests the partial molar volumes of each form would have to be the same, in contrast to the result expected from Eq. (7). Clearly, the above expressions involve some subtle issues. As we will be using an exact theory of solutions during the current analysis it is therefore important to be as thorough and rigorous as possible. The present analysis essentially follows the approach outlined by Ben-Naim [21].

The pressure derivatives of Eq. (5) that lead to Eq. (7) are performed at constant T and composition (N_1 and N_2). However, true partial molar quantities require T , N_1 , N_N and N_D to be constant. Fortunately, while these derivatives are not the same, the relationship between them is known and is given by,

$$\left(\frac{\partial X}{\partial p} \right)_{m_2} = \left(\frac{\partial X}{\partial p} \right)_{m_N, m_D} + \left[f_N \left(\frac{\partial X}{\partial \ln m_D} \right)_{p, m_N} - f_D \left(\frac{\partial X}{\partial \ln m_N} \right)_{p, m_D} \right] \left(\frac{\partial \ln K}{\partial p} \right)_{m_2} \quad (9)$$

for a quantity $X = \mu_D, \mu_D^*, V_D^*, \text{etc.}$ at constant T. Here, $f_D = N_D/N_2$ is the fraction of denatured protein and we have used the fact that,

$$\left(\frac{\partial f_D}{\partial p}\right)_{m_2} = -\left(\frac{\partial f_N}{\partial p}\right)_{m_2} = f_D f_N \left(\frac{\partial \ln K}{\partial p}\right)_{m_2} \quad (10)$$

Eq. (9) is valid at any solute concentration as long as the equilibrium constant is measured by concentrations and not activities. The above expression indicates that the derivatives at constant m_2 and at constant m_N and m_D are not equal, but are related through a series of additional derivatives. These derivatives are not available experimentally, although they could be obtained from simulation if necessary.

Fortunately, the additional derivatives in Eq. (9) take a simple form when m_2 (and therefore m_N and m_D) are present at infinite dilution. If we set $X = \mu_D^*$, then Eq. (9) provides,

$$\begin{aligned} \left(\frac{\partial \mu_D^{*,\infty}}{\partial p}\right)_{m_2} &= \left(\frac{\partial \mu_D^{*,\infty}}{\partial p}\right)_{m_D, m_N} = V_D^{*,\infty} \\ \left(\frac{\partial^2 \mu_D^{*,\infty}}{\partial p^2}\right)_{m_2} &= \left(\frac{\partial V_D^{*,\infty}}{\partial p^2}\right)_{m_2} = \left(\frac{\partial V_D^{*,\infty}}{\partial p}\right)_{m_D, m_N} = -K_D^{*,\infty} \end{aligned} \quad (11)$$

and similarly for the native state. Hence, the expected results are obtained for Eq. (7). But, if we set $X = \mu_D$ or μ_N and use the fact that $(\mu_\alpha / \ln m_\beta)_p = RT \delta_{\alpha\beta}$ as $m_\beta \rightarrow 0$, with $\delta_{\alpha\beta}$ being the Kronecker delta function, then Eq. (9) gives,

$$\begin{aligned} \left(\frac{\partial \mu_D^\infty}{\partial p}\right)_{m_2} &= \bar{V}_D^\infty + f_N RT \left(\frac{\partial \ln K}{\partial p}\right)_{m_2} \\ \left(\frac{\partial \mu_N^\infty}{\partial p}\right)_{m_2} &= \bar{V}_N^\infty - f_D RT \left(\frac{\partial \ln K}{\partial p}\right)_{m_2} \end{aligned} \quad (12)$$

Consequently, the derivatives on the left hand side are not the formal partial molar volumes of the denatured and native states, even at infinite dilution, unless the pressure dependence of the equilibrium is negligible. Using Eq. (12) with the equilibrium condition ($\mu_D = \mu_N$) provides the relatively simple relation,

$$-RT \left(\frac{\partial \ln K}{\partial p}\right)_{m_2} = \bar{V}_D^\infty - \bar{V}_N^\infty \quad (13)$$

The above derivative is also given by the difference in the pseudo volumes, as indicated by Eq. (3) [4]. Hence, the appropriate derivative of the equilibrium constant is given by the

difference in partial molar, standard or pseudo volumes as long as we have an infinitely dilute solute. Using this result in Eq. (12) provides,

$$\left(\frac{\partial \mu_2^\infty}{\partial p}\right)_{m_2} = f_D \bar{V}_D^\infty + f_N \bar{V}_N^\infty = \bar{V}_2^\infty \quad (14)$$

which is expected as $\mu_2 = \mu_D = \mu_N$ for all pressures. The use of Eq. (12) in Eq. (8) now provides the same results as Eq. (11) in Eq. (7) and the conflict outlined earlier has been removed.

Second derivatives of the equilibrium constant can be treated in the same way. By choosing $X = \bar{V}_D$ in Eq. (9) we obtain,

$$\left(\frac{\partial \bar{V}_D^\infty}{\partial p}\right)_{m_2} = \left(\frac{\partial \bar{V}_D^\infty}{\partial p}\right)_{m_N, m_D} = -K_D^\infty \quad (15)$$

Taking derivatives of Eq. (12) and using Eq. (10) provides,

$$\begin{aligned} \left(\frac{\partial^2 \mu_D^\infty}{\partial p^2}\right)_{m_2} &= \left(\frac{\partial \bar{V}_D^\infty}{\partial p}\right)_{m_2} - f_N f_D RT \left[\left(\frac{\partial \ln K}{\partial p}\right)_{m_2}\right]^2 + f_N RT \left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2} \\ \left(\frac{\partial^2 \mu_N^\infty}{\partial p^2}\right)_{m_2} &= \left(\frac{\partial \bar{V}_N^\infty}{\partial p}\right)_{m_2} - f_N f_D RT \left[\left(\frac{\partial \ln K}{\partial p}\right)_{m_2}\right]^2 - f_D RT \left(\frac{\partial^2 \ln K}{\partial p^2}\right)_{m_2} \end{aligned} \quad (16)$$

The above relationships link the second expression in Eq. (7) to the formal partial molar compressibilities. Hence, one can use the partial molar volumes and compressibilities alone to characterize changes in the equilibrium constant, but only at infinite dilution of the solute. Finally, we note that use of Eq. (16) indicates that,

$$\left(\frac{\partial^2 \mu_2^\infty}{\partial p^2}\right)_{m_2} = f_D \left(\frac{\partial^2 \mu_D^\infty}{\partial p^2}\right)_{m_2} + f_N \left(\frac{\partial^2 \mu_N^\infty}{\partial p^2}\right)_{m_2} \quad (17)$$

The above compressibility relationships can also be used for the pressure derivative of Eq. (8) and then provide results consistent with Eq. (7).

