Decomposition of Protein Experimental Compressibility into Intrinsic and Hydration Shell Contributions

Voichita M. Dadarlat and Carol Beth Post
Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology, Purdue Cancer Center, Purdue University, West Lafayette, Indiana

ABSTRACT The experimental determination of protein compressibility reflects both the protein intrinsic compressibility and the difference between the compressibility of water in the protein hydration shell and bulk water. We use molecular dynamics simulations to explore the dependence of the isothermal compressibility of the hydration shell surrounding globular proteins on differential contributions from charged, polar, and apolar protein-water interfaces. The compressibility of water in the protein hydration shell is accounted for by a linear combination of contributions from charged, polar, and apolar solvent-accessible surfaces. The results provide a formula for the deconvolution of experimental data into intrinsic and hydration contributions when a protein of known structure is investigated. The physical basis for the model is the variation in water density shown by the surface-specific radial distribution functions of water molecules around globular proteins. The compressibility of water hydrating charged atoms is lower than bulk water compressibility, the compressibility of water hydrating apolar atoms is somewhat larger than bulk water compressibility, and the compressibility of water around polar atoms is about the same as the compressibility of bulk water. We also assess whether hydration water compressibility determined from small compound data can be used to estimate the compressibility of hydration water surrounding proteins. The results, based on an analysis from four dipeptide solutions, indicate that small compound data cannot be used directly to estimate the compressibility of hydration water surrounding proteins.

INTRODUCTION

The change in physical properties of water in the vicinity of biological molecules is of interest to both experimentalists and computational biologists ((1) and references therein). Protein compressibility obtained from sound velocity measurements is a combination of protein intrinsic compressibility and the change in the compressibility of water in the protein hydration shell. Separating the two contributions is valuable for characterizing both the internal dynamics and flexibility of protein molecules (2,3), as well as the physical properties of water at the protein surface. Protein compressibility measurements provide unique information on the general thermodynamic properties of proteins (4–7), including insights into the behavior of the protein interior (8), conformational transitions (9,10), and molecular interactions (11,12).

In this article, we devise a method to estimate the compressibility of hydration water as a function of the varied chemical nature of the protein-water interface. A description of hydration based on the protein surface composition is attractive because it reflects the chemical nature of the protein-water interface (13). The average solvent-exposed atomic-type composition estimated for a set of 16 globular proteins includes not only charged (12.2%) and polar (37.4%) atoms, but also a substantial number of apolar atoms (50.4%) (14). One argument in favor of an approach based on surface composition is the fact that when only a limited number of water molecules is available, the molecules do not distribute uniformly on the protein solvent-accessible surface. Molecular dynamics simulations have found that 350 water molecules that fully hydrate myoglobin, originally distributed uniformly on the protein surface, arrange in such a way that they hydrate every charged group and cover 83% of the charged surface, 65% of the hydrophilic surface, and only 53% of the hydrophobic surface (15).

Protein-water interactions are fundamental to the description of protein physical properties (1,16,17). The hydration properties of water play an important role in the formation of the three-dimensional native structures of proteins and nucleic acids, and the dynamics of biological molecules (18). Proteins do not perform their functions until a critical hydration level is achieved, and in some cases, the level of hydration that is required to restore enzymatic activity is less than one layer of water molecules (19). A detailed description of the structure and properties of water in the protein hydration shell is essential for our understanding of the processes taking place in the interior of a biological cell, a gel-like matrix (20,21) where most water molecules are within a few solvation layers of proteins and other macromolecules.

There has been extensive work through both experimental (22–26) and computer simulations methods (27–32) to explain how solutes, and proteins in particular, change the properties of water layers adjacent to their surface. Theories of hydrophobic hydration (33–36) relate the structural ordering of water by the nonpolar solutes to the hydration thermodynamics through solute-water pair correlations. The special properties of water molecules in the hydration shell

Submitted April 24, 2006, and accepted for publication August 31, 2006.
Address reprint requests to C. B. Post, E-mail: cbp@purdue.edu.

0006-3495/06/12/4544/11 $2.00
doi: 10.1529/biophysj.106.087726

of proteins are reflected in the compressibility and density, when compared to bulk water. Others have estimated the compressibility of hydration water (5,37–42) to be, on average, 22% lower than bulk water compressibility. Data from small-angle x-ray scattering combined with neutron scattering studies of lysozyme, thioredoxin reductase, and protein R1 (24) show that the average density of the water in the protein first hydration shell is 10% higher than the density of bulk water, a result that is consistent with experimental findings from x-ray crystallography (43). The estimated values for the hydration water compressibility and density are consistent with each other in that higher density restricts density fluctuations and therefore lowers compressibility.

