THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

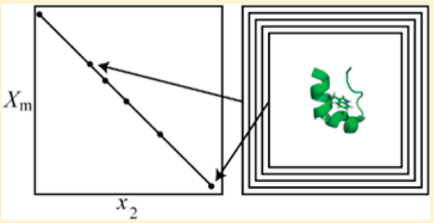
Infinitely Dilute Partial Molar Properties of Proteins from Computer Simulation

Elizabeth A. Ploetz and Paul E. Smith*

Department of Chemistry, Kansas State University, 213 CBC Building, Manhattan, Kansas 66506-0401, United States

Supporting Information

ABSTRACT: A detailed understanding of temperature and pressure effects on an infinitely dilute protein's conformational equilibrium requires knowledge of the corresponding infinitely dilute partial molar properties. Established molecular dynamics methodologies generally have not provided a way to calculate these properties without either a loss of thermodynamic rigor, the introduction of nonunique parameters, or a loss of information about which solute conformations specifically contributed to the output values. Here we implement a simple method that is thermodynamically rigorous and possesses none of the above disadvantages, and we report on the method's feasibility and computational demands. We calculate infinitely dilute partial molar properties for two proteins and attempt to distinguish the thermodynamic differences between a native and a denatured conformation of a designed miniprotein. We conclude that simple ensemble average properties can be calculated with very reasonable amounts of computational power. In contrast, properties corresponding to fluctuating quantities are computationally demanding to calculate precisely, although they can be obtained more easily by following the temperature and/or pressure dependence of the corresponding ensemble averages.



INTRODUCTION

Equilibrium constants (and their derivatives) that describe the delicate balance between the native and denatured states of proteins are frequently measured to understand if and why certain proteins are more stable than others and how mutations can alter this stability.[1] However, under physiological conditions, the ratio of [denatured]/[native] is usually very small, making it difficult to measure accurately. Often the equilibrium must be perturbed, for example by changing the temperature or pressure, in order to achieve an equilibrium constant that is larger in magnitude and can, therefore, be calculated more precisely. Afterward, the results are usually extrapolated back to physiological or standard conditions.[2] Thus, it is important to understand the effects of temperature and pressure on the equilibrium constant.[1,2] One way of learning more about these effects is through computer simulation. However, structural properties are usually the sole or primary focus of most simulations studies, and the corresponding thermodynamic properties are often ignored.[2,3] When thermodynamic properties are determined from simulation, it is typically by way of a method that does not retain information concerning which conformation gave rise to the specific values, meaning the simulation does not really provide any more insight than an experiment would, or the properties are calculated from phenomenological relationships rather than rigorously derived thermodynamic expressions.

Thermal and Pressure Effects on Conformational Equilibria. Although multiple environmental variables affect the conformational equilibrium of a protein,[4] here we limit our scope to thermal and pressure effects. Hence, the effects of cosolvents and pH on protein equilibrium are not discussed here, although a theoretical framework for understanding their effects on an equilibrium has recently been advanced.[5−8]

To understand the effects of the environment on a protein's thermodynamic properties, it is common to study an infinitely dilute protein in a single solvent. To keep the situation tractable, we will assume that a protein molecule adopts one of only two forms, the native state, N, or the denatured state, D. (The process of classifying a specific conformation as being N or D is subjective, but addressing this subjectivity is beyond the focus of this paper. When studying a specific system's equilibrium by computer simulation, it would be wise to use the same definition as was used experimentally, e.g., measuring tryptophan fluorescence, if possible.) When considering the equilibrium between N and D, N ⇌ D, the standard state Gibbs free energy of unfolding ($\Delta_N^D G^\circ$, hereafter ΔG°) determines the population of the protein conformations through the usual relationship to the thermodynamic equilibrium constant $K = a_D/a_N$, where a_i is the equilibrium activity of form i, through

$$\Delta G^\circ = -RT \ln K = \mu_D^\circ - \mu_N^\circ \quad (1)$$

where R is the gas constant, T is the temperature, μ_i is the chemical potential of the protein in conformation i, and the superscripted o denotes the standard state.[9] The activity coefficients at infinite dilution are assumed to be one, which leads to the replacement of the activities with their concentrations in the equilibrium constant expression.[10]

Received: August 26, 2014
Revised: October 16, 2014
Published: October 17, 2014

While ΔG° quantifies the ratio of species at equilibrium, derivatives of ΔG° reveal increasingly more detailed information the higher their order. The bivariate Taylor expansion of the standard state free energy change for unfolding, $-\beta\Delta\Delta G^\circ(\beta, p) = \ln(K/K_{ref})$, up to second order around a reference (ref) inverse temperature ($\beta = 1/RT$) and pressure (p) provides an avenue to interpret experimental thermal and pressure denaturation data via

$$-\beta\Delta\Delta G^\circ(\beta, p) \approx \mathbf{k}^T\mathbf{d} + \frac{1}{2!}\mathbf{d}^T\mathbf{K}\mathbf{d} \tag{2}$$

where $\mathbf{k}$ is the column vector whose elements are the first derivatives of ln K with respect to β and/or p evaluated at a reference β and p and $\mathbf{K}$ is the matrix whose elements are the corresponding second partial derivatives of ln K. $\mathbf{d}$ is the column vector of deviations from the reference β and p,

$$\mathbf{d} = \begin{bmatrix} \Delta\beta \\ \Delta p \end{bmatrix} \tag{3}$$

where $\Delta X = X - X_{ref}$.

It can be shown that

$$\mathbf{k} = \begin{bmatrix} -\Delta H^\circ \\ -\beta\Delta V^\circ \end{bmatrix}_{p=p_{ref},\beta=\beta_{ref}}$$

$$\mathbf{K} = \begin{bmatrix} RT^2\Delta C_p^\circ & T\Delta\alpha_p^\circ - \Delta V^\circ \\ T\Delta\alpha_p^\circ - \Delta V^\circ & \beta\Delta\kappa_T^\circ \end{bmatrix}_{p=p_{ref},\beta=\beta_{ref}} \tag{4}$$

where H denotes the enthalpy, V the volume, C_p the heat capacity, α_p the thermal expansion, and κ_T the isothermal compressibility. The changes are given by $\Delta H^\circ \equiv H_D^\circ - H_N^\circ$, $\Delta V^\circ \equiv V_D^\circ - V_N^\circ$, $\Delta C_p^\circ \equiv C_{p,D}^\circ - C_{p,N}^\circ$, $\Delta\alpha_p^\circ \equiv V_D^\circ\alpha_{p,D}^\circ - V_N^\circ\alpha_{p,N}^\circ$, and $\Delta\kappa_T^\circ \equiv V_D^\circ\kappa_{T,D}^\circ - V_N^\circ\kappa_{T,N}^\circ$.

First order Taylor expansions provide the following temperature and pressure dependencies of ΔH°, the standard state entropy change (ΔS°), and ΔV° via,[1]

$$\begin{aligned} \Delta H^\circ(\beta) &\approx \Delta H_{ref}^\circ - RT_{ref}^2\Delta C_{p,ref}^\circ\Delta\beta \\ \Delta S^\circ(\beta) &\approx \Delta S_{ref}^\circ - RT_{ref}\Delta C_{p,ref}^\circ\Delta\beta \\ \Delta V^\circ(\beta) &\approx \Delta V_{ref}^\circ - RT_{ref}^2\Delta\alpha_{p,ref}^\circ\Delta\beta \\ \Delta V^\circ(p) &\approx \Delta V_{ref}^\circ - \kappa_{T,ref}^\circ\Delta p \end{aligned} \tag{5}$$

Thus, to use MD simulations to understand how temperature and pressure affect the conformational equilibrium of a protein in significant detail, that is, to map $\Delta\Delta G^\circ$ as a function of β and p, one requires ΔH°, ΔV°, ΔC_p°, $\Delta\alpha_p^\circ$, and $\Delta\kappa_T^\circ$. Similarly, to map ΔH°, ΔS°, and ΔV° as a function of β and p, one requires ΔC_p°, $\Delta\alpha_p^\circ$, and $\Delta\kappa_T^\circ$.

The second derivative properties can also be expressed by

$$\begin{aligned} C_{p,i}^\circ &= (\partial H_i^\circ/\partial T)_p \\ \kappa_{T,i}^\circ &= -(1/V_i^\circ)(\partial V_i^\circ/\partial p)_\beta \\ \alpha_{p,i}^\circ &= (1/V_i^\circ)(\partial V_i^\circ/\partial T)_p \end{aligned} \tag{6}$$

The relationships in eq 6 will be used later. All of the above properties are standard state properties. Standard state properties are usually associated with those of an infinitely dilute protein (for which the activity is approximated by the concentration) in either the native or denatured forms,[11] which is helpful since most molecular dynamics simulations model the infinitely dilute case. In fact, the single protein molecules studied by simulation should really be considered as *pseudo* infinitely dilute (but we will drop the use of pseudo for the sake of brevity), since the amount of protein in a given volume (the formal or effective concentration) will vary depending upon the total system size even though the number of protein molecules is fixed at one.