1.3. Fluctuation solution theory

In the previous section, we related the pseudo chemical potential, volume and compressibility factor to the corresponding thermodynamic quantities that characterize pressure denaturation. What remains, from a statistical mechanics viewpoint, is to express the pseudo volumes and their pressure derivatives in terms of molecular distributions, or

something equivalent, that provides a clear physical picture of these effects. This can then be applied to analyze computer simulation data. To achieve this, we turn to KB theory or, as it is more generally known, FST.

The KB expression for the pseudo volume of an infinitely dilute solute, in terms of a spatial integral, G_{21}^{∞} , over the solute-solvent distribution function, g_{21} , is well known [5,7–12,14,22]. However, this is not the most convenient form for irregular shaped objects such as proteins. It is much more illuminating to invoke the particle number fluctuation (FST) viewpoint, which provides the same results in terms of simple ensemble averages and also facilitates a residue-based decomposition. In addition, we have recently extended the theory to provide expressions for the compressibility of a solute in a binary solution, which has generated additional insight regarding this important thermodynamic property [4,6].

The pseudo molar volume for an infinitely dilute solute in a pure solvent of any size is given by the expression [4-6],

$$V_2^{*,\infty} = - [\langle N_1 \rangle_2 - \langle N_1 \rangle_o] V_1^o \quad (18)$$

where V_1^o is the molar volume of the pure solvent. The same expression is obtained in the presence of other infinitely dilute solutes. The FST expression for the pseudo volume of a solute is, in fact, very simple. The volume is determined by the number of waters that are found within a fixed local volume centered on the protein, $\langle N_1 \rangle_2$, compared to the number of waters observed for the same shape and volume of bulk solution, $\langle N_1 \rangle_o$. This definition is not subjective and has been obtained without approximation. Hence, there is no need to integrate a solute-solvent distribution function that is difficult to compute for protein solutes. The volume is determined by an intensive difference between two extensive ensemble averages. Clearly, the water distribution around the solute becomes the sole focus in this approach.

The corresponding expression for the pressure derivative of the pseudo volume is given by [4-6],

$$K_2^{*,\infty} = \kappa_{T,1}^o V_2^{*,\infty} + \beta (V_1^o)^2 [\langle \delta N_1 \delta N_1 \rangle_o] \quad (19)$$

where $\langle \delta N_1 \delta N_1 \rangle_X = \langle N_1^2 \rangle_X - \langle N_1 \rangle_X^2$ and $\beta = 1/RT$. The square brackets in the second expression clearly indicate that the compressibility is determined by solvent molecule fluctuations in the local vicinity of the protein compared to those in the same shape and volume of bulk solution. Low compressibilities result from small water fluctuations in the vicinity of the solute. It should be noted that the above expression is valid for a solute of any size, not just a protein, and that negative compressibilities are easily accommodated.

1.4. Test System - Ubiquitin

Ubiquitin was chosen for this study because it is a relatively small (76 residue) protein that we have previously analyzed [4]. More importantly, it is often considered a model system for both experimental and computational studies of protein structure and dynamics; its denaturation has been extensively studied by temperature, pressure, and cosolvent denaturation as well as combinations of these perturbations [23-24].

Ubiquitin is a very thermostable protein [25]. It is neutral at pH 7 and therefore no counterions are required in the simulations. It also contains no disulfides. While a net charge or disulfides does not eliminate a protein from candidacy for an FST analysis, their absence does simplify the analysis. Ubiquitin is physicochemically composed of 12 acidic, 12 basic, 20 polar, and 32 non-polar residues. In the native conformation (PDB ID: 1UBQ), its secondary structure composition is 11 hairpin, 12 helical, 11 sheet, and 42 coil residues [26]. These two decompositions (physicochemical and secondary structure type) will be used in the subsequent analysis and the individual residue assignments are listed in Table S1 (see Supporting Information).

Specifically, we have simulated native (N) and fully extended (FE) ubiquitin at pressures up to 10 kbar, which reflects the experimental pressure range used to investigate biochemical systems, with most globular proteins denaturing at above ~4–6 kbar [27-28]. Pressures in this range affect the intermolecular and conformational protein properties, but not the bond distances or angles or the primary structure [28]. Furthermore, pressure denaturation is typically reversible [28]. Herberhold and Winter used Fourier Transform – Infrared (FT-IR) spectroscopy to show that ubiquitin denatures at 5.4 kbar at 294 K and pD 7, that secondary structural elements are present in the pressure-generated DSE, and that the denaturation is reversible at least up to 9 kbar (the maximum pressure used in the study) [29]. Consistent with these findings, the NMR structures of ubiquitin at 3 kbar are very similar to the ambient pressure structure [30-31].

We have performed standard, classical MD simulations of the N and FE conformations, compared their volumes and compressibilities as a function of pressure, and broken these properties down into various contributions. The FE conformation was chosen not because it is a realistic model of (or even a significantly populated member of) the pressure DSE. Indeed, Herberhold and Winter have shown that the pressure-generated DSE for ubiquitin has an increase in α -helical and turn structural elements and a decrease in disorder as compared to the native structure [29]. These changes are probably due to the known strengthening of the majority of hydrogen bonds in ubiquitin for pressures up to at least 2.5 kbar, and the known reduced favorability of hydrophobic interactions at elevated pressures [32-35]. The FE conformation was chosen simply because it is a well-defined extreme (fully exposed to solvent), and easily conceptualized starting point to begin thinking about conformation-specific thermodynamic properties. Furthermore, the FE conformation includes many of the specific short-ranged inter-residue interactions present in all plausible conformations of the protein, while lacking conformation-specific, long-ranged interactions.

2. Methods

2.1. General simulation

The initial structure of N ubiquitin was centered in a rhombic dodecahedron box where the distance between any two parallel box faces (image distance) was 12 nm. The FE ($\phi = \Psi = 180^\circ$) structure was centered in an $\sim 11.5 \text{ nm} \times 11.5 \text{ nm} \times 36.5 \text{ nm}$ ($x \times y \times z$) rectangular box with the principle axis of ubiquitin oriented along z . Relatively large systems, such as those used here, are required to help ensure that bulk water is present in the simulations and therefore the properties will converge with respect to distance away from the protein surface given that sufficient sampling is obtained. The titratable residues were protonated to correspond to neutral pH. A pure water system, consisting of a rhombic dodecahedron box with an image distance of 12 nm was also simulated at each pressure and these systems were used to calculate the corresponding properties of the pure solvent.