We report here a model for hydration water compressibility derived from analysis of molecular dynamics (MD) simulations of protein solutions. The approach for estimating protein compressibility from volume fluctuations was established previously based on MD simulations for four globular proteins: trypsin, ribonuclease A, hen egg white (HEW) lysozyme, and α-lactalbumin (44). In our model, the residual compressibility of the water in the protein hydration shell is accounted for by a linear combination of contributions from charged, polar, and apolar solvent-accessible surfaces. The physical basis for the hydration water compressibility model is the variation in water density indicated by surface-specific radial distribution functions (RDFs) of water around charged, polar, and apolar protein surfaces. The outcome is a formula to interpret experimental data on protein solution compressibility by deconvolution of intrinsic protein and hydration contribution when a protein of known structure is investigated. Further, we compare the hydration water compressibility for solvation of proteins with that of small peptides to assess whether hydration water compressibility determined from small compound data can be used to estimate the compressibility of hydration water surrounding proteins. The results indicate that small compound data cannot be directly used to estimate the compressibility of hydration water surrounding proteins.

METHODS

Theoretical methods

The isothermal compressibility of a protein solution, β_T^{sol}, is defined as the relative change in the solution volume, V^{sol}, under pressure (P) at constant temperature (T):

$$\beta_T^{sol} = -\frac{1}{V^{sol}}\left(\frac{\partial V^{sol}}{\partial P}\right)_T. \quad (1)$$

The total volume is a sum of contributions from the protein molecular volume, V^p, the volume of the bulk water, V^w, and the volume of hydration water, V^h, as follows:

$$V^{sol} = V^p + V^h + V^w.$$

Equation 1 can be rewritten as a sum over component volume contributions,

$$\beta_T^{sol} = -\frac{1}{V^{sol}}\left[\left(\frac{\partial V^p}{\partial P}\right)_T + \left(\frac{\partial V^h}{\partial P}\right)_T + \left(\frac{\partial V^w}{\partial P}\right)_T\right], \quad (2)$$

and again as

$$\beta_T^{sol} = \frac{V^p}{V^{sol}}\left[-\frac{1}{V^p}\left(\frac{\partial V^p}{\partial P}\right)_T\right] + \frac{V^h}{V^{sol}}\left[-\frac{1}{V^h}\left(\frac{\partial V^h}{\partial P}\right)_T\right] + \frac{V^w}{V^{sol}}\left[-\frac{1}{V^w}\left(\frac{\partial V^w}{\partial P}\right)_T\right]. \quad (3)$$

The ratios between the solution volume components and the total solution volume are the volume fractions: $\Phi^p = \frac{V^p}{V^{sol}}\sqrt{b^2 - 4ac}$, $\Phi^h = \frac{V^h}{V^{sol}}$, and $\Phi^w = \frac{V^h}{V^{sol}}$. In Eq. 3, we identify the protein compressibility,

$$\beta_T^p = -\frac{1}{V^p}\left(\frac{\partial V^p}{\partial P}\right)_T,$$

hydration water compressibility,

$$\beta_T^h = -\frac{1}{V^h}\left(\frac{\partial V^h}{\partial P}\right)_T,$$

and bulk water compressibility,

$$\beta_T^w = -\frac{1}{V^w}\left(\frac{\partial V^w}{\partial P}\right)_T.$$

Finally, Eq. 2 for the compressibility of the protein solution becomes

$$\beta_T^{sol} = \Phi^p\beta_T^p + \Phi^h\beta_T^h + \Phi^w\beta_T^w. \quad (4)$$

The hydration contribution to solution compressibility is estimated assuming that all water in the protein solution has the properties of bulk water, and then compare this estimate with the actual β_T^{sol} (Eq. 4). In this procedure, $\beta_T^h = \beta_T^w$ and the ideal solution compressibility is

$$\beta_T^0 = \Phi^p\beta_T^p + (\Phi^h + \Phi^w)\beta_T^w. \quad (5)$$

Any difference between the actual (Eq. 4) and the ideal (Eq. 5) solution compressibility,

$$\Delta\beta_T^{sol} = \Phi^h(\beta_T^h - \beta_T^w) = \Phi^h\Delta\beta_T^h, \quad (6)$$

is due to the fact that hydration water has different compressibility than bulk water. The value $\Delta\beta_T^h$ is the difference between the compressibility of hydration and bulk water, respectively.

For the experimental determination of protein compressibility, the adiabatic compressibility of a solution is given by the Laplace equation, $\beta_S = \frac{1}{\rho u^2}$, where ρ is the density of the solution and u is the velocity of sound through the solution. The experimental, apparent partial adiabatic compressibility of a protein, is defined as the limit at 0 protein concentration of the following expression (38):

$$\beta_S^{exp} = \frac{\beta_0^w}{V_0^p}\lim_{c\to 0}\frac{\frac{\beta_S^{sol}}{\beta_0^w} - \Phi_w^0}{c}. \quad (7)$$

β_S^{sol} is the adiabatic solution compressibility, β_0^w is the compressibility of water, V_0^p is the partial specific volume of the protein, and Φ_w^0 is the apparent volume fraction of the solvent in solution. The apparent protein compressibility has contributions from both intrinsic protein compressibility and a hydration contribution.

The adiabatic compressibility is related to isothermal compressibility (5):

$$\beta_T^{exp} = \beta_S^{exp} + f(T, \alpha, \alpha_0, C_P, C_{P0}). \quad (8)$$

Here f is a function of the thermal expansion coefficients of water and protein solutions, α and α_0, and the heat capacity of water and protein solutions, C_P and C_{P0}, respectively.