Most simulations do not readily provide the above infinitely dilute partial molar properties, because the expressions relating the output simulation results to these properties have generally been considered unknown. In contrast, it is well-known that the system fluctuations are related to the compressibility, thermal expansion, and heat capacity of the whole system. In the NpT ensemble the corresponding expressions are

$$\begin{aligned} VRT\kappa_T &= \langle(\delta V)^2\rangle \\ VRT^2\alpha_p &= \langle\delta V\delta H\rangle \\ RT^2C_p &= \langle(\delta H)^2\rangle \end{aligned} \tag{7}$$

where the angled brackets indicate time or ensemble averages, $\delta X = X - \langle X\rangle$ and $\langle\delta X\delta Y\rangle = \langle XY\rangle - \langle X\rangle\langle Y\rangle = \mathrm{cov}[X,Y]$, where cov is the covariance. We emphasize that these are all bulk system properties, which means that when applied to a system containing an infinitely dilute solute, the properties obtained are those of the solvent, perturbed to some extent by the presence of the solute molecule. For a binary mixture in which both species 1 and 2 are at finite concentrations, the process of extracting partial molar properties is straightforward. One needs only to fit the property of interest as a function of composition and then take the partial derivatives with respect to each component.[12] However, for the infinitely dilute case, an appropriate method for determining partial molar properties has not been clear or computationally intuitive.

Current Computer Simulation Methodologies for Determining Infinitely Dilute Partial Molar Quantities. Computationally speaking, we are only aware of one method in the literature that has been used to obtain all of the infinitely dilute partial molar properties and can be considered thermodynamically rigorous, namely, the enhanced sampling technique of replica exchange molecular dynamics (REMD).[13] REMD simulations map out the phase space and allow one to obtain K by determining the fraction of protein molecules that are folded or unfolded as a function of temperature. Once this is known, the free energy of unfolding as a function of temperature can be computed according to[13]

$$\Delta_N^D G^\circ(\beta, p) = -\beta^{-1}\ln\left(\frac{\rho_D}{\rho_N}\right) \tag{8}$$

where ρ_i is the number density of i ($N_i/\langle V\rangle$). After the free energy of unfolding has been computed, the difference (unfolded minus folded) in all of the thermodynamic properties may be fitted simultaneously according to the approach illustrated by, for example, Garcia et al.[13] Note, however, that all that is obtained using REMD is the difference between the thermodynamic properties in the N and D forms, not the values of the thermodynamic properties for each of the native and the many denatured states that collectively gave rise to that difference.

Thus, the REMD approach provides an objective way to calculate the change in an infinitely dilute partial molar quantity, $\Delta\bar{X}_2^\infty$, and is therefore extremely useful from a force field validation or a property prediction point of view. What remains missing, however, is the ability to assign thermodynamic properties to specific conformations, that is, the ability to rank different conformations according to their $\bar{X}_2^\infty$ values. Many approaches for the determination of partial molar quantities of individual biomolecule conformations have been proposed in the literature.

Protein Volume and Compressibility. From a computational viewpoint, many subjective definitions for the volume of a protein exist. Levy and co-workers have used three possible definitions to calculate the volume of 15 proteins.[14] The most striking volume variability was, naturally, for the smallest protein they studied, the insulin monomer (51 residues). They reported the van der Waals (VDW) volume to be 5.564 nm^3, the molecular volume (calculated using a sphere with probe radius of 0.14 nm) to be 6.848 nm^3, and the excluded volume (also calculated using a sphere with probe radius of 0.14 nm) to be 11.142 nm^3.[14] This corresponds to a 100% increase in the volume on going from a VDW volume definition to an excluded volume definition. Even the largest protein they studied, carboxypeptidase A (307 residues), had a 17% increase in the volume on going from a VDW to an excluded volume definition (reported volumes: VDW = 33.571 nm^3, molecular = 42.098 nm^3, excluded = 58.055 nm^3).[14] While there is nothing wrong with calculating these quantities and using them to probe the properties of proteins, it is unclear which approach best corresponds to the rigorously defined partial molar volume. Unfortunately, for those interested in further calculations that depend upon the protein volume, it is also unclear which volume definition should then be used. This has been an issue in several publications.[15,16]

We have found several examples[17−25] where an objective definition of the protein volume has been used via the apparent molar approach usually encountered in experimental studies, that is,[26]

$$\bar{V}_2^\infty = \langle V \rangle - V_1^* N_1 \tag{9}$$

where $\langle V \rangle$ is the average system volume from an NpT simulation, V_1^* is the molar volume of pure 1, N_1 is the number of molecules of species 1, and the angled brackets, again, refer to time or ensemble averages.

Another rigorous $\bar{V}_2^\infty$ definition comes from the Kirkwood−Buff theory of solutions via the expression[9]

$$\bar{V}_2^\infty = RT\kappa_T^* - G_{21}^\infty \tag{10}$$

where the Kirkwood−Buff integral, G_{21}^∞, is given by[27]

$$G_{21}^\infty = \int_\infty^0 4\pi r^2 [g_{21}^{\infty,\mu VT}(r) - 1]\mathrm{d}r \tag{11}$$

and where $g_{21}(r)$ is the radial distribution function of the solvent (1) around the solute (2) defined in the grand canonical (μVT) ensemble. Equation 10 has been used by several authors.[5,20,23,24,28−30] Due to the nonspherical geometry of proteins, the integration in eq 11 is more difficult to perform if one is interested in a surface based approach.[28] Similarly, Hirata and colleagues have derived statistical mechanical expressions for the infinitely dilute partial molar volumes using the Reference Interaction Site Model (RISM) integral equation theory coupled to KB theory.[31,32]

It is worth noting that while traditional KB theory provides a rigorous expression for the volume only, the generalization of KB theory to Fluctuation Solution Theory (FST) provides rigorous expressions for all of the thermodynamic properties considered in this work. However, FST is still in its infancy, and therefore, while the expressions have appeared, they have not yet been applied to obtain the full set of possible thermodynamic properties.[5,33]

Most calculations of $\bar{\kappa}_{T,2}^\infty$ we have found cannot be considered rigorous for one of two main reasons. First, they may have relied on subjective definitions of the protein volume. Second, they often assume that the equation that relates the bulk system compressibility to fluctuations in the volume of the entire system can be applied to a component of the system (i.e., the protein) to obtain its compressibility, that is, that $\bar{\kappa}_{T,2}^\infty = \beta\langle(\delta V_2)^2\rangle/\langle V_2\rangle$,[15,16,34−38] where V_2 is some measure of the protein volume. This may or may not provide a measure of the intrinsic compressibility of the protein, but it is unlikely that it corresponds to the thermodynamically defined infinitely dilute partial molar compressibility of the protein. This idea has been used in the literature many times and appears to have originated from the work of Cooper.[35] It has been further supported by Hinz and co-workers,[39,40] but has also been strongly contested by Imai and Hirata.[31] Furthermore, while many researchers have been able to successfully obtain reasonable agreement with the experimental $\bar{\kappa}_{T,2}^\infty$ by breaking the simulated compressibility into different contributions, such as the "intrinsic," "hydrational," and/or "thermal" components (which involves the use of nonunique volume definitions),[41] here we seek to calculate the thermodynamically defined property, $\bar{\kappa}_{T,2}^\infty$.

A few noteworthy exceptions to the above compressibility studies exist, such as the work of Yu and co-workers mentioned in the previous paragraphs explaining protein volume calculation methods. Yu used eq 10 and then looked at the pressure dependence of the volume to obtain the corresponding compressibility.[28] Likewise, Hirata and co-workers also used the pressure derivative of the volume that was obtained from RISM combined with KB theory to obtain the isothermal compressibility.[31,32] In contrast to these studies, we seek a method that can be used to analyze computer simulations, does not involve integral equations, and directly provides any first or second derivative of the free energy, rather than only providing the volume (obtained from the first pressure derivative of the free energy).