The simulations used the standard CHARMM22* force field [36] with the (Gromacs recommended) rigid TIP3P without Lennard-Jones parameters on the hydrogens. The simulations were performed in Gromacs version 4.6 or 4.6.1 [37–38]. Periodic boundary conditions and the minimum image convention were employed. The system coordinates were written out every 10 ps and used at this frequency for the subsequent analysis. The temperature was 300 K in all simulations. All bonds were constrained using the LINCS [39] algorithm for the protein and the Settle algorithm [40] for water. The use of bond constraints allowed for a 2 fs time step to be used for the integration of the equations of motion, which was performed using the Leap-Frog algorithm [41]. The particle-mesh Ewald technique was used to calculate electrostatic interactions with cutoff distances of 1.05 nm for the real space electrostatic and van der Waals interactions [42], a convergence parameter of 2.975 nm^{-1} , cubic interpolation, a maximum fast Fourier transform grid spacing of 0.12 nm for the reciprocal space sum, and tinfoil boundary conditions [43]. The grid-based neighbor list was updated every step and the following flags were set: $\text{nstscenergy} = \text{nsttcoupl} = \text{nstpcoupl} = 1$.

2.2. Native simulations

For the N simulations (PDB 1UBQ, an X-Ray crystal structure) [26], the system was minimized for over 1000 steps using the steepest descent algorithm followed by 100 ps of equilibration at 300 K and the appropriate pressure (1, 3,000, 6,000, or 10,000 bar) using Berendsen T -coupling and Berendsen isotropic P -coupling with $\tau_T = 0.1 \text{ ps}$ and $\tau_P = 0.5 \text{ ps}$. Nosé-Hoover (chain length of one) T -coupling and Parrinello-Rahman isotropic P -coupling were then used for the production simulations with $\tau_T = 2 \text{ ps}$ and $\tau_P = 2.5 \text{ ps}$ [42]. Two temperature coupling groups were used (protein and water), and a $4.5 \times 10^{-5} \text{ bar}^{-1}$ compressibility was used. For the production simulations, the N conformation was simulated with no position restraints for 100 ns at each of the pressures, and the results are presented as the average and, when error bars are shown, standard deviation/(number of subaverages)^{1/2} from four 20-ns subaverages of the last 80 ns, based upon the assumption that the error is proportional to $1/t_{\text{sim}}^{1/2}$.

2.3. Fully extended simulations

For the FE simulations, the system was minimized for over 1000 steps using the steepest descent algorithm followed by 100 ps of C^α position restrained equilibration and then at least 1 ns of unrestrained equilibration. Both sets of equilibrations were at 300 K and the appropriate pressure (1, 3,000, 6,000, or 10,000 bar) using Berendsen *T*-coupling and Berendsen semiisotropic *p*-coupling with $\tau_T = 0.1$ ps and $\tau_p = 0.5$ ps and a compressibility of 0 in the *x*- and *y*-dimensions and 4.5×10^{-5} bar⁻¹ in the *z*-dimension, such that only the *z*-dimension could change length. The second, unrestrained, equilibration was performed to obtain the correct bulk water density at each pressure. Following this equilibration of the box size, the output box size from the previous 100 ps restrained equilibration was adjusted to the *z*-dimension obtained from the unrestrained equilibration. This configuration was used to start NVT production simulations with weak C^α position restraints (force constants of 10 kJ mol⁻¹ nm⁻²) for 100 ns at each of the pressures, to ensure that the FE conformation remained fully extended. This protocol was devised to avoid potential artifacts from the use of position restraints with a barostat, where strong restraint forces can affect the pressure and thereby the bulk solvent density. The results are presented as the average and, when error bars are shown, standard deviation/(number of subaverages)^{1/2} from four 20-ns subaverages.

2.4. Stability of conformations

The initial radius of gyration was 1.18 and 7.98 nm for the N and FE conformations, respectively. The initial solvent accessible surface area (SASA) values are shown in Table 1. The SASA drops by ~4% for N and ~9% for FE over the course of the simulations (not shown). The native secondary structural elements remained intact in all N simulations. Following 100 ns of production simulation for each pressure, the C^α root mean squared deviation (RMSD) for the final structures, after a translational and rotational fit to the initial structure, was 0.15 nm, 0.19 nm, 0.20 nm, and 0.11 nm, for 1 bar, 3 kbar, 6 kbar, and 10 kbar, respectively. When considering residues 1–70 only, as the C-terminus is flexible in the low (and high) pressure experimental NMR models [30], these RMSD values dropped to 0.08 nm, 0.08 nm, 0.07 nm, and 0.08 nm, for 1 bar, 3 kbar, 6 kbar, and 10 kbar, respectively. When instead fit to model 1 from the 3 kbar NMR experimental structure (PDB ID: 1V81) [30], the C^α RMSD at 100 ns for residues 1–70 was 0.12 nm, 0.14 nm, 0.12 nm, and 0.13 nm, for 1 bar, 3 kbar, 6 kbar, and 10 kbar, respectively. Clearly, the N protein simulations remain close to the native structure even under high pressures.

2.5. Analysis of the native conformation

This section describes how to analyze the native conformation, but it would also apply to most any other conformation. The FE conformation is an extreme case and was analyzed differently as described in Section 2.6. Instantaneous measures of the number of waters around the protein, $(N_1)_2$, and in an equivalent volume of pure water, $(N_1)_o$, are made for each snapshot of the simulation. From these values, $\langle N_1 \rangle_2$ and $\langle N_1 \rangle_o$ may be obtained after averaging over all the snapshots. For the pure water analysis, the protein coordinates for a given snapshot are superimposed on a pure water configuration and the waters are counted the same way they are counted in the protein system. Obviously, many more waters are found in this case. Then, the protein volume as determined by the water distribution to a

distance d away from the protein surface, $V_{2^{*,\infty}}(d)$, is obtained by comparing the number of waters found in each system and scaling the result by the volume of a water molecule in pure water as given by Eq. (18). This method avoids the need to integrate an irregular shape, which has been a problem for the simulation community in the past when attempting to use the KB method to determine protein volumes [10]. The center of mass of each water was used to determine if it fell within a given distance from the surface of the protein.

2.6. Analysis of the fully extended conformation

Due to the large size of the FE system and its reduced computational efficiency due to its long z -dimension, it was desirable to not have to simulate another pure water system of the same geometry and size as the FE system if it was not required. Since, in this work, only the pseudo residue volumes were obtained from the FST expressions, all that was needed was the number density of pure water. Since the simulated V_1^0 as a function of pressure was already determined in the pure water simulations of the same geometry and size system as the native ubiquitin simulations, for the analysis of $\langle N_1 \rangle^0$ for the FE system, a random number generator was used to generate random water configurations given the previously simulated (rhombic dodecahedron) pure water V_1^0 value at each pressure. If the protein compressibility were to be obtained from the FST expressions, then this method would not work due to the need for the true pure water fluctuations, for which a random number generator cannot be used as a substitute.