The following relationship (45,46) may be used to decompose the apparent experimental protein compressibility into an intrinsic protein compressibility, β_T^p, and a hydration water contribution:

$$\beta_T^{exp} = \beta_T^p + \frac{\Phi^h}{\Phi^p}(\beta_T^h - \beta_T^w). \quad (9)$$

The hydration contribution reflects the change in the compressibility of the solvent as a result of interactions with the surface-exposed protein atoms. Finally, protein intrinsic compressibility, β_T^p, can be calculated as

$$\beta_T^p = \beta_T^{exp} - \frac{\Phi^h}{\Phi^p}\Delta\beta_T^h. \quad (10)$$

In this article, we outline a procedure for straightforward determination of the protein intrinsic compressibility, β_T^p, from the apparent compressibility determined experimentally, β_T^{exp}. The central idea of the procedure is the evaluation of $\Delta\beta_T^h$ for any protein of known three-dimensional structure.

Computational methods

Simulations

MD simulations of trypsin, ribonuclease A, lysozyme, and α-lactalbumin were performed for 1 ns in an isothermal-isobaric (NPT) ensemble (constant number of particles, N, constant pressure, $P = 1$ atm, and constant temperature, $T = 300$ K), with periodic boundary conditions and the particle-mesh Ewald method (47) for the calculations of the electrostatic interactions. Constant temperature and pressure were maintained using the Nosé-Hoover method of coupling to a heat bath and extended system algorithm for controlling the pressure, with a coupling constant of 25 ps. The computer simulations and post-processing of the generated trajectories are performed with CHARMM (48). More details of the MD simulations for the four globular proteins are described elsewhere (44). MD simulations were also performed for four small dipeptides: alanine dipeptide, glutamic acid dipeptide, aspartic acid dipeptide, and lysine dipeptide. Each $-CH_3$-blocked dipeptide was solvated in a cubic box of 1000 TIP3P water molecules and MD simulations were performed using the same protocol as for proteins (details of the simulations have appeared (49)).

Compressibility calculation

Isothermal compressibility, β_T, for proteins, protein and dipeptide solutions, and pure water (TIP3P model), is determined from volume fluctuations, δV^2, and average volumes, $\langle V\rangle$, according to

$$\beta_T = \frac{1}{k_B T}\frac{\delta V^2}{\langle V\rangle_{NPT}}. \quad (11)$$

The protein molecular volume, $V(t)$, is determined using a grid-based, atomic volume expansion algorithm, that allows for the protein interstitial spaces to be included in the total volume together with the protein van der Waals volume (44). Protein volumes are evaluated every ps from a 1-ns equilibrium trajectory with a 0.2 Å grid spacing. The average and standard deviations of the volume distribution are then used to calculate protein intrinsic compressibility, β_T^p. The isothermal compressibility evaluated in this manner correlates well with the experimental compressibility. As pointed out in a recent review (7), the suitability of any method for protein volume calculation is validated by the correlation with experiment. The compressibility of the protein and dipeptide solutions is also determined from Eq. 11, using the histogram of the time series of the whole system (simulation box) volume, $V^{sol}(t)$, rather than molecular volumes.

Errors due to finite grid spacing

We have calculated protein total volume using a grid spacing, $\delta = 0.2$ Å. An upper bound for the errors introduced in the volume calculation by the finite grid spacing can be obtained as follows. Let us assume that our protein is spherical, with a radius of 15 Å. An estimate of the error in volume calculation is given by the ratio of the spherical shell volume at the protein surface with a width of $\Delta R = \delta$, $4\pi R^2\delta$, and the volume of the whole sphere, $4\pi R^3/3$. For our particular case, this ratio is $3 \times 0.2/15 = 0.04$, i.e., 4%. Following the rationale for error calculation described earlier (44), the error in compressibility calculation due to the calculation of volume with a finite grid spacing formula is 4%. This is a maximal error that assumes that all the protein surface is convex, such as the case with a spherical protein. In fact, the protein surface has both convex and concave local geometry (peaks and valleys) and some of the errors (overestimates and underestimates) will cancel out. The total error in the calculation of the total protein volume with a grid spacing of 0.2 Å for a globular protein with a radius of gyration of 15 Å is smaller than 4%.

Atomic categories

To calculate the charged, polar, and apolar contributions to $\Delta\beta_T^h$, surface-exposed protein atoms are placed in three categories: charged (crg); polar (pol); or apolar (apl). Charged atoms include the side-chain carboxyl groups of Asp and Glu, the guanidinium group of Arg, the side-chain amino group of Lys, the amino and carboxylate termini of the polypeptide chain, and the charged Ca^{2+} ions in α-lactalbumin and trypsin. The polar category includes atoms with an absolute partial charge $|q_i| > 0.3\ e$, excluding those atoms defined as charged. All other atoms are apolar atoms. The solvent-accessible surface areas contributed by charged, polar, and apolar atoms (SAS^{type}, $type = crg, plr, aplr$), as well as the total SAS are calculated with CHARMM with a probe radius of 1.4 Å (based on the algorithm of Lee and Richards (50)).