Protein Enthalpy, Heat Capacity, and Thermal Expansion. In a computer simulation, the absolute partial molar enthalpy is available, whereas experiments only provide the excess partial molar enthalpy. When a single protein conformation is simulated, it is not obvious that there would be any value in quoting the absolute enthalpy of that protein form. This may explain why we have found considerably fewer attempts to extract $\bar{H}_2^\infty$ or $\bar{C}_{p,2}^\infty$ from simulations as compared to the preceding properties, outside of the REMD studies. For interesting explanations as to the dearth of simulations targeting $\bar{C}_{p,2}^\infty$, the reader is referred to a review by Prabhu and Sharp[2] and an article by Cooper.[42]

Shing and Chung did calculate the internal energy of an infinitely dilute Lennard−Jones (LJ) sphere in a LJ binary mixture by the finite difference of the system internal energy in the absence or presence of the infinitely dilute solute.[21] However, the method was not generalized to all properties or other types of systems such as biomolecules. As we will show

here, it is trivial to make these generalizations, but the literature from the intervening years suggests that this approach has not been popular. In the same vein, Smit and co-workers recently calculated the enthalpy of a lipid bilayer, with respect to the enthalpy of the solvent background, in order to make a connection to differential scanning calorimetry experiments.[3] Unlike nucleic acids, proteins, and polysaccharides, lipids are not polymeric. Therefore, Smit calculated the (extensive) enthalpy of a lipid aggregate, not an (intensive) infinitely dilute partial molar lipid enthalpy. However, their method is in the same spirit as the method used in this work and it could be applied to any infinitely dilute solute (considering the lipid aggregate as a single object) and to any property, not just the enthalpy.

Lastly, very few examples of protein thermal expansion values calculated from computer simulations studies can be found. They have all used nonunique protein volume definitions and/or decomposed the thermal expansion into different contributions, such as the "intrinsic," "hydrational," and/or "thermal" components, which introduces several additional subjective parameters.[17,43−45]

Aim of This Study. The purpose of this paper is to investigate the computational feasibility of using a general method for extracting the thermodynamically defined $\bar{X}_2^\infty$ that does not result in a loss of information about the conformational state of the protein that gave rise to these properties. The significance of this work is that it allows MD simulations to better fulfill their claim of providing exquisite details about systems, often unavailable from experiment, and to do so using thermodynamically exact relationships and approaches.

As mentioned in the Introduction, obtaining thermodynamically rigorous partial molar properties for species whose concentrations are finite is trivial by computer simulation. However, finite concentrations are typically not feasible for biological system simulations. Thus, it is the infinitely dilute limit that is of primary interest for this work. The current approaches found in the literature to determine the relevant properties are often reasonable, but are typically subjective and often rather involved. In comparison, we show here that it is often simpler to use a thermodynamically rigorous apparent molar approach.

THEORY

As discussed in the Introduction, obtaining partial molar properties is, in principle, straightforward when all species are at finite concentrations. For a multicomponent system of 1 ≡ solvent, 2 ≡ solute, 3, ..., n ≡ cosolvents, a molar property is defined as $X_m = X/N$, where X is the corresponding extensive thermodynamic property of interest and N is the total number of molecules in the system. Here we will be focused on X = H, V, C_p, $V\kappa_T$, and $V\alpha_p$ only, where we have multiplied the κ_T and α_p by the system volume. In addition, we have also investigated results for the kinetic energy, K, but only as a test of the procedure, so as to ensure that the equipartition results were achieved. The partial molar property of 2 is then obtained from

$$\bar{X}_2(\{N\}, p, T) = \left(\frac{\partial X}{\partial N_2}\right)_{\{N\}',p,T} \tag{12}$$

where the subscripted $\{N\}'$ indicates that all N other than N_2 are held constant.

Now consider that 2 is a protein plus its counterions. Generally, it would not be feasible to perform simulations in which one varies the number of protein molecules, because the required system sizes would become elephantine rather quickly. Furthermore, researchers are often more interested in the infinitely dilute system where the inclusion of protein−protein interactions would often be undesirable. A finite difference approach can be used if N_2 is set equal to zero and to one in two separate simulations while $\{N\}'$ is held fixed. As mentioned in the Introduction, this finite difference calculation has been used before for calculations of a LJ sphere's volume and internal energy, for determining protein volumes, and for a lipid aggregate enthalpies.[3,21,46] It has also been noted that such an approach can suffer from large statistical errors.[46]

We instead have constructed a series of infinitely dilute systems in which we vary x_2 by changing N_1, where N_2 is always one. If we choose systems such that even the smallest system contains water that is beyond the "sphere of influence" of the protein, that is, some waters exhibit the thermodynamic properties of bulk water, then the addition of more waters will correspond to a simple dilution process, and the corresponding plot of X_m versus x_2 will be linear. An example plot (created using real data), which illustrates the method, is

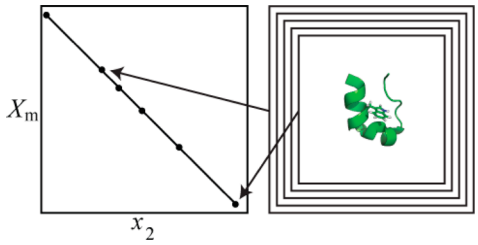


Figure 1. Illustration of the apparent molar approach used to calculate infinitely dilute partial molar properties. A protein is simulated in various box sizes so that a plot of X_m vs x_2 may be achieved from which $\bar{X}_2^\infty$ may be calculated. (Simulation boxes not to scale.)

shown in Figure 1. Due to the linearity, it is easy to fit X_m. The line is then given by

$$X_m = X_m^* + ax_2 \tag{13}$$

where X_m^* is the molar property of pure 1 (always water here) and a is the slope. After taking the partial derivative of eq 13 with respect to N_2, multiplying both sides by N, and taking the limit $x_2 \to 0$, one finds that

$$a = \bar{X}_2^\infty - X_m^* \tag{14}$$

Thus, the linear regression of X_m provides a and subsequently $\bar{X}_2^\infty$, the thermodynamically rigorous infinitely dilute partial molar property of the protein plus its counterions. Rearrangement of eq 13, after substituting eq 14 in for a, indicates that this method is equivalent to the apparent molar approach often used experimentally,[26]

$$\bar{X}_2^\infty = \langle X \rangle - X_m^* N_1 \tag{15}$$

Although we have seen a few molecular dynamics studies that use this approach to calculate $\bar{V}_2^\infty$,[5,17−19,28,46] to our knowledge, it has never been presented generally or applied for the determination of higher derivatives.

In using eq 13, one has a choice of fitting both X_m^* and a, or of fixing the y-intercept and only fitting the slope. We have chosen to fix the y-intercept for the following two reasons: (1) So that we could use the same values of X_m^* regardless of the

dx.doi.org/10.1021/jp508632h | J. Phys. Chem. B 2014, 118, 12844−12854

protein under study and (2) because we have performed the pure water simulation for four times longer than the simulations of systems containing proteins, in an effort to increase the precision on the molar properties of water. It should be noted that one could use this method with any number of components as long as the ratio of 1:3:...:n is held constant and only the ratio of 2:$(N-1)$ varies.

The advantages of the above approach include the following:

(i) It involves an interpolation of data between pure water and some value of x_2, not an extrapolation.
(ii) It is thermodynamically exact.
(iii) No subjective definitions are required.
(iv) It can be applied to extract any partial molar property for any solute in any solvent system regardless of the number of components.
(v) Partial molar properties can be assigned to specific conformations.

However, regardless of the quality of the approach, in computer simulations, the ultimate arbiter remains the quality of the force field and the need for sufficient sampling of the relevant phase space. Errors in our results will be due to these factors. Specifically, there will be errors in our results due to a classical treatment of those X which have significant quantum mechanical contributions, such as the constant pressure heat capacity.

The disadvantage of the approach is that, as alluded to above, multiple simulations are required to obtain a good fit for the calculation of $\overline{X}_2^{\infty}$. Since it is a linear fit, in principle, only two compositions are required, namely, a pure water box and one protein system. However, if any noise is present in the results, it can be better identified when multiple compositions are used. Furthermore, systematic deviations from linearity in the plot of X_m versus x_2 would indicate that a simulation box was too small, such that bulk water was not present. Figure 1 schematically illustrates the use of multiple simulation boxes.

METHODS

Systems Studied and Molecular Dynamics Simulations. We illustrate the above method using basic pancreatic trypsin inhibitor (58 amino acids, PDB ID 5pti), abbreviated as BPTI, and hen egg white lysozyme (129 amino acids, PDB ID 4lzt), abbreviated as HEW lysozyme. These proteins are relatively small, which allows smaller simulation boxes to be used, and experimental data is available for several of the partial molar properties for comparison with the simulated results.