To obtain plateaued values for the FE properties, the V_1^0 value had to be adjusted for some pressures. It was found that small changes in the V_1^0 value (tenths of 1%) were enough to affect the flatness of the integrated water distribution around the protein (not shown). This is understandable because the system contains nearly 161,000 water molecules; thus, it is necessary that V_1^0 is known precisely, and small changes can lead to significant problems.

2.7. Hill/Schnell analysis

In previous studies of solute and solvent mixtures, it has been observed that the particle fluctuations can display a significant local volume size dependence [44–46]. Fortunately, Schnell et al. have analyzed these finite system size effects, using approaches similar to Hill [44,47], and observe that the fluctuations display a $1/L$ size dependence, where L is some measure of the size of the local region of interest [44]. Hence, infinite system limiting values can easily be obtained by extrapolation. More details are provided elsewhere [45–46,48]. In this study, total protein and residue volumes were obtained using the Hill/Schnell approach taking the linear region to be $(1\text{--}2\text{ nm})^{-1}$ [44–47]. For the N conformation, this produces nearly identical results to averaging the distance-dependent volume values between 2 and 3 nm. The general noise on the FE conformation data prevented the two methods from producing identical results for this extreme conformation. However, no systematic difference was found that would suggest that system size effects were present that needed to be corrected for using a Hill/Schnell approach. We conclude that the Hill/Schnell approach used to obtain all the volumes used in this study was not in fact necessary. The comparison of the two methods is presented in Figs. S1–S5 of the Supporting Information.

2.8. Residue volume and compressibility calculations

The residue volumes were obtained by assigning each water molecule to a residue according to that water's distance from the nearest heavy atom of a residue, *i.e.* we have used a simple proximity criterion. We emphasize that this type of decomposition is not unique and the residue-based properties are not thermodynamically accessible. However, the sum of all the residue volumes is the well-defined protein volume that can be compared to the experimentally measured thermodynamic protein volume. The results presented here suggest this simple approach can be very informative. The protein compressibility factor was obtained from the pressure derivative of a quadratic fit of the Hill/ Schnell protein volume as a function of pressure instead of using the FST expression shown in Eq. (19). This is due to the absence of a clear method for decomposing Eq. (19) into residue-based compressibility factors. Consequently, the total and residue volumes, compressibility factors, and compressibilities are related by,

$$\begin{aligned} V_2^{*,\infty} &= \sum_i V_{2,i}^* \\ K_2^{*,\infty} &= - \sum_i (\partial V_{2,i}^* / \partial p)_{m_2} \\ \kappa_2^{*,\infty} &= \sum_i \phi_i \kappa_{2,i}^* \end{aligned} \quad (20)$$

where $V_{2,i}^0$ is the contribution to the total volume of the protein from residue i , with $\phi_i = V_{2,i}^* / V_2^{*,\infty}$ and $\kappa_{2,i}^* = (\partial \ln V_{2,i}^* / \partial p)_{m_2}$ and the sum is over all i residues. Finally, we note that the pseudo compressibility of a solute is related, using an alternative definition of the partial molar compressibility, through the following relationship,

$$\bar{\kappa}_2^\infty \equiv \frac{1}{V_2^\infty} \left(\frac{\partial \bar{V}_2^\infty}{\partial p} \right)_{T,m_2} = - \frac{RT}{V_2^\infty} \left(\frac{\partial \kappa_{T,1}^0}{\partial p} \right)_{T,m_2} + \frac{V_2^{*,\infty}}{V_2^\infty} \kappa_2^{*,\infty} \quad (21)$$

for an infinitely dilute solute, after using the definitions in Eq. (3). The partial molar and pseudo volumes, and the partial molar and pseudo compressibilities, differ by terms that are typically negligible for proteins in aqueous solutions under ambient conditions.

3. Results

Fig. 1 shows that the water probability distribution, g_{21} , reaches the bulk distribution beyond ~ 1 nm from the surface of the protein. Several solvation shells are observed at smaller distances and the water appears to be slightly more structured around the FE conformation, although not at contact. The simulated pseudo protein volume $V_2^{*,\infty}$ is also displayed. A constant value for the volume is obtained for distances beyond which the solvent distribution reaches the bulk random distribution. We have previously established that a center of mass based analysis provides the same converged properties as does the present surface-based analysis [4]. However, the surface analysis is favored as it reaches the required plateau value at shorter distances. Small differences between the two protein conformations are also

observable, although the FE results exhibit a general increase in noise compared to the native results. This is possibly because the FE conformation is unrealistic and therefore the water distribution around this conformation is too unfavorable. As a crude proxy for truly quantifying this possible unfavorability, we note that the B-factors for the FE residues were significantly higher than for the N residues, and the time to break a given protein-water oxygen contact was around three fold smaller for the FE conformation (data not shown) [49–50]. In addition, restraining all heavy atoms instead of just the C α 's did reduce the FE B-factors, but it did not reduce the noise on the FE volume. Our previous study analyzing two different denatured state conformations, generated from a high temperature (390 K) unfolding simulation by the D. E. Shaw group [42], did not exhibit this same general increase in noise. In fact, they were statistically comparable to the native conformation [4]. Thus, we conclude that the FST method presented here is, in general, applicable to both native and denatured conformations, with the FE conformation representing an extreme, and apparently a more problematic, example.

Fig. 2 shows $V_2^{*,\infty}$ as a function of pressure for both protein conformations and also the volume change upon unfolding. For comparison, the volume of pure TIP3P water at 300 K and 1 bar is 0.01847 L/mol. Specific volumes can be determined by noting the mass of Ubiquitin is 8565 g/mol. Differences between the two conformations are observed and are on the order of a few water molecules (out of the ~ 346 water molecule-equivalent $V_N^{*,\infty}$ value at 1 bar), which agrees with the volume changes expected experimentally [51]. The FE conformation has a smaller volume than N, although only at higher pressures, and the magnitude of this difference increases as the pressure increases. The negative sign seems to indicate that the FE conformation is consistent with pressure denaturation, i.e. that N ubiquitin would be driven to unfold to the FE conformation at elevated pressure if the two states studied here were indeed the only two options. Experimentally, the standard state volume change for unfolding (assuming a two-state process) determined by Herberhold and Winter was - 0.05(2) L/mol, or - 2(1) water molecules [29]. These results are consistent with the findings of Hinz and others that the change in volume upon unfolding under excess pressure is negative [52-53]. The change in sign upon unfolding at ambient pressures has been debated in the literature, with Royer arguing that it should also always be negative as long as the temperature is not elevated, and Chalikian arguing that either sign is allowed [53-54].