Radial distribution functions

The radial distribution function (RDF), $g_{AB}(R)$, of atom A, at a distance R from a central atom, B, can be calculated from the normalized histogram of the numbers of atoms, $n_A(R + 0.5\Delta R)$, in a bin of width ΔR:

$$g_{AB}(R + 0.5\Delta R) = \frac{3V}{4\pi N}\frac{n_A(R + 0.5\Delta R)}{((R + \Delta R)^3 - R^3)}. \quad (12)$$

N is the total number of atoms in the system and V is the volume of the system.

We calculate the $g_{OA}^{type}(R)$ of water oxygen atoms surrounding charged, polar, and apolar groups using CHARMM. The space around each solvent-exposed protein (heavy) atom is divided into concentric spherical shells with a width of 0.1 Å for a total of 100 shells, with R ranging from 0 to 10 Å from the center of each heavy atom on the protein surface. The oxygen-oxygen $g_{OO}(R)$ values in ambient water are calculated from a 20-ns simulation of 2000 water molecules at constant pressure and temperature with periodic boundary conditions and the particle-mesh Ewald (47) summation for the electrostatic interactions. The protein-water RDFs do not converge to 1; instead, they converge to ~0.62. This is because the volume occupied by the protein is not accessible to the solvent. We renormalized the protein-water RDFs (i.e., multiplied the original curves by 1.61) to converge to 1.0 for easier comparison with $g_{OO}(R)$ of pure water-TIP3 model.

RESULTS

Radial distribution functions of water molecules around charged, polar, and apolar groups in proteins

The radial distribution functions of water oxygen atoms around charged, polar, and apolar protein surfaces were

calculated to assess the differences in water structure at different types of protein-water interfaces and to gain insights into the physical basis for hydration water compressibility. A comparison of protein-water RDFs with the bulk water RDFs provides the guidelines for calculating the volume of the protein hydration water, V^{h}. The distance from the protein surface at which no differences between hydration and bulk water are observed is then used to estimate the hydration water volume and the volume fraction Φ^{h}.

Interface-specific RDFs

The interface-specific, cumulative RDF, $g^{c}(R)$, is a sum over all same-type (charged, polar, or apolar) surface-exposed atomic RDFs:

$$g^{c}(R) = \sum_{N=1}^{N^{type}} g_{O^{wat}A^{type}}(R). \quad (13)$$

A^{type}, $type = crg, plr, aplr$ is a heavy atom on the protein surface (the central atom in Eq. 12) and O^{wat} is the Oxygen atom of water molecules. For this calculation, the exposed surface atoms are those atoms with the time-average $\langle SAS \rangle$ larger than 10.0 Å^2, to allow full contact with one water molecule.

Differences in the cumulative RDFs around charged (*red curves*), polar (*green curves*), and apolar (*black curves*) atoms are shown for trypsin (Fig. 1 *A*) and α-lactalbumin (Fig. 1 *B*). The sum of the converged values of the cumulative RDFs is proportional to the total number of heavy atoms that are solvent-exposed in each globular protein. Radial distribution functions, $g(R)$, for the three types of interface are obtained from the normalized cumulative RDFs by scaling with the trajectory-averaged numbers of atoms of each type that are surface-exposed, $\langle N_S^{type} \rangle$. Three distinct $g(R)$ values, each representing the hydration of a particular type of protein-water interface, are obtained in Fig. 1, *C* (trypsin) and *D* (α-lactalbumin).

The perturbation of water structure by atom types is shown by a comparison of the protein-water $g(R)$ with the oxygen-oxygen $g_{OO}(R)$ for bulk water, the TIP3P model (51) (*blue solid curves* in Fig. 1, *C* and *D*). The width and position of the first peak in the RDFs around charged and polar surfaces (the *red* and *green curves* in Fig. 1) are roughly the same as the width and position of the first peak in the $g_{OO}(R)$ in water, whereas their intensities are different: the peak is larger for charged protein groups, indicating that these groups are surrounded by more water molecules than oxygen atoms in bulk water. The peak is less intense for the RDF from the polar atomic groups in proteins indicating that they are less hydrated than charged atoms. In contrast, the position of the main peak in the RDFs of the apolar surface (*black curves* in Fig. 1) is shifted away from the surface, and the peak is wider and smaller than the first peak in the RDFs of the charged and polar surfaces.

In liquid water under ambient conditions, each water molecule is surrounded on average by 3.8 closest neighbors. At the protein-water interface, approximately half of the configurational space is available to water, which means that each heavy atom should be in the proximity of ~1.9 water molecules. We calculate the average number of hydration waters in the protein first hydration shell (to a radial distance of 3.2 Å from the protein atoms), and find that there are 2.7 water molecules hydrating charged groups, 1.5 water molecules for polar groups, and 1.1 for apolar groups.