The initial structures for BPTI and HEW lysozyme were each centered in a series of truncated octahedron boxes where the distances between any two parallel box faces were 8, 9, 10, 11, and 12 nm. The titratable residues were protonated to correspond to a neutral pH and counterions were added to ensure neutral systems (6 Cl^- for BPTI systems, 9 Cl^- for HEW lysozyme systems). A pure water system, consisting of a truncated octahedron box (with the distance between any two parallel box faces equal to 7 nm) was used to calculate the corresponding molar properties of the pure solvent.

All simulations were performed using classical MD techniques at a temperature of 300 K and pressure of 1 bar unless otherwise noted. In an effort to ensure that the simulations would sample from the NpT ensemble and the corresponding system fluctuations would be correct, the Nosé–Hoover (chain length of one) and Parrinello–Rahman T and p baths were used.[47−50] The protein and ions were modeled using the Kirkwood–Buff Force Field (http://kbff.chem.k-state.edu, KBFF)[51−53] in explicit solvent with the SPC/E[54] water model. Following 25 ns of production simulation for each system size, the average C^{α} root mean squared deviation (RMSD) for the final BPTI (lysozyme) structures, after a translational and rotational fit to the initial structure, was 0.22 nm (0.27 nm). These values correspond to the average over the final structure from each of the different system sizes. Additional details of the simulations are provided in the Supporting Information.

Due to the promising initial results for BPTI and HEW lysozyme, and due to our overarching questions, we also asked whether the method could be used to distinguish between different conformations of a protein, that is, if it could be used to determine $\Delta_N^D\overline{X}_2^{\infty}$. To test this, we simulated a native and an extended denatured conformation of trp-cage (PDB ID 2jof, construct TCb10) using the KBFF models (see the Supporting Information). Trp-cage is a 20 amino acid designed miniprotein. The system setup was the same as for BPTI and HEW lysozyme, except that the distance between any two parallel box faces was 7, 8, 9, and 10 nm for the native conformation, and 8, 9, and 10 nm for the denatured conformation. The simulation protocol for trp-cage was similar to that for BPTI and HEW lysozyme. The specific trp-cage construct used here is neutral, so no counterions were added to the systems. Following 25 ns of production simulation, the average C^{α} RMSD for the final trp-cage native structures, after a translational and rotational fit to the initial structure, was 0.17 nm.

In addition to modeling trp-cage using the KBFF in explicit solvent with the SPC/E water model, we performed identical trp-cage simulations with the AMBER99sb force field and the TIP3P water model to allow for a force field comparison.[55] A TIP3P truncated octahedron box, with the distance between any two parallel box faces equal to 7 nm, was used to calculate the properties of pure water for the AMBER99sb simulations. Following 25 ns of production simulation, the average C^{α} RMSD for the final trp-cage native structures, after a translational and rotational fit to the initial structure, was 0.09 nm using AMBER99sb.

The initial and final denatured conformations are shown in Figure 2. Since the experimental $\Delta_N^D\overline{X}_2^{\infty}$ corresponds to a difference between properties of an ensemble of native and denatured states, our goal here was not necessarily to match the experimental value, although certainly that is of long-term interest. In this study, we instead specifically sought to establish how precisely the differences could be calculated, and to see if we achieved the correct sign and order of magnitude for the values using only two conformations.

Extracting X_m from the Simulations. The properties U_m, K_m, H_m, and V_m were extracted from the Gromacs energy file since they simply correspond to time averages of system properties. The other properties ($C_{p,m}$, κ_T, and α_p) were calculated from their bulk system fluctuations definitions in eq 7. The standard deviation for each property was calculated using block averages where the block size was 5 ns.

The system fluctuations displayed a large standard deviation (see Results and Discussion). In an effort to increase the signal-to-noise ratio, we additionally calculated these properties using the expressions in eq 6 (where the standard state superscripts should be removed). To do this, each system was run at four additional temperatures (290, 295, 305, and 310 K) and four additional pressures (250, 500, 750, and 1000 bar) for 10 ns per

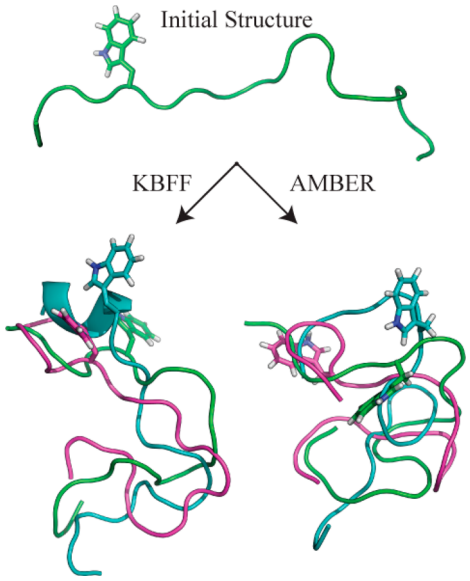


Figure 2. Trp-cage denatured conformations used in this study. The denatured conformation was obtained after clustering a 100 ns, 500 K, *NVT* simulation using the AMBER99sb force field. Subsequent simulations of this structure using KBFF and AMBER99sb resulted in the denatured conformations shown. The three structures for each force field correspond to the structures from the three compositions (i.e., the three system sizes) studied.

new state point. The $C_{p,m}$, κ_T, and α_p were then obtained by fitting the simulated molar volumes or enthalpies to a polynomial function and taking the derivatives of the polynomial at the temperature or pressure of interest (300 K or 1 bar). Quadratic fits were used for the calculation of κ_T and α_p. Linear fits were used for the calculation of $C_{p,m}$. The standard deviations for each property using this "polynomial fitting method" were calculated using block averages where the block size was again 5 ns. Although the production simulation lengths for the system fluctuation approach (25 ns) and polynomial fitting approach (10 ns per state point) are not equal and correspond to 25 ns per property for the system fluctuation approach and 50 ns per property for the polynomial fitting approach, the reported error estimates (see Results and Discussion) indicate that the polynomial fitting method greatly reduced the noise beyond the type of noise reduction one would see by doubling the simulation length alone.

RESULTS AND DISCUSSION

Figure 3 shows that ensemble average properties were well behaved for both BPTI and HEW lysozyme. In contrast, there was significant noise in the three properties that are obtained from fluctuating properties. This is clearly illustrated by the scattering of the filled points around the solid lines in Figure 4.

It is well-known that the isothermal compressibility of proteins is less than that of pure water.[15] Lines with negative slopes in the top panel of Figure 4 correspond to protein partial molar isothermal compressibilities that are less than the isothermal compressibility of pure water. If the slope is too negative (see, e.g., the system fluctuation result for BPTI), then the compressibility of the protein will actually become negative (Table 2). It is known that the isothermal compressibilities of zwitterionic amino acids are negative.[34] So, in general, this is not an impossible result. However, most reported isothermal compressibilities of proteins are positive.[34]

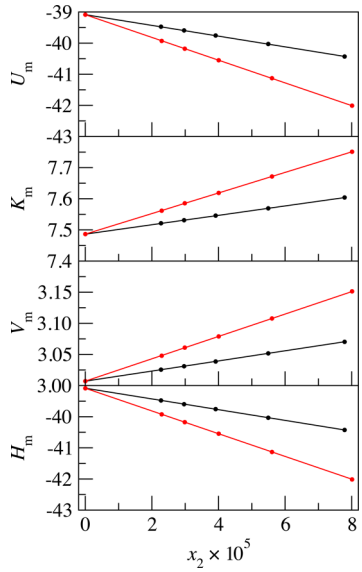


Figure 3. System properties used to calculate the infinitely dilute partial molar properties of basic pancreatic trypsin inhibitor (BPTI) and hen egg white (HEW) lysozyme in pure water at 300 K and 1 bar according to eq 14. Black: BPTI. Red: HEW lysozyme. Circles: The property at each composition. Lines: Linear fits (fixed *y*-intercept) through the set of points. The properties shown correspond to time averages from the simulations: the molar internal energy, U_m (kJmol^{-1}), kinetic energy, K_m (kJmol^{-1}), volume, V_m (nm^3 × 10^2), and enthalpy, H_m (kJmol^{-1}).

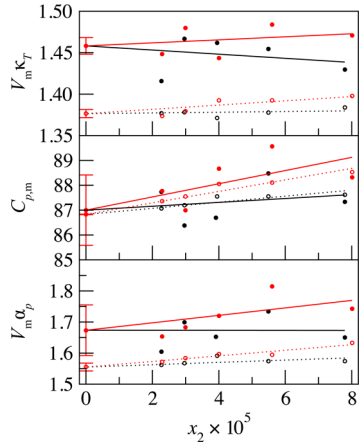


Figure 4. System properties used to calculate the infinitely dilute partial molar properties of basic pancreatic trypsin inhibitor (BPTI) and hen egg white (HEW) lysozyme in pure water at 300 K and 1 bar according to eq 14. Black: BPTI. Red: HEW lysozyme. Results from the bulk system fluctuations are shown as solid lines with filled circles, while results from the polynomial fitting method are shown as dotted lines with open circles. The property at each composition is shown as a circle, and the linear fit through the set of points, in which the *y*-intercept was fixed, is shown as a line. The top panel corresponds to $V_m\kappa_T$ (×10^6 nm^3 bar^{-1}), the middle panel to $C_{p,m}$ (J mol^{-1} K^{-1}), and the bottom panel to $V_m\alpha_p$ (×10^5 nm^3 K^{-1}).