We have also decomposed the total conformation-specific pseudo volumes into residue-based contributions. Fig. 3 shows the distance-dependent residue pseudo volumes for N ubiquitin analyzed using the approach described in Sections 2.5 and 2.7. The values plateau by ~ 1 nm from the protein surface. Buried residues will never have a water assigned to them, because that water is always closer to a surface residue, and thus they do not exhibit the usual “peaks and valleys,” due to the presence of solvation shells, that the solvent accessible residues exhibit. Instead, the volume for the buried residues is comparatively featureless; it increases from zero and reaches its plateau value at very short distances (see also Fig. S1 and S3).

Fig. 4 shows the limiting values of the residue pseudo volumes for the native conformation. The errors on the residue volumes are small and suggest that meaningful results can be

obtained and clear differences can be seen between different residues. While the equivalent FE data is noisier (not shown), the figures below show that statistically meaningful differences between different residue types are still obtainable in many cases.

It should be noted that the error on the residue volumes is correlated. For example, at 300 K and 1 bar, the total native protein volume and standard deviation is 6.395 ± 0.013 L/mol, but using the propagation of errors formula for the sum of the individual residue standard deviations (assuming no covariance) would give a standard deviation that is 2.4 times larger, at ± 0.031 L/mol. There is a simple physical reason for such a correlation. As a water molecule moves from the vicinity of one residue to another neighboring residue the instantaneous volume of the initial residue increases while that of the new residue decreases. Hence, not only are the residue volumes correlated, they are anti-correlated in the FST approach.

Figs. 5 and 6 show the raw Hill/Schnell residue volumes as a function of pressure and simple quadratic fits to these volumes, which are required to obtain residue pseudo compressibilities in the absence of the use of Eq. (19). The expected trends are observed, i.e. residues composed of more atoms (e.g., Arg, Tyr) have larger volumes than residues composed of fewer atoms (e.g., Gly, Ala). However, different residues of a given type also show variations, as well as variations in their response to pressure (e.g., Arg, Lys, Thr) and, therefore, their compressibility. The data for N ubiquitin is very well-behaved, while the data for FE is statistically more challenging and thus the fits are more uncertain.

The residue-based contributions to the total protein volume provide exquisite detail and information that should help to reveal important contributions to pressure denaturation. Several decompositions and plausible re-groupings can be imagined and the decomposition/regrouping that provides the most insights will no doubt differ for different systems under study and different state points. Our first attempts at re-grouping residue-based contributions have involved grouping nonpolar, polar, positively charged, and negatively charged residues, as well as grouping residues according to their secondary structure assignment in the N protein (see Table S1).

Figs. 7–8 show this regrouping for the change in volume for a given conformation upon changing the pressure, as well as the volume difference for unfolding as a function of pressure. As already noted, pressure decreases the protein volume. However, with this further analysis, it is revealed that the volume change is dominated by the nonpolar groups for both the N and FE conformations. Our results also indicate that the volume difference is dominated by changes in the polar and positively charged residues. These results are consistent with some of the findings of Mitra et al. that, in general, at ambient temperatures a negative sign comes about by a positive contribution from nonpolar groups that is outweighed by a negative contribution from polar groups as well as the hydration of cavities (not explicitly investigated here, but accounted for) [55]. However, the use of a fully extended denatured state clouds this comparison. These findings are inconsistent with studies of the transfer of model hydrocarbon groups to water [53]. The acidic groups are anomalous and their behavior with pressure corresponds to a negative compressibility (see also Fig. 9). Eq. (19) indicates that this is due to a significantly reduced fluctuation of water

molecules around these groups, as would be expected if electro-striction is an important factor.

In Fig. 8 it is shown that, for this case, all the secondary structure types seem similar when looking at the volume at x bar minus the volume at 1 bar for a given conformation. However, when looking at the volume change upon unfolding, the shape of the overall protein volume change appears to be largely determined by the change in the volume of the hairpin and coil residues, shifted by the change in volume of the residues in the other well-defined secondary structures.

If we assume that the FE conformation is a reasonable representation of the pressure unfolded state, that a negative volume change for unfolding signals instability at high pressures, and that the current simulations are sufficiently accurate, then several conclusions can be drawn from the above observations. From the observed volume changes in Fig. 7, we see that, at high pressures, one would expect the hydrophobic contacts to be maintained, with larger changes to polar residues. Positively charged residues would also tend to favor a more solvent exposed environment but, as most charged residues occupy surface sites, this is unlikely to have as significant of an effect on the protein conformation. From the observed volume changes in Fig. 8, we see that the hairpin and coil regions would only be perturbed at high pressure, while the helix and sheet residues would be slightly destabilized. The observed destabilization of the sheet is supported by high-pressure (max 2.5 kbar) NMR findings indicating that the last hydrogen bond in the parallel β -sheet between strands $\beta 3$ and $\beta 5$ (R42 $\rightarrow$ V70) was strongly destabilized by pressure, although adjacent hydrogen bonds were marginally stabilized by high pressure [35]. The same NMR study did not predict that the helical hydrogen bonds would be destabilized by high pressure [35]. However, it has already been determined that the ubiquitin helix is too unstable with CHARMM22* at 300 K and 1 bar [42], which may partially account for the discrepancy between our observations and those of the high-pressure NMR hydrogen bond strength study.

In contrast to the negative sign observed for the helix to extended “transition” shown in the bottom row of Fig. 8, Krobath et al. calculated a positive sign for the helix to coil transition of hexadeca-alanine [56]. This study involved implicit solvent simulation followed by a protein volume that was defined by combining information from the van der Waals volume, Connolly molecular surface volume [57], and the solvent excluded volume. The volume was calibrated against models of residue partial molar volumes extracted from experimental data. However, they note that the sign change they observe is specific to peptides, i.e. it is opposite to the sign change observed experimentally for globular protein volume changes upon unfolding (disregarding the proteins' secondary structure compositions) [56]. They attributed the opposite sign to the lack of cavities in peptides, not to attributes of helices in particular. Because ubiquitin is a globular protein, the opposite sign for the helical residues observed here is not in disagreement with Krobath et al.'s overall conclusions, and their isolated helical peptide results may not serve as general predictors of the sign of the helical contributions to the unfolding of globular proteins [56]. Alternatively, the difference could be due to the explicit inclusion of long-range solvent effects in the present analysis.

It should be noted that, in this work, we have not performed any normalization for the number of residues within a certain group (physicochemical or secondary structure). Since the coil residues correspond to 55% of the total residues, the dominance of this group could be attributed to the sheer number of coil residues. In this instance, it appears that more insight is gained by looking at the physicochemical groupings. A further complication involves the fact that hydrophobic residues, for example, include contributions from the peptide backbone and are not purely nonpolar in character. These issues will be addressed in a future publication.