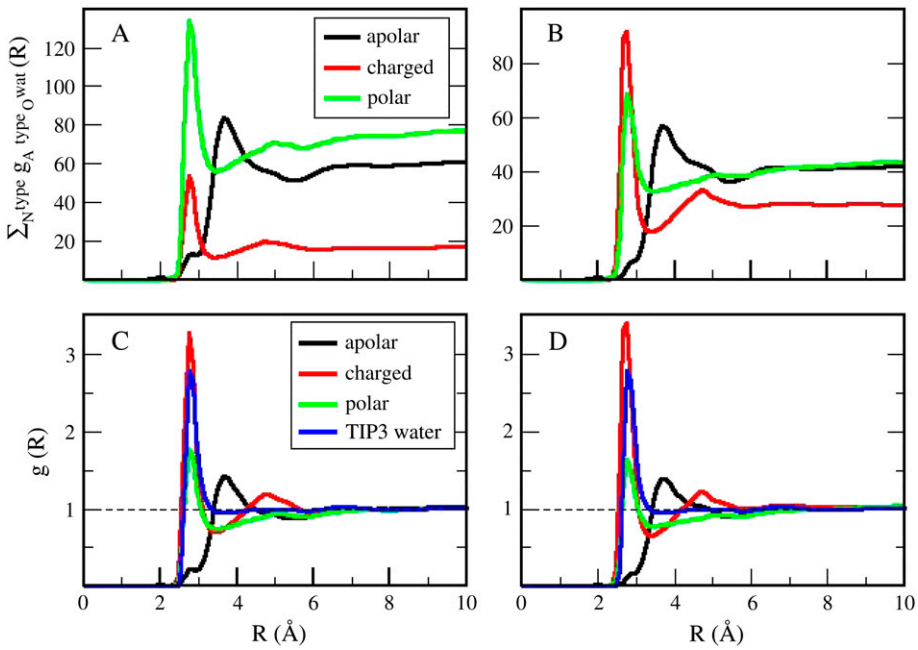

FIGURE 1 Cumulative RDFs for trypsin (*A*) and α-lactalbumin (*B*). Charged RDFs are the red curves, polar RDFs are the green curves, and apolar RDFs are the black curves. Normalized RDFs for trypsin (*C*) and *a*-lactalbumin (*D*): charged-groups hydration, red curves; polar hydration, green curves; and apolar hydration, black curves. The oxygen-oxygen RDF for water, TIP3P model, is shown in blue.

The results presented here provide a physical basis for the decomposition of the protein hydration volume into contributions from charged, polar, and apolar hydration volumes. No discernable differences in the protein-water RDFs and bulk water are detected beyond 8 Å from the protein surface. MD simulations of proteins typically include 2–3 layers of water molecules separating the protein from the wall of the simulation box. As such, the majority of the water in the simulation box can be approximately characterized as hydration water.

Local environment

At the protein-water interface, the individual atomic group-specific RDFs are influenced by two factors of the local environment: the unique configuration of neighboring amino acids, and the local topology of the solvent-accessible surface, which could be convex or concave. Effects of the local environment are manifested in the differences on the RDFs for the same chemical group from individual residues. The RDFs of same-type atomic group vary somewhat in magnitude but the structural distinctions between atomic types apparent in the RDFs in Fig. 1 are consistently manifested for individual residues (results given in Supplementary Material). Difference in positions of the first hydration shell are due partly to local surface topology. That is, the maximum position at smaller distances for charged groups reflects the convex surface where these residues are generally located, while the shift toward larger distances is consistent with a concave surface for apolar groups.

Model: hydration shell compressibility as a function of the chemical nature of the protein-water interface

To assess the change in compressibility due to protein hydration we formulate a simple model where the compressibility of the hydration water, β_T^h, has contributions dependent on the nature of the protein surface exposed to water: charged, polar, and apolar. To estimate these contributions, we first identify the hydration contribution to the solution compressibility as the residual compressibility, $\Delta\beta_T^{sol}$ (Eq. 6)—the difference between the actual isothermal compressibility of the protein-water solution (β_T^{sol} estimated from simulation)—and an ideal protein-water solution compressibility, which does not take into account changes in the solvent compressibility due to protein solvation, β_T^0. If water and the protein molecule are assumed to be independent components of the protein-water system, β_T^0 can be estimated from the intrinsic protein compressibility, β_T^p, and the compressibility of the bulk water, β_T^w, weighted by their corresponding volume fractions, Φ^p and $\Phi^h + \Phi^w$, according to Eq. 5.

The actual isothermal compressibility of the protein-water solutions and bulk water systems are calculated from the whole simulation box average volume, $\langle V_{sol}\rangle$, and fluctuations, δV_{sol}^2, according to Eq. 11. The calculated protein-water solution compressibility, β_T^{sol} (Table 1), is between 52 and 55 10^{-6} atm^{-1}. The compressibility of bulk water (TIP3P model) calculated from whole system volume fluctuations is 61 10^{-6} atm^{-1} (44,51). Intrinsic, isothermal protein compressibility, β_T^p, is calculated from protein volume fluctuations, δV^2, and average volumes, $\langle V\rangle$, as described previously (44). The values β_T^{sol} and β_T^0 are listed in Table 1.

The resulting values for $\Delta\beta_T^{sol}$ range from 2.7×10^{-6} atm^{-1} for trypsin to -0.83×10^{-6} atm^{-1} for HEW lysozyme (Table 1). Such a range in $\Delta\beta_T^{sol}$ implies a nonuniform contribution from the protein surface that could arise from differences in the hydration of surface groups with different polarity.