Tables 1 and 2 show the values and standard deviations of the $\bar{X}_2^\infty$ obtained for BPTI and HEW lysozyme. While the internal energy, kinetic energy, and enthalpy of the proteins are not comparable to experiments, it may be of interest to study trends in these properties with, for example, protein mass, for a given force field. One application of such a study might be that these trends could provide an additional metric for the

dx.doi.org/10.1021/jp508632h | *J. Phys. Chem. B* 2014, 118, 12844−12854

Table 1. $\bar{K}_2^\infty$, $\bar{H}_2^\infty$, and $\bar{V}_2^\infty$ of Native BPTI and Lysozyme in Pure Water at 300 K and 1 bar[a]

		BPTI		lysozyme	
		sim	exptl/pred*	sim	exptl/pred*
$\bar{K}_2^\infty$	$\times 10^{-3}$ kJ mol^{-1}	1.515(8)	1.511*	3.309(8)	3.306*
$\bar{H}_2^\infty$	$\times 10^{-4}$ kJ mol^{-1}	−1.725(1)		−3.655(5)	
$\bar{V}_2^\infty$	nm^3	8.164(8)	7.8	18.00(3)	16.92(2)

[a]Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. The "experimental" BPTI volume is actually a calculation based upon BPTI's amino acid composition.[56] The experimental HEW lysozyme volume was measured using a solution of isoionic lysozyme in distilled and deionized water.[57] Simulations were performed using the KBFF force field.[51−53] For reference, the volume of a single SPC/E water molecule is ∼0.03 nm^3, and 1 nm^3 molecule^{-1} = 602.2 cm^3 mol^{-1}.

Table 2. $\bar{C}_{p,2}^\infty$, $\bar{\kappa}_{T,2}^\infty$, and $\bar{\alpha}_{p,2}^\infty$ of Native BPTI and Lysozyme in Pure Water at 300 K and 1 bar[a]

		BPTI			lysozyme		
		SF	PF	exptl	SF	PF	exptl
$\bar{C}_{p,2}^\infty$	kJ mol^{-1} K^{-1}	7(15)	11.2(4)	9.5	33(38)	23.39(9)	17.97, 21.22
$\bar{\kappa}_{T,2}^\infty$	$\times 10^6$ bar^{-1}	−25(32)	5(2)	$O(1-10)$	8(13)	12(4)	7.73
$\bar{\alpha}_{p,2}^\infty$	$\times 10^4$ K^{-1}	0(22)	4(4)	6	6(12)	5(1)	4.26

[a]SF, system fluctuations; PF, polynomial fitting. Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. Experimental BPTI heat capacity value from Makhatadze et al.,[58] and thermal expansion value from Lin et al.[59] Experimental lysozyme heat capacity values from Gekko and Noguchi[57] and Yang and Rupley,[60] isothermal compressibility value from Gekko and Hasegawa,[36] and thermal expansion value from Gekko and Noguchi.[57] Simulations were performed using the KBFF force field.[51−53]

Table 3. $\bar{H}_2^\infty$ and $\bar{V}_2^\infty$ of Native (N) and Denatured (D) Trp-Cage in Pure Water at 300 K and 1 bar[a]

		KBFF $\bar{X}_2^\infty$			AMBER99sb $\bar{X}_2^\infty$			exptl
		N	D	Δ_N^D	N	D	Δ_N^D	Δ_N^D
no position restraints								
$\bar{H}_2^\infty$	$\times 10^{-3}$ kJ mol^{-1}	−5.120(8)	−5.03(2)	0.09(2)	−1.817(9)	−1.77(2)	0.05(2)	0.065(2)
$\bar{V}_2^\infty$	nm^3	2.524(3)	2.496(7)	−0.028(8)	2.375(8)	2.377(6)	0.00(1)	
position restraints on denatured conformation								
$\bar{H}_2^\infty$	$\times 10^{-3}$ kJ mol^{-1}		−5.00(1)	0.12(1)		−1.74(2)	0.08(2)	0.065(2)
$\bar{V}_2^\infty$	nm^3		2.48(2)	−0.04(2)		2.37(1)	0.00(1)	

[a]Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages. The kinetic energy contributions agreed with the equipartition results and were the same, within statistical uncertainty, for the native and denatured confirmations. Experimental $\Delta_N^D\bar{H}_2^\infty$ from Barua et al.[65]

characterization of protein force fields. It is possible that such a characterization, for a given force field, could be used to distinguish between folded and unfolded conformations using thermodynamic, rather than structural, criteria. The simulated protein kinetic energy agreed with the predicted equipartition result in every case, calculated as $\bar{K}_2^\infty = 1/2k_BT(3N_{atoms} - N_{bonds})$, where N_{atoms} is the number of protein + ion atoms and N_{bonds} is the number of constrained bonds in the protein.

The simulated BPTI volume, 8.164(8) nm^3, was larger than the approximate value calculated based upon the amino acid composition, 7.8 nm^3.[61] This difference of 0.364 nm^3 corresponds to approximately 12 waters (the volume of an SPC/E water molecule is ∼0.03 nm^3). The simulated HEW lysozyme volume, 18.00(3) nm^3, was larger than the experimental volume, 16.92(2) nm^3 (experimental measurement was of isoionic lysozyme in completely deionized and distilled water at 25 °C).[57] This difference of 1.08 nm^3 for HEW lysozyme corresponds to approximately 36 waters. For both BPTI and HEW lysozyme, the disagreement with the literature results is partly due to the inclusion of counterion volumes in our simulated values (6 for BPTI, 9 for HEW lysozyme). For HEW lysozyme, this is clearly not the only issue.

Using the system fluctuation method, the fluctuating properties for both proteins displayed too much noise to determine if there was good agreement with experiment or not (Table 2). To increase the signal-to-noise ratio, we used a second approach, in which we looked at the system volume and enthalpy as a function of pressure and or temperature and calculated the κ_T, $C_{p,m}$, and α_p using their partial derivative definitions instead of their fluctuation definitions (see Theory). This greatly reduced the error (shown visually in Figure 4 and tabulated in Table 2).

The BPTI and HEW lysozyme $C_{p,m}$ overpredict the experimental values. Since our simulations were classical and there are significant quantum corrections to the heat capacity, perfect agreement with experiment, even if achieved, may have been for the wrong reasons. Specifically, real molecules store energy in low frequency bonds, whereas all bonds were constrained in our simulations. Second, in real molecules, only the low frequency bond angles would store energy, whereas all angles (except those in water) are flexible and store energy in our simulations. Quantum corrections could be made,[62] but they are difficult for proteins and have not been made here.

The thermal expansion and isothermal compressibility were in reasonable agreement with experiment, but displayed large error bars. The magnitude of the errors on the fluctuating properties is a downside of our approach that has not been observed using more subjective methods.[15,16]

Table 4. $\bar{C}^{\infty}_{p,2}$, $\bar{\kappa}^{\infty}_{T,2}$, and $\bar{\alpha}^{\infty}_{p,2}$ of Native (N) and Denatured (D) Trp-Cage in Pure Water at 300 K and 1 bar[a]

		KBFF $\bar{X}^{\infty}_2$			AMBER99sb $\bar{X}^{\infty}_2$		
		N	D	Δ^{D}_{N}	N	D	Δ^{D}_{N}
$\bar{C}^{\infty}_{p,2}$	kJ mol^{-1} K^{-1}						
exptl				−0.2(1)			−0.2(1)
SF		14(21)	23(18)	9(28)	19(6)	18(16)	−1(17)
PF		3.6(1)	3.5(5)	−0.1(5)	5.8(5)	6.4(1)	0.6(5)
PF PRD		3.6(1)	3.23(5)	−0.4(2)	5.8(5)	6.1(4)	0.3(6)
PF (1 μs)					6.2(1)		
$\bar{\kappa}^{\infty}_{T,2}$	×10^{6} bar^{-1}						
SF		9(72)	−100(100)	−100(100)	−78(168)	65(160)	144(233)
PF		3(8)	33(11)	29(13)	−16(10)	−12(9)	−9(10)
$\bar{\alpha}^{\infty}_{p,2}$	×10^{4} K^{-1}						
SF		30(40)	20(40)	−10(60)	−10(50)	40(70)	50(90)
PF		6(6)	8(15)	2(16)	7(2)	6.7(2)	−1(2)
PF PRD		6(6)	8(5)	2(8)	7(2)	6(3)	−2(4)
PF (1 μs)					8.6(4)		