In the top and middle rows of Figs. 9–10, the pseudo compressibility factors ($K_2^{*,\infty}$) of the protein are shown and decomposed into physicochemical and secondary structure group-based contributions, respectively. As expected, the protein compressibility is approximately one order of magnitude smaller than the compressibility of pure water [58–60]. The positive value for the FE compressibility disagrees with an empirical, group-additivity based, calculation of fully extended protein compressibilities, which were negative in sign [61]. Fig. 9 shows that this positive sign is predominantly due to the hydrophobic residues at lower pressures. Their contribution decreases as pressure increases, but this is essentially countered by an increase in compressibility of the negatively charged residues. The trends in the contributions from the polar and positively charged residues differ for the N and FE conformations.

The pressure dependence of the FE compressibility is different than that of the N. The N conformation becomes more compressible as the pressure increases, whereas the FE remains the same to within the error. The nonpolar residues become less compressible with increasing pressure for both the N and FE conformations. In the folded conformation, this is counteracted by polar and charged residues, whereas in the FE conformation, only the acidic residues cancel out these contributions. Furthermore, the negative compressibility of the acidic residues (extractable from the data presented in Fig. 7) is borne out in Fig. 9, for pressures of 1 bar–6 kbar. Fig. 10 shows that the overall change in compressibility for the protein with pressure can be attributed predominantly to the coil residues.

Chalikian states that the change in compressibility upon unfolding can be either sign at ambient pressure [62], but that for pressure denaturation it is generally positive [63]. The bottom row of Figs. 9–10 show a positive (but decreasing with pressure) change in the compressibility factor upon unfolding that can largely be attributed to the polar and basic residues or the coil residues. Chalikian has also summarized that the change in compressibility upon going from native to a “molten globule” or “compact intermediate” is positive, but to a completely unfolded conformation is negative [63]. These findings seem to agree with the general consensus that pressure denaturation is not compatible with a fully extended structure, as indicated experimentally [28,32].

Although the observation that $V_2^{*,\infty} > 0$ in this study indicated that FE could be consistent with pressure denaturation, the observation that $K_2^{*,\infty} > 0$ does contradict the consensus of the community that a fully extended structure is an inappropriate model for pressure denaturation. Future studies are required to check if $K_2^{*,\infty} > 0$ is observed at

pressures above the denaturation pressure when more realistic conformations are used, as well as to check for consistency across different force fields.

4. Conclusions

The surface-based volume of specific protein conformations can be obtained from simulation using FST. The use of this approach allows for the assignment of contributions to the thermodynamic properties from individual residues using the “proximity criterion”. This approach should allow for new insights into protein stability and the nature of the denatured state ensemble. Reasonable results appear to be obtained, although further analysis is required to establish their full validity. For the system studied herein, large contributions to the overall protein properties seem to be due to the properties of the nonpolar, acidic, and coil residues. Future work will examine more realistic (and statistically less noisy) denatured state conformations instead of the more extreme FE conformation, additional decomposition methods (such as backbone vs. sidechain), as well as additional proteins and biomolecular force fields to determine if consistent trends emerge. It appears that the Hill/ Schnell approach is not needed for the total protein or residue volumes, and that it is sufficient to obtain plateau values from averaging over a distance sufficiently far from the protein surface (2-3 nm this case).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project described was supported by grant R01GM118719 from the National Institute of General Medical Sciences to PES and grant 1F32AG048690 from the National Institute on Aging to EAP. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Science, the National Institute on Aging, or the National Institutes of Health.

References

1. Prigozhin MB, Liu Y, Wirth AJ, Kapoor S, Winter R, Schulten K, Gruebele M. Misplaced helix slows down ultrafast pressure-jump protein folding. *Proc Natl Acad Sci*. 2013; 110(20):8087–8092. [PubMed: 23620522]
2. Ploetz EA, Smith PE. Infinitely dilute partial molar properties of proteins from computer simulation. *J Phys Chem B*. 2014; 118(45):12844–12854. [PubMed: 25325571]
3. Canchi DR, Paschek D, Garcia AE. Equilibrium study of protein denaturation by urea. *J Am Chem Soc*. 2010; 132(7):2338–2344. [PubMed: 20121105]
4. Ploetz EA, Smith PE. Particle and energy pair and triplet correlations in liquids and Liquid mixtures from experiment and simulation. *J Phys Chem B*. 2015; 119(25):7761–7777. [PubMed: 25961333]
5. Jiao Y, Smith PE. Fluctuation theory of molecular association and conformational equilibria. *J Chem Phys*. 2011; 135(1):014502. [PubMed: 21744905]
6. Ploetz, EA., Smith, PE. Local fluctuations in solution: theory and applications. In: Rice, SA., Dinner, AR., editors. *Advances in Chemical Physics*. second. Vol. 153. John Wiley & Sons, Inc; Hoboken: 2013. p. 311-372.
7. Patel N, Dubins DN, Pomes R, Chalikian TV. Parsing partial molar volumes of small molecules: A molecular dynamics study. *J Phys Chem B*. 2011; 115(16):4856–4862. [PubMed: 21466176]
8. Meng B, Ashbaugh HS. Pressure reentrant assembly: direct simulation of volumes of Micellization. *Langmuir*. 2013; 29(48):14743–14747. [PubMed: 23930981]