We use a simple, linear model to decompose $\Delta\beta_T^h$ (Eq. 6), into contributions from changes in the hydration water compressibility due to charged, polar, and apolar surfaces. The volume of the hydration water is decomposed into the volume near charged surfaces, $V^{h\text{-}crg}$, the volume near polar surfaces, $V^{h\text{-}plr}$, and the volume near apolar surfaces, $V^{h\text{-}aplr}$. The corresponding volume fractions are $\Phi^{h\text{-}crg} = V^{h\text{-}crg}/V^h$, $\Phi^{h\text{-}plr} = V^{h\text{-}plr}/V^h$, and $\Phi^{h\text{-}aplr} = V^{h\text{-}aplr}/V^h$. In a first approximation, the hydration volume fractions can be replaced by the respective surface fractions as $V^{h\text{-}crg}/V^h \approx SAS^{crg}/SAS = f_S^{crg}$, $V^{h\text{-}plr}/V^h \approx SAS^{plr}/SAS = f_S^{plr}$, and $V^{h\text{-}aplr}/V^h \approx SAS^{aplr}/SAS = f_S^{aplr}$. With this approximation, $\Delta\beta_T^h$ becomes

$$\Delta\beta_T^h = \Delta\beta_T^{h\text{-}crg} f_S^{crg} + \Delta\beta_T^{h\text{-}plr} f_S^{plr} + \Delta\beta_T^{h\text{-}aplr} f_S^{aplr}. \quad (14)$$

Best-fitted values for $\Delta\beta_T^{h\text{-}type}$ are calculated from f_S^{type} and $\Delta\beta_T^{sol}$ values (Table 1) by using a least-squares maximum likelihood method, $\Delta\beta_T^{h\text{-}crg} = -49.3$, $\Delta\beta_T^{h\text{-}plr} = -1.7$, and $\Delta\beta_T^{h\text{-}aplr} = 16.6$ (in units of 10^{-6} atm^{-1}; see Table 2). The compressibility of hydration water surrounding charged, polar, and apolar surfaces are the values $\beta_T^{h\text{-}crg} = 11.7\,10^{-6}$ atm^{-1}, $\beta_T^{h\text{-}plr} = 59.3\,10^{-6}$ atm^{-1}, and $\beta_T^{h\text{-}aplr} = 77.6\,10^{-6}$ atm^{-1}. The estimates for $\Delta\beta_T^{h\text{-}type}$ reproduce the original $\Delta\beta_T^{sol}$ with <10% error (last row in Table 1).

TABLE 1 Solution compressibilities, β_T^{sol}, calculated from simulation box volume fluctuations, $(\Delta V^2)_{sol}^{\frac{1}{2}}$, and average, $\langle V\rangle_{sol}$

	Trypsin	RNAse A	HEW Lyso	α-Lacta
β_T^{sol}	55	54	52	53
β_T^0*	52.3	53	52.8	53.6
$\Delta\beta_T^{sol}$†	2.7	1.1	−0.8	−0.6
f_S^{crg} (%)	9.0	14.0	16.0	18.0
f_S^{plr} (%)	40.0	35.0	38.0	29.0
f_S^{aplr} (%)	51.0	51.0	46.0	53.0
Error $\Delta\beta_T$ (%)‡	4	10	1.7	4.3

Volumes are reported as Å^3 and compressibilities as 10^{-6} atm^{-1}.
*Solution compressibility using the calculated protein compressibility, $\beta_T^p = \beta_T^{corr}$ (44), bulk water compressibility, $\beta_T^w = 61\times10^{-6}$ atm^{-1}, and volume fractions, Φ^p and $\Phi^w = 1 - \Phi^p$.
†$\Delta\beta_T^{sol} = \beta_T^{sol} - \beta_T^0$.
‡Errors in the calculation of residual hydration water compressibility from the estimated changes in hydration water compressibility: $\Delta\beta_T^{h\text{-}crg}$, $\Delta\beta_T^{h\text{-}plr}$, and $\Delta\beta_T^{h\text{-}aplr}$.

TABLE 2 Local compressibilities around charged, polar, and apolar groups

	Charged	Polar	Apolar
$\Delta\beta_T^{h\text{-type}}$(prot)*	−49.3	−1.7	16.6
$\Delta\beta_T^{h\text{-type}}$(prot-renorm)†	−36.4	−1.3	12.2
$\Delta\beta_T^{h\text{-type}}$(pep)‡	−48.5	−6.6	−3.4
$\Delta\beta_T^{h\text{-type}}$(pep-renorm)§	−35.8	−4.9	−2.5
n_h^{first}(prot)	2.76	1.51	1.11

Errors estimated by maximum likelihood analysis of three proteins are 3.3% for charged, 6.3% for polar, and 0.3% for the apolar hydration compressibility.

*Change in types of hydration water compressibility around proteins.

†Change in types of hydration water compressibility around proteins renormalized to experimental water compressibility, 45 10^{-6} atm^{-1}.

‡Change in types of hydration water compressibility around peptides from maximum likelihood method applied to dipeptide data from Table 4, with $\Phi^h = 0.5$.