[a]Error estimates are reported in parentheses and correspond to the standard deviation from 5 ns block averages, excluding the 1 μs simulation, which corresponds to the standard deviation on ∼333 ns block averages. For the 1 μs simulations, the error on the y-intercept (pure water value) was taken as zero. SF: System Fluctuations, PF: Polynomial Fitting. PRD: Position Restrained Denatured conformation. Experimental $\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ from Barua et al.[65]

Literature values are available for all of the pure water (SPC/E) properties. However, since it is critical that all of the simulation conditions are consistent, we simulated a pure SPC/E system instead of simply quoting previously reported SPC/E properties. The reason for the difference between the pure SPC/E values obtained using the system fluctuation method versus the polynomial fitting method for the isothermal compressibility is currently unknown (Figure 4). However, the value obtained using the polynomial fitting method is in better agreement with experiment [polynomial fitting, 4.48(2) × 10^{-5} bar^{-1}; system fluctuation, 4.75(3) × 10^{-5} bar^{-1}; experiment,[63] 4.525 × 10^{-5} bar^{-1}].

The results for the trp-cage simulations with both force fields and with both methods (the system fluctuation and the polynomial fitting method) are summarized in Tables 3 and 4. As shown in Table 3, the KBFF results suggest that the native state has a larger volume than the denatured state by a value of approximately one water molecule or ∼1% of the protein volume, regardless of whether or not the denatured conformation was position restrained. This is the typical sign observed for the volume change of a protein upon denaturation.[64] In contrast, the AMBER99sb simulations produced no statistically significant volume change.

The KBFF simulation produced the correct sign for the change in enthalpy upon denaturation. The enthalpy of the denatured state was more positive, or unfavorable, than the enthalpy of the native state, as expected for heat denaturation.[65,66] The average KBFF result was in better agreement with experiment when the position restraints were removed, although the difference was not statistically significant and the native state was slightly overstabilized in both cases. Since the starting conformation for both simulations was taken from a 500 K simulation using the AMBER99sb force field (see the Supporting Information), better agreement with experiment may be achieved if the KBFF simulations were extended to allow for more conformational rearrangements. The AMBER99sb result was within the error of the experimental value with or without position restraints on the denatured conformation.

Globular protein heat capacity changes upon denaturation are typically positive; a positive (negative) change is considered to be due to apolar (polar) solvation upon denaturation.[2] Since trp-cage is considered globular and its name stems from its buried tryptophan side chain, one might predict that the sign of $\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ would be positive. Indeed, the experimental $\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ for the original trp-cage construct (TC5b) was reported to be small and positive (0.3 ± 0.1 kJ mol^{-1} K^{-1}).[67] However, the experimental results of Barua et al. were small and negative ($\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ = −0.2 ± 0.1 kJ mol^{-1} K^{-1}) for the trp-cage construct simulated here (TC10b),[65] indicating that considering tryptophan exposure is not enough to predict the correct sign.

While the statistical uncertainty in our results was too great to determine whether or not the correct sign was obtained, our results may suggest that the KBFF and AMBER99sb simulations provide opposite signs (negative and positive, respectively). It is encouraging that the KBFF simulations may be able to detect this atypical $\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ sign. The positive sign for the AMBER results reported here agrees with the recent 1 μs per replica REMD simulations of the TC10b construct by English and Garcia, who also used the AMBER99sb/TIP3P force fields and reported $\Delta^{D}_{N}\bar{C}^{\infty}_{p,2}$ = 0.35 ± 0.51 kJ mol^{-1} K^{-1}.[68]

The results for the isothermal compressibility and thermal expansion changes upon denaturation are too noisy to determine the simulated sign; however, the error was greatly reduced using the polynomial fitting method.

Visualization of the denatured trp-cage trajectories showed that the KBFF and AMBER99sb denatured simulations without position restraints behaved differently from each other. The KBFF denatured trp-cage conformation remained more fully extended throughout the 25 ns, whereas the AMBER99sb denatured trp-cage collapsed quickly and remained collapsed for most of the simulation (although it became more extended at the very end of the simulation). The final snapshot of the denatured simulations is shown in Figure 2. Despite this, the two force fields gave similar enthalpy differences [experiment, $\Delta^{D}_{N}\bar{H}^{\infty}_{2}$ = 65(2) kJ mol^{-1}; KBFF, $\Delta^{D}_{N}\bar{H}^{\infty}_{2}$ = 90(20) kJ mol^{-1}; AMBER99sb, $\Delta^{D}_{N}\bar{H}^{\infty}_{2}$ = 50(20) kJ mol^{-1}]. Certainly, the simulations did not access all the denatured conformations with

dx.doi.org/10.1021/jp508632h | J. Phys. Chem. B 2014, 118, 12844−12854

only 25 ns of production, so it was remarkable that the $\Delta_N^D\overline{H}_2^\infty$ values were so close to the experimental value.

To test whether the different denatured conformations that were sampled in the simulations of trp-cage with these two different force fields accounted for the different enthalpy differences, we ran a new set of denatured simulations for both FFs using soft harmonic position restraints (100 kJ mol^{-1} nm^{-2}) on the C^α atoms of trp-cage. The average position restraint energy was small, approximately 2 J mol^{-1}, so we did not repeat the simulations of the native trp-cage with position restraints. Thus, we assumed that any artifacts from the addition of the position restraints were negligible when calculating the $\Delta_N^D\overline{X}_2^\infty$ using simulations of the native structure without position restraints. The $\Delta_N^D\overline{X}_2^\infty$ results were the same with and without position restraints to within the statistical error. This is consistent with a two-state folding pathway in which the energy landscape is relatively flat for the (high free energy) unfolded conformations and drops down into one minimum for the folded conformation. Remarkably, the enthalpy differences remained essentially correct when calculated from the difference between the native conformation and just one restrained denatured conformation.

Error Analysis. Lin et al. and references therein have reported that typical magnitudes for the volume change associated with protein unfolding are <0.5% of the protein volume.[59] Using pressure perturbation calorimetry, Lin determined the $\Delta_N^D\overline{V}_2^\infty$ value was −0.06% of the volume of native BPTI in water at pH 4 and −0.08% of the volume of native HEW lysozyme in water at pH 2.5.[59] The errors on our $\Delta_N^D\overline{V}_2^\infty$ value for BPTI and HEW lysozyme were 0.1% and 0.2%; thus, simulation times longer than the currently used times would be necessary to calculate $\Delta_N^D\overline{V}_2^\infty$ with precision for these proteins. The corresponding $\Delta_N^D\overline{\alpha}_{p,2}^\infty$ values reported by Lin were 0.5×10^{-4} K^{-1} and 1.4×10^{-4} K^{-1} for BPTI and HEW lysozyme, respectively. To calculate $\Delta_N^D\overline{\alpha}_{p,2}^\infty$ with sufficient precision using the polynomial fitting method, we would need to reduce the error on $\overline{\alpha}_{p,2}^\infty$ by a factor of 8 for BPTI, which would require simulations approximately 64 times as long, that is, 640 ns at each temperature, assuming that the error $\propto t_{sim}^{-1/2}$. With the current simulation time, the $\overline{\alpha}_{p,2}^\infty$ value for HEW lysozyme was at the minimal precision necessary for detection of $\Delta_N^D\overline{\alpha}_{p,2}^\infty$, if simulations of denatured conformations were performed.