9. Floris FM. Nonideal effects on the excess volume from small to large cavities in TIP4P water. *J Phys Chem B*. 2004; 108(41):16244–16249.
10. Yu I, Tasaki T, Nakada K, Nagaoka M. Influence of hydrostatic pressure on dynamics and spatial distribution of protein partial molar volume: time-resolved surficial Kirkwood-Buff approach. *J Phys Chem B*. 2010; 114(38):12392–12397. [PubMed: 20822124]
11. Patel N, Dubins DN, Pomes R, Chalikian TV. Size dependence of cavity volume: a molecular dynamics study. *Biophys Chem*. 2012; 161:46–49. [PubMed: 22133917]
12. Sangwai AV, Ashbaugh HS. Aqueous partial molar volumes from simulation, and individual group contributions. *Ind Eng Chem Res*. 2008; 47(15):5169–5174.
13. Smith, PE., Matteoli, E., O'Connell, JP. *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*. Taylor & Francis; Boca Raton: 2013.
14. Kirkwood JG, Buff FP. The statistical mechanical theory of solutions. I. *J Chem Phys*. 1951; 19(6): 774–777.
15. Mark AE, van Gunsteren WF. Decomposition of the free energy of a system in terms of specific interactions. *J Mol Biol*. 1994; 240(2):167–176. [PubMed: 8028000]
16. Mehrotra PK, Beveridge DL. Structural analysis of molecular solutions based on quasi-component distribution functions. Application to H₂CO]Aq at 25 °C. *J Am Chem Soc*. 1980; 102(13):4287–4294.
17. Baynes BM, Trout BL. Proteins in mixed solvents: a molecular-level perspective. *J Phys Chem B*. 2003; 107(50):14058–14067.
18. Canchi DR, García AE. Cosolvent effects on protein stability. *Annu Rev Phys Chem*. 2013; 64(1): 273–293. [PubMed: 23298246]
19. Canchi DR, García AE. Backbone and side-chain contributions in protein denaturation by urea. *Biophys J*. 2011; 100(6):1526–1533. [PubMed: 21402035]
20. Ben-Naim, A. *Molecular Theory of Solutions*. Oxford University Press; New York: 2006.
21. Ben-Naim A. Solute and solvent effects on chemical-equilibria. *J Chem Phys*. 1975; 63(5):2064–2073.
22. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*. Plenum Press; New York: 1992.
23. Jackson SE. Ubiquitin: a small protein folding paradigm. *Org Biomol Chem*. 2006; 4(10):1845–1853. [PubMed: 16688326]
24. Vajpai N, Nisius L, Wiktor M, Grzesiek S. High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin. *Proc Natl Acad Sci*. 2013; 110(5):E368–E376. [PubMed: 23284170]
25. Ibarra-Molero B, Makhatadze GI, Sanchez-Ruiz JM. Cold denaturation of ubiquitin. *Biochim Biophys Acta Protein Struct Mol Enzymol*. 1999; 1429(2):384–390.
26. Vijay-Kumar S, Bugg CE, Cook WJ. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol*. 1987; 194(3):531–544. [PubMed: 3041007]
27. Erlkamp M, Marion J, Martinez N, Czeslik C, Peters J, Winter R. Influence of pressure and crowding on the sub-nanosecond dynamics of globular proteins. *J Phys Chem B*. 2015; 119(14): 4842–4848. [PubMed: 25781064]
28. Winter R. Synchrotron X-ray and neutron small-angle scattering of lyotropic lipid mesophases, model biomembranes and proteins in solution at high pressure. *Biochim Biophys Acta Protein Struct Mol Enzymol*. 2002; 1595(1-2):160–184.
29. Herberhold H, Winter R. Temperature- and pressure-induced unfolding and refolding of ubiquitin: a static and kinetic Fourier transform infrared spectroscopy study. *Biochemistry*. 2002; 41(7): 2396–2401. [PubMed: 11841233]
30. Kitahara R, Yokoyama S, Akasaka K. NMR snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 Kbar. *J Mol Biol*. 2005; 347(2):277–285. [PubMed: 15740740]
31. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, Wand AJ. Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc*. 2012; 134(20):8543–8550. [PubMed: 22452540]

32. Ravindra R, Winter R. On the temperature-pressure free-energy landscape of proteins. *ChemPhysChem*. 2003; 4(4):359–365. [PubMed: 12728550]
33. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc Natl Acad Sci U S A*. 1998; 95(4):1552–1555. [PubMed: 9465053]
34. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR. Hydrophobic effects on a molecular scale. *J Phys Chem B*. 1998; 102(51):10469–10482.
35. Nisius L, Grzesiek S. Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network. *Nat Chem*. 2012; 4(9):711–717. [PubMed: 22914191]
36. Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? *Biophys J*. 2011; 100(9):L47–L49. [PubMed: 21539772]
37. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013; 29(7):845–854. [PubMed: 23407358]
38. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM force field in Gromacs: analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J Chem Theory Comput*. 2010; 6(2):459–466. [PubMed: 26617301]
39. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. Lincs: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997; 18(12):1463–1472.
40. Miyamoto S, Kollman PA. Settle: an analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem*. 1992; 13(8):952–962.
41. Hockney RW, Goel SP, Eastwood JW. Quiet high resolution computer models of a plasma. *J Comput Phys*. 1974; 14(2):148–158.
42. Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. *Proc Natl Acad Sci U S A*. 2013; 110(15):5915–5920. [PubMed: 23503848]
43. Darden T, York D, Pedersen L. Particle mesh Ewald: an n.Log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993; 98(12):10089–10092.
44. Schnell SK, Liu X, Simon JM, Bardow A, Bedeaux D, Vlugt TJH, Kjelstrup S. Calculating thermodynamic properties from fluctuations at small scales. *J Phys Chem B*. 2011; 115(37):10911–10918. [PubMed: 21838274]
45. Krüger P, Schnell SK, Bedeaux D, Kjelstrup S, Vlugt TJH, Simon JM. Kirkwood-Buff integrals for finite volumes. *J Phys Chem Lett*. 2013; 4(2):235–238. [PubMed: 26283427]
46. Kjelstrup S, Schnell SK, Vlugt TJH, Simon JM, Bardow A, Bedeaux D, Trinh T. Bridging scales with thermodynamics: from nano to macro. *Adv Nat Sci Nanosci Nanotechnol*. 2014; 5(2):023002.
47. Hill, TL. *Thermodynamics of Small Systems, Parts I & II*. W. A. Benjamin, Inc.; New York: p. 1963-1964.
48. Ploetz EA, Rustenburg AS, Geerke DP, Smith PE. To polarize or not to polarize? Charge-on-Spring versus KBFF models for water and methanol bulk and vapor-liquid interfacial mixtures. *J Chem Theory Comput*. 2016; 12(5):2373–2387. [PubMed: 27045390]
49. Luzar A, Chandler D. Hydrogen-bond kinetics in Liquid water. *Nature*. 1996; 379(6560):55–57.
50. van der Spoel D, van Maaren PJ, Larsson P, Timneanu N. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. *J Phys Chem B*. 2006; 110(9):4393–4398. [PubMed: 16509740]
51. Lin LN, Brandts JF, Brandts JM, Plotnikov V. Determination of the volumetric properties of proteins and other solutes using pressure perturbation calorimetry. *Anal Biochem*. 2002; 302(1):144–160. [PubMed: 11846388]
52. Hinz HJ, Vogl T, Meyer R. An alternative interpretation of the heat capacity changes associated with protein unfolding. *Biophys Chem*. 1994; 52(3):275–285. [PubMed: 7999977]
53. Royer CA. Revisiting volume changes in pressure-induced protein unfolding. *Biochim Biophys Acta Protein Struct Mol Enzymol*. 2002; 15951-2:201–209.