§Change in hydration water compressibility around peptides renormalized to experimental water compressibility, 45 10^{-6} atm^{-1}.

The values obtained for the compressibility of water around charged, polar, and apolar surfaces from maximum likelihood estimates indicate a clear trend: the compressibility of water hydrating charged atoms is lower than the compressibility of bulk water, the compressibility of water hydrating polar atoms is about the same as the compressibility of bulk water, and the compressibility of water surrounding apolar atoms tends to be higher than the compressibility of bulk water. The differences in hydration water compressibility for the three types of surfaces are consistent with the differences in hydration water density indicated by the corresponding RDFs (Figs. 1 and 2).

Independent estimate of the compressibility of the first hydration shell

A qualitative check of the results obtained for the compressibility of hydration water from the decomposition of the hydration water compressibility as a function of the chemical nature of the protein-water interface can be readily obtained. In particular, the prediction that more charged protein surface induces a reduced compressibility of the water in the hydration shell of proteins can be verified by calculating the isothermal compressibility of the first hydration layer surrounding trypsin and HEW-lysozyme from MD simulations. The first hydration layer is the main contributor to the hydration shell compressibility. Use of these two proteins is instructive because they have similar fractions of exposed polar (40% and 38%) and apolar (51% and 46%) surface but different fractions of exposed charged surface (8% for trypsin and 16% for HEW-lysozyme). An alternative to Eq. 11 for compressibility used here to estimate the compressibility of hydration water is (3)

$$\beta_T = \frac{v_0}{k_B T}\frac{\delta n_h^2}{\langle n_h\rangle}. \quad (15)$$

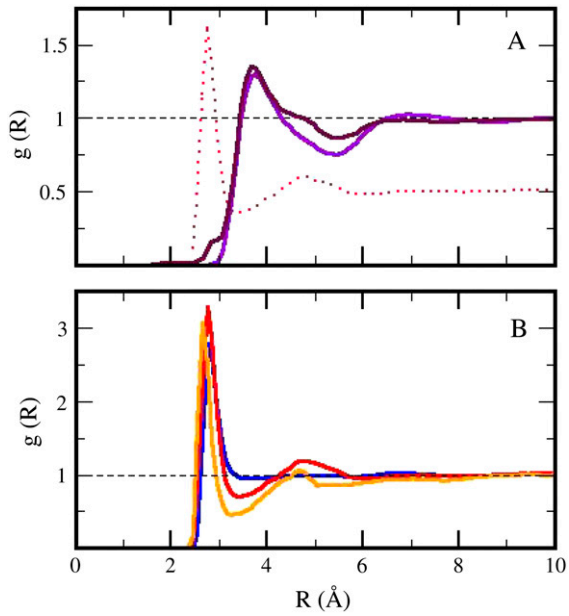

FIGURE 2 A comparison of RDFs for apolar surfaces (*A*) in proteins (*maroon curve*) and peptides (*violet curve*). The red (*dotted*) curve in Panel *A* represents the scaled (by half) RDF for charged protein surfaces. Panel *B* represents the RDFs for charged surfaces in proteins (*red curve*) and peptides (*orange curve*). The blue curve represents the oxygen-oxygen RDF for TIP3 water.

The value v_0 is the average volume of one water molecule in the hydration shell and δn_h^2 and $\langle n_h\rangle$ are the fluctuations and averages of the number of hydration water molecules in the first hydration shell, n_h. The value $n_h(t)$ is a time series of the number of water molecules within a shell of 3.1 Å surrounding the protein.

The average volume of a water molecule in the first hydration layer is estimated using the average solvent-accessible surface area, the width of the layer, and the average number of hydration waters: $v_0 = \langle SAS\rangle \times 3.1/\langle n_h\rangle$. The average of v_0 for the four globular proteins is 23 $Å^3$. This value is smaller than the average volume of a water molecule in bulk, 29.7 $Å^3$, but close to the volume of water molecules hydrating proteins obtained from high resolution crystal structures of 22 globular proteins (52), 23.8 $Å^3$.

We calculated β_T^h for trypsin and HEW lysozyme from $\langle n_h\rangle = 676$ for trypsin and $\langle n_h\rangle = 456$ for HEW lysozyme, the fluctuations in the number of water molecules in the first hydration shell, $\sigma_{n_h} = 7.58$ (trypsin) and 5.84 (HEW lysozyme), and the average volume of one water molecule in the hydration shell, v_0. With these values, the calculated compressibility of the first hydration layer is 47.2 10^{-6} atm^{-1} for trypsin and 41.6 10^{-6} atm^{-1} for HEW lysozyme. The calculated hydration compressibility values for the first solvation layer are 77% of bulk water compressibility for trypsin and 68% of bulk water compressibility for HEW-lysozyme. The lower compressibility of the first hydration layer of HEW lysozyme relative to trypsin is consistent with the model

(Eq. 14), given the difference in the charged solvent-accessible surface areas exposed by the two proteins to water. However, this is only an approximation for the compressibility of the first hydration shell and does not represent the whole hydration water compressibility. Equation 15 is strictly valid for compressibility calculation in an NVT ensemble.