In a compendium of protein thermodynamic properties Makhatadze has reported that BPTI $\Delta_N^D\overline{H}_2^\infty$ = 130 kJ mol^{-1} and $\Delta_N^D\overline{C}_{p,2}^\infty$ = 3.0 kJ K^{-1} mol^{-1} at 298 K.[69] These numbers are larger than the uncertainty on the native enthalpy and native heat capacity calculated here, 10 kJ mol^{-1} and 0.4 kJ mol^{-1} K^{-1}, respectively. The same source reports that HEW lysozyme $\Delta_N^D\overline{H}_2^\infty$ = 242 kJ mol^{-1} and $\Delta_N^D\overline{C}_{p,2}^\infty$ = 9.1 kJ mol^{-1} K^{-1} at 298 K.[69] Again, these numbers are larger than the uncertainty on the native HEW lysozyme enthalpy and heat capacity determined here, 50 kJ mol^{-1} and 0.09 kJ mol^{-1} K^{-1}, respectively. Thus, we predict that similar length simulations of either of these proteins in their denatured conformations would allow for $\Delta_N^D\overline{H}_2^\infty$ and $\Delta_N^D\overline{C}_{p,2}^\infty$ to be calculated with statistical precision. The challenge in these situations, heightened as the number of residues in the protein of interest increases, would be in choosing protein conformations that were "representative" of the denatured state. Although the trp-cage results from these simulations surprisingly showed that the values were not sensitive to the protein conformation, much more work on different proteins would be necessary before confirming that behavior as a general conclusion. Indeed, one would anticipate that it is not generally true. However, Shaw and co-workers recently reported similar observations for ubiquitin.[70]

We could have simulated all of the systems for even longer periods of time. However, even a multiplication of the production simulation length by four would only result in a noise reduction by a factor of 2 (assuming that the error $\propto t_{sim}^{-1/2}$), which would not be sufficient in most cases to determine these properties with sufficient precision. So, in an effort to estimate how long one would need to simulate trp-cage, in order to obtain the heat capacity and thermal expansion with sufficient precisely, we extended the AMBER99sb native trp-cage simulations to 1 μs at temperatures of 290, 295, 300, 305, and 310 K. Since we had already established the linear behavior of X_m versus x_2 for the 7, 8, 9, and 10 nm box sizes, we only extended the simulations for the 8 nm box. Extension of a simulation from 10 ns to 1 μs is an increase of t_{sim} by a factor of 100; thus, we would have predicted a reduction in the noise by a factor of 10. As shown in Table 4, the errors were actually reduced by a factor of 5 for $\overline{C}_{p,2}^\infty$ and only a factor of 2.5 for $\overline{\alpha}_{p,2}^\infty$.

CONCLUSIONS

We have investigated the ability of a simple method to extract infinitely dilute partial molar properties of biomolecules from molecular dynamics simulations using basic pancreatic trypsin inhibitor and hen egg white lysozyme as our test case proteins. We then used the method to distinguish between the protein volume and enthalpy of a native and a denatured trp-cage conformation with sufficient precision. The method is most similar to that used recently by Smit to determine the enthalpy of a lipid aggregate, and to Shing and Chung's work in the late 1980s for the infinitely dilute partial molar volume and internal energy of a Lennard−Jones solute,[21,71] but to our knowledge it has never been used or presented for any general solute and for any general partial molar property. The strengths of the approach are that it is thermodynamically exact, no ambiguous parameters are introduced, and knowledge of the specific conformations that contributed to the output values is retained. The method is general, and this approach may be used to calculate any partial molar property of any solute in any solution. It could be employed in the future as a test of the quality of a FF.

The current limitation of the approach is the high level of noise in the isothermal compressibility and thermal expansion coefficient results, and the moderate level of noise in the isobaric heat capacity results. To increase the signal-to-noise ratio, a polynomial fitting method was used. Clearly, the system fluctuation approach was not the method of choice for the calculation of the fluctuating properties, and the polynomial fitting method provided higher precision results. However, we were still unable to precisely distinguish between native and denatured trp-cage conformations' heat capacities, thermal expansions, and isothermal compressibilities.

Based upon our results, we predict that for a protein the size of HEW lysozyme (129 residues), the following amounts of time are needed to calculate specific thermodynamic properties with statistical precision: $\Delta_N^D\overline{V}_2^\infty$, ∼200 ns; $\Delta_N^D\overline{H}_2^\infty$, achieved necessary precision using 25 ns simulation; $\Delta_N^D\overline{C}_{p,2}^\infty$, achieved necessary precision using 10 ns simulations at each of 5 temperatures; $\Delta_N^D\overline{\kappa}_{T,2}^\infty$, difficult to predict due to the large noise; $\Delta_N^D\overline{\alpha}_{p,2}^\infty$, achieved the necessary precision using 10 ns simulations at each of 5 temperatures. Currently, the methods

that rely upon subjective definitions, if they can be trusted, are more competitive than our approach for the isothermal compressibility due to the high amounts of noise in the current method. Continued increases in computational power should make the method presented here more usable, with ease, over time.

Lastly, the method presented here does not provide a means of directly decomposing the thermodynamic properties into contributions, from specific surface groups for example. The solute−solute and solute−solvent energy terms decay as a function of distance away from the solute, which makes them directly calculable. Since the solvent−solvent contribution does not decay, it must be solved for indirectly. This is a commonly encountered issue; see, for example, McCammon and co-workers.[72] We are currently working to develop and test a method to calculate all of the above properties that uses expressions from Fluctuation Solution Theory and is decomposable, analogous to the group contribution decomposition for partial molar volumes already achieved using traditional Kirkwood−Buff theory.[23,30]

■ ASSOCIATED CONTENT

Supporting Information

Additional details of the simulation methods. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 785-532-5109. Fax: 785-532-6666. E-mail: pesmith@ksu.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The project described was supported by Grant R01GM079277 from the National Institute of General Medical Sciences to P.E.S. E.A.P. was supported by the KSU NSF GK-12 program under Grant NSF DGE-0841414 and the NSF GRF program under Grant NSF DGE-0750823. 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 Institutes of Health, or the National Science Foundation.

■ REFERENCES

(1) Holthauzen, L. M. F.; Auton, M.; Sinev, M.; Rosgen, J. Protein stability in the presence of cosolutes. *Methods Enzymol.* **2011**, *492*, 61−125.

(2) Prabhu, N. V.; Sharp, K. A. Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **2005**, *56*, 521−548.

(3) Rodgers, J. M.; Sorensen, J.; de Meyer, F. J. M.; Schiott, B.; Smit, B. Understanding the phase behavior of coarse-grained model lipid bilayers through computational calorimetry. *J. Phys. Chem. B* **2012**, *116*, 1551−1569.

(4) Davis-Searles, P. R.; Saunders, A. J.; Erie, D. A.; Winzor, D. J.; Pielak, G. J. Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 271−306.

(5) Jiao, Y. F.; Smith, P. E. Fluctuation theory of molecular association and conformational equilibria. *J. Chem. Phys.* **2011**, *135*, 014502.

(6) Smith, P. E. Local chemical potential equalization model for cosolvent effects on biomolecular equilibria. *J. Phys. Chem. B* **2004**, *108*, 16271−16278.

(7) Smith, P. E. Cosolvent interactions with biomolecules: Relating computer simulation data to experimental thermodynamic data. *J. Phys. Chem. B* **2004**, *108*, 18716−18724.

(8) Ben-Naim, A. Theoretical aspects of pressure and solute denaturation of proteins: A Kirkwood-Buff-theory approach. *J. Chem. Phys.* **2012**, *137*, 235102.

(9) Ben-Naim, A. Y. *Statistical thermodynamics for chemists and biochemists*; Plenum Press: New York, 1992.

(10) Glasstone, S. *Thermodynamics for chemists*; D. Van Nostrand Company, Inc.: New York, 1947.

(11) Majer, V.; Sedlbauer, J.; Wood, R. H. Calculation of standard thermodynamic properties of aqueous electrolytes and non-electrolytes. In *Aqueous Systems at Elevated Temperatures and Pressures: Physical Chemistry in Water, Steam, and Hydrothermal Solutions*; Palmer, D. A., Fernandez-Prini, R., Harvey, A. H., Eds. Elsevier: Amsterdam, 2004; pp 99−147.

(12) Lewis, G. N. Outlines of a new system of thermodynamic chemistry. *Proc. Am. Acad. Arts Sci.* **1907**, *43*, 259−293.

(13) Canchi, D. R.; Paschek, D.; Garcia, A. E. Equilibrium study of protein denaturation by urea. *J. Am. Chem. Soc.* **2010**, *132*, 2338−2344.

(14) Murphy, L. R.; Matubayasi, N.; Payne, V. A.; Levy, R. M. Protein hydration and unfolding: Insights from experimental partial specific volumes and unfolded protein models. *Folding Des.* **1998**, *3*, 105−118.

(15) Dadarlat, V. M.; Post, C. B. Insights into protein compressibility from molecular dynamics simulations. *J. Phys. Chem. B* **2001**, *105*, 715−724.

(16) Dadarlat, V. M.; Post, C. B. Decomposition of protein experimental compressibility into intrinsic and hydration shell contributions. *Biophys. J.* **2006**, *91*, 4544−4554.

(17) Mitra, L.; Smolin, N.; Ravindra, R.; Royer, C.; Winter, R. Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution: Experiments and theoretical interpretation. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1249−1265.

(18) Sarupria, S.; Ghosh, T.; Garcia, A. E.; Garde, S. Studying pressure denaturation of a protein by molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1641−1651.