54. Chalikian TV. Volumetric properties of proteins. *Annu Rev Biophys Biomol Struct.* 2003; 32:207–235. [PubMed: 12543706]
55. Mitra L, Hata K, Kono R, Maeno A, Isom D, Rouget JB, Winter R, Akasaka K, García-Moreno B, Royer CA. Vi-value analysis: a pressure-based method for mapping the folding transition state ensemble of Proteins. *J Am Chem Soc.* 2007; 129(46):14108–14109. [PubMed: 17960902]
56. Krobath H, Chen T, Chan HS. Volumetric physics of polypeptide coil-helix transitions. *Biochemistry.* 2016; 55(45):6269–6281. [PubMed: 27775315]
57. Connolly M. Analytical molecular surface calculation. *J Appl Crystallogr.* 1983; 16(5):548–558.
58. Kharakoz DP. Protein compressibility, dynamics, and pressure. *Biophys J.* 2000; 79(1):511–525. [PubMed: 10866977]
59. Dadarlat VM, Post CB. Insights into protein compressibility from molecular dynamics simulations. *J Phys Chem B.* 2001; 105(3):715–724.
60. Høiland H, Hedwig GR. 6 compressibilities of amino acids, peptides and proteins in aqueous solution. Springer Materials-The Landolt-Börnstein Database. 2003:1–17.
61. Kharakoz DP. Partial volumes and compressibilities of extended polypeptide chains in aqueous solution: additivity scheme and implication of protein unfolding at normal and high pressure. *Biochemistry.* 1997; 36(33):10276–10285. [PubMed: 9254626]
62. Chalikian, TV. Partial molar volumes of proteins in solution. In: Wilhelm, E., Letcher, T., editors. *Volume Properties: Liquids, Solutions and Vapours.* Vol. 2014. The Royal Society of Chemistry; Cambridge, UK: p. 542-574.
63. Taulier N, Chalikian TV. Compressibility of protein transitions. *Biochim Biophys Acta Protein Struct Mol Enzymol.* 2002; 1595(1-2):48–70.

Fig. 1.

Top: Probability of finding a water as a function of distance (d) in nm from the protein surface relative to the probability of finding a water in bulk water (g_{21}). Bottom: Pseudo protein volume ($V_2^{*,\infty}$) in L/mol as a function of distance from the protein surface.

Fig. 2.

Top: Pseudo volume ($V_2^{*,\infty}$) for native (N) and fully extended (FE) ubiquitin conformations as a function of pressure in kbar. Units are L/mol. Bottom: Change in pseudo volume upon unfolding, $\Delta V_2^{*,\infty} = V_{FE}^{*,\infty} - V_N^{*,\infty}$. Units are cm³/mol. The lines connecting the points are simply a guide for the eyes. The volumes plotted correspond to the pressure-fitted residue Hill/Schnell volume values not the raw Hill/Schnell values.

Fig. 3. Residue-based contributions (cm^3/mol) to the N ubiquitin pseudo volume at 300 K and 1 bar, averaged over 20–100 ns, as a function of distance (d) in nm. Residues are colored by physicochemical type. Black: Nonpolar, Green: Polar, Blue: Positive, Red: Negative. The smallest volume residue is the C-terminal glycine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Residue-based contributions (cm^3/mol) to the N ubiquitin pseudo volume at 300 K and 1 bar. Black: Nonpolar, Green: Polar, Blue: Positive, Red: Negative. The secondary structure of the N conformation of ubiquitin is shown at the top of the figure for reference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. Residue-based contributions (cm^3/mol) to the N ubiquitin pseudo volume as a function of pressure (in kbar) at 300 K grouped by residue name. The points with error bars are the raw Hill/Schnell values and the curves are quadratic fits of those values. The number of residues of each type is shown in parentheses following the name of the residue. Each residue of a certain type is colored uniquely according to its order in the primary sequence, as noted in the legend. There are no Cys or Trp residues in the Ubiquitin sequence.

Fig. 6. Residue-based contributions (cm^3/mol) to the FE ubiquitin pseudo volume as a function of pressure (in kbar) at 300 K grouped by residue name. The points are the raw Hill/ Schnell values and the curves are quadratic fits of those values. The number of residues of each type is shown in parentheses following the name of the residue. Each residue of a certain type is colored uniquely according to its order in the primary sequence, as noted in the legend.

Fig. 7. Physicochemical group-based contributions to pseudo volume changes. Top&Middle: Changes in volume for a given conformation (N or FE) upon a change in pressure (x bar – 1 bar) at 300 K. Bottom: Changes in volume upon unfolding as a function of pressure. Volume is in cm³/mol and pressure is in kbar. The lines connecting the points are simply a guide for the eyes. These values were generated using the pressure-fitted residue Hill/Schnell volume values not the raw Hill/Schnell values.

Fig. 8.

Secondary structure group-based contributions to pseudo volume changes. Top&Middle: Changes in volume for a given conformation (N or FE) upon a change in pressure (x bar–1 bar) at 300 K. Bottom: Change in volume upon unfolding as a function of pressure. Volume is in cm^3/mol and pressure is in kbar. The lines connecting the points are simply a guide for the eyes. These values were generated using the pressure-fitted residue Hill/Schnell volume values not the raw Hill/Schnell values.

Fig. 9.

Physicochemical group-based contributions to the pseudo compressibility factors. Top & Middle: Contributions to the N and FE pseudo compressibility factors. Bottom: Change in pseudo compressibility factor contributions upon unfolding, $K_2^{*,\infty} = K_{FE}^{*,\infty} - K_N^{*,\infty}$. The pseudo compressibility factor is in $\text{cm}^3/\text{mol/kbar}$ and pressure is in kbar. The colors are the same as those shown in Fig. 7. The lines connecting the points are simply a guide for the eyes.

Fig. 10.

Secondary structure group-based contributions to the pseudo compressibility factors. Top & Middle: Contributions to the N and FE pseudo compressibility factors. Bottom: Change in pseudo compressibility factor contributions upon unfolding, $K_2^{*,\infty} = K_{FE}^{*,\infty} - K_N^{*,\infty}$. The pseudo compressibility factor is in cm³/mol/kbar and pressure is in kbar. The colors are the same as those shown in Fig. 8. The lines connecting the points are simply a guide for the eyes.

Solvent Accessible Surface Area (SASA) using a solvent probe radius of 0.14 nm for initial conformations of Native (N) and Fully Extended (FE) Ubiquitin at 300 K and 1 bar.

Table 1

Residues		SASA (nm ²)			SASA (%)		
No.	%	N	FE	N	N	FE	FE
Residue physicochemical type							
Nonpolar	32 42%	13	42	27%	38%		
Polar	20 26%	13	28	27%	26%		
Positive	12 16%	12	23	25%	21%		
Negative	12 16%	11	17	21%	16%		
Total	76 100%	50	110	100%	100%		
Residue secondary structure type							
Hairpin	11 14%	6	16	11%	15%		
Helix	12 16%	7	18	15%	17%		
Sheet	11 14%	5	17	10%	15%		
Coil	42 55%	32	59	64%	53%		
Total	76 100%	50	110	100%	100%		