(19) DeVane, R.; Ridley, C.; Larsen, R. W.; Space, B.; Moore, P. B.; Chan, S. I. A molecular dynamics method for calculating molecular volume changes appropriate for biomolecular simulation. *Biophys. J.* **2003**, *85*, 2801−2807.

(20) Patel, N.; Dubins, D. N.; Pomes, R.; Chalikian, T. V. Parsing partial molar volumes of small molecules: A molecular dynamics study. *J. Phys. Chem. B* **2011**, *115*, 4856−4862.

(21) Shing, K. S.; Chung, S. T. Calculation of infinite-dilution partial molar properties by computer simulation. *AlChE J.* **1988**, *34*, 1973−1980.

(22) Surampudi, L. N.; Ashbaugh, H. S. Direct evaluation of polypeptide partial molar volumes in water using molecular dynamics simulations. *J. Chem. Eng. Data* **2014**, *59*, 3130−3135.

(23) Meng, B.; Ashbaugh, H. S. Pressure reentrant assembly: Direct simulation of volumes of micellization. *Langmuir* **2013**, *29*, 14743−14747.

(24) Floris, F. M. Nonideal effects on the excess volume from small to large cavities in TIP4P water. *J. Phys. Chem. B* **2004**, *108*, 16244−16249.

(25) Ashbaugh, H. S.; Truskett, T. M. Putting the squeeze on cavities in liquids: Quantifying pressure effects on solvation using simulations and scaled-particle theory. *J. Chem. Phys.* **2011**, *134*, 014507.

(26) Franks, F.; Johnson, H. H. Accurate evaluation of partial molar properties. *Trans. Faraday Soc.* **1962**, *58*, 656−661.

(27) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical theory of solutions. I. *J. Chem. Phys.* **1951**, *19*, 774−777.
(28) 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*, 12392−12397.
(29) Patel, N.; Dubins, D. N.; Pomes, R.; Chalikian, T. V. Size dependence of cavity volume: A molecular dynamics study. *Biophys. Chem.* **2012**, *161*, 46−49.
(30) Sangwai, A. V.; Ashbaugh, H. S. Aqueous partial molar volumes from simulation, and individual group contributions. *Ind. Eng. Chem. Res.* **2008**, *47*, 5169−5174.
(31) Imai, T.; Hirata, F. Partial molar volume and compressibility of a molecule with internal degrees of freedom. *J. Chem. Phys.* **2003**, *119*, 5623−5631.
(32) Imai, T.; Kinoshita, M.; Hirata, F. Theoretical study for partial molar volume of amino acids in aqueous solution: Implication of ideal fluctuation volume. *J. Chem. Phys.* **2000**, *112*, 9469−9478.
(33) Ploetz, E. A.; Smith, P. E. Local fluctuations in solution: Theory and applications. In *Advances in Chemical Physics*; Rice, S. A., Dinner, A. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 2013; pp 311−372.
(34) Høiland, H.; Hedwig, G. R. 6 Compressibilities of amino acids, peptides and proteins in aqueous solution. In *SpringerMaterials-The Landolt-Börnstein Database*; Springer: New York, **2003**; pp 1−17.
(35) Cooper, A. Thermodynamic fluctuations in protein molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 2740−2741.
(36) Gekko, K.; Hasegawa, Y. Compressibility structure relationship of globular proteins. *Biochemistry* **1986**, *25*, 6563−6571.
(37) Marchi, M. Compressibility of cavities and biological water from Voronoi volumes in hydrated proteins. *J. Phys. Chem. B* **2003**, *107*, 6598−6602.
(38) Lopez, C. F.; Darst, R. K.; Rossky, P. J. Mechanistic elements of protein cold denaturation. *J. Phys. Chem. B* **2008**, *112*, 5961−5967.
(39) Rosgen, J.; Hinz, H. J. Response functions of proteins. *Biophys. Chem.* **2000**, *83*, 61−71.
(40) Hallerbach, B.; Hinz, H. J. Protein heat capacity: Inconsistencies in the current view of cold denaturation. *Biophys. Chem.* **1999**, *76*, 219−227.
(41) Chalikian, T. V. Volumetric measurements in binary solvents: Theory to experiment. *Biophys. Chem.* **2011**, *156*, 3−12.
(42) Cooper, A. Protein heat capacity: An anomaly that maybe never was. *J. Phys. Chem. Lett.* **2010**, *1*, 3298−3304.
(43) Mitra, L.; Oleinikova, A.; Winter, R. Intrinsic volumetric properties of trialanine isomers in aqueous solution. *ChemPhysChem* **2008**, *9*, 2779−2784.
(44) Brovchenko, I.; Andrews, M. N.; Oleinikova, A. Volumetric properties of human islet amyloid polypeptide in liquid water. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4233−4238.
(45) Medvedev, N. N.; Voloshin, V. P.; Kim, A. V.; Anikeenko, A. V.; Geiger, A. Culation of partial molar volume and its components for molecular dynamics models of dilute solutions. *J. Struct. Chem.* **2013**, *54*, 271−288.
(46) Kim, A. V.; Medvedev, N. N.; Geiger, A. Molecular dynamics study of the volumetric and hydrophobic properties of the amphiphilic molecule C8E6. *J. Mol. Liq.* **2014**, *189*, 74−80.
(47) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(48) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(49) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(50) Nosé, S.; Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(51) Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. Developing force fields from the microscopic structure of solutions: The Kirkwood−Buff approach. In *Modeling Solvent Environments*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2010; pp 55−76.
(52) Ploetz, E. A.; Bentenitis, N.; Smith, P. E. Developing force fields from the microscopic structure of solutions. *Fluid Phase Equilib.* **2010**, *290*, 43−47.
(53) Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. Recent applications of Kirkwood-Buff theory to biological systems. *Cell Biochem. Biophys.* **2008**, *50*, 1−22.
(54) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(55) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(56) Chiou, S. H.; Azari, P.; Himmel, M. E.; Squire, P. G. Isolation and physical characterization of bovine lens crystallins. *Int. J. Pept. Protein Res.* **1979**, *13*, 409−417.
(57) Gekko, K.; Noguchi, H. Compressibility of globular proteins in water at 25 °C. *J. Phys. Chem.* **1979**, *83*, 2706−2714.
(58) Makhatadze, G. I.; Kim, K. S.; Woodward, C.; Privalov, P. L. Thermodynamics of BPTI folding. *Protein Sci.* **1993**, *2*, 2028−2036.
(59) 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.* **2002**, *302*, 144−160.
(60) Yang, P. H.; Rupley, J. A. Protein-water interactions: Heat capacity of the Lysozyme-water system. *Biochemistry* **1979**, *18*, 2654−2661.
(61) Squire, P. G.; Himmel, M. E. Hydrodynamics and protein hydration. *Arch. Biochem. Biophys.* **1979**, *196*, 165−177.
(62) Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R. Thermodynamics and quantum corrections from molecular-dynamics for liquid water. *J. Chem. Phys.* **1983**, *79*, 2375−2389.
(63) Vedamuthu, M.; Singh, S.; Robinson, G. W. Properties of liquid water. 4. The isothermal compressibility minimum near 50 °C. *J. Phys. Chem.* **1995**, *99*, 9263−9267.
(64) Frye, K. J.; Royer, C. A. Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. *Protein Sci.* **1998**, *7*, 2217−2222.
(65) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-cage: Optimizing the stability of a globular miniprotein. *Prot. Eng., Des. Sel.* **2008**, *21*, 171−185.
(66) Romero-Romero, M. L.; Ingles-Prieto, A.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M. Highly anomalous energetics of protein cold denaturation linked to folding-unfolding kinetics. *PLoS One* **2011**, *6*, e23050.
(67) Streicher, W.; Makhatadze, G. Unfolding thermodynamics of Trp-cage, a 20 residue miniprotein, studied by differential scanning calorimetry and circular dichroism spectroscopy. *Biochemistry* **2007**, *46*, 2876−2880.
(68) English, C. A.; Garcia, A. E. Folding and unfolding thermodynamics of the TC10b Trp-cage miniprotein. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2748−2757.
(69) Makhatadze, G. I. Thermodynamic properties of proteins. In *Physical Properties of Polymers Handbook*; Mark, J. E., Ed.; Springer: New York, 2007; pp 103−143.
(70) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5915−5920.
(71) Rogers, P. S. Z.; Pitzer, K. S. Volumetric properties of aqueous sodium chloride solutions. *J. Phys. Chem. Ref. Data* **1982**, *11*, 15−81.
(72) Setny, P.; Baron, R.; McCammon, J. A. How can hydrophobic association be enthalpy driven? *J. Chem. Theory Comput.* **2010**, *6*, 2866−2871.