Application to experiment: a formula for obtaining intrinsic protein compressibility from the experimentally determined compressibility

We can now outline a simple algorithm for the calculation of protein intrinsic isothermal compressibility, β_T^p from the experimental isothermal protein compressibility, β_T^{exp}. First, one must estimate f_S^{type} from the charged, polar, apolar, and total SAS of the three-dimensional structure of the protein. The value $\Delta\beta_T^h$ is then calculated from f_S^{type} and $\Delta\beta_T^{h\text{-}type}$ using Eq. 14. Next, protein intrinsic compressibility, β_T^p can be calculated from Eq. 9 by subtracting $\Phi^h/\Phi^p\Delta\beta_T^h$ from the experimental isothermal compressibility.

The algorithm can be used to decompose the experimental compressibility for any protein of known three-dimensional structure. For the application, it is necessary to reconcile the difference in the calculated and experimental isothermal compressibility of bulk water. Bulk water compressibility estimated from sound velocity measurements is 45×10^{-6} atm^{-1} in ambient conditions (53) while the calculated compressibility for pure water, the TIP3P model, is 61×10^{-6} atm^{-1}. The calculated $\Delta\beta_T^{h\text{-}type}$ must be rescaled by a factor of 45:61 to be compatible with the experiment (Table 2).

The use of Eq. 9 for the calculation of the intrinsic protein compressibility, β_T^p, requires an estimate of the ratio between the hydration and protein volume fractions, Φ^h/Φ^p. Here we use a simple approximation dictated by the convergence of the RDFs around protein molecules: no differences in the RDFs between the bulk water and hydration water can be observed at distances, D, larger than 8 Å from the protein surface (Figs. 1 and 2). For a globular protein with a radius of gyration, R_{gyr}, the ratio between the hydration and protein volumes can be approximated by

$$V^h/V^p \approx ((R_{gyr} + D)^3 - R_{gyr}^3)/R_{gyr}^3.$$

For a typical protein, with $R_{gyr} = 15$ Å and $D = 8$ Å, the Φ^h/Φ^p ratio is 2.6.

NMR absorption equilibrium experiments (54) have found that for HEW-lysozyme, the onset of drastic decrease in relaxation rate occurs at 1.7 g of water for 1 g of protein. If the water density is 1 g/cm^3 and protein density is 1.43 g/cm^3, $\Phi^h/\Phi^p = 1.7 \times 1.43 = 2.43$. The ratio of Φ^h/Φ^p obtained using the guidelines provided by the RDFs is in good agreement with estimates from NMR experiments. Both methods estimate that hydration of a regular protein such as lysozyme requires a hydration water volume that is approximately two-and-a-half times larger than the protein volume.

Examples

The method is illustrated using three proteins of known structure: lysozyme, α-lactalbumin, and myoglobin. The value β_T^{exp} have been estimated by Gekko and Hasegawa using Eq. 8 (38). The $\Delta\beta_T^h$ values are estimated here using the adjusted (to experiment) $\Delta\beta_T^{h\text{-}type}$ from Table 2 and the corresponding fractions of protein-water interfaces, f_S^{type}. To obtain the intrinsic protein compressibility we estimate the corrections due to protein hydration using the ratio of $\Phi^h/\Phi^p = 2.43$. This leads to intrinsic compressibility corrections ranging from 1.13 10^{-6} atm^{-1} for α-lactalbumin and 4.37 10^{-6} atm^{-1} for myoglobin. Final protein intrinsic compressibilities are listed in Table 3.

The first two proteins considered here, lysozyme and α-lactalbumin, are part of the set of proteins for which the intrinsic isothermal compressibility was calculated from volume fluctuations in MD simulations (44) (Table 3). The intrinsic protein compressibilities calculated from experimental data are in good agreement with the intrinsic compressibility calculated from molecular volume fluctuations, $\beta_T^p(\delta V^2)$ (Table 3) (44) with a relative error of 7% for lysozyme and 12% for α-lactalbumin. For myoglobin, our method predicts a large contribution to the intrinsic isothermal compressibility from hydration, 4.4 10^{-6} atm^{-1}. This estimate remains to be verified through other methods.

The match between the intrinsic protein compressibility calculated from molecular volume fluctuations and from estimates of hydration contributions to protein compressibility provides a first validation of the theoretical model for protein hydration compressibility.

Hydration water compressibility around small compounds

Others have discussed the possibility that the hydration characteristics of atomic groups at the protein surface can be modeled from hydration properties of similar groups in small molecules. It was proposed that protein hydration water

TABLE 3 Examples: the calculation of protein intrinsic isothermal compressibility from experimental compressibility

	α-Lactalbumin	Lysozyme	Myoglobin
β_T^{exp}	12.4	7.73	11.2
$\beta_T^p(\delta V^2)$*	11.8	10.2	—
f_S^{crg}(%)	18	16	20
f_S^{plr}(%)	29	38	30
f_S^{aplr}(%)	53	46	50
$\Delta\beta_T^h$	−0.46	−0.71	−1.8
$2.43\Delta\beta_T^h$	−1.12	−1.72	−4.37
$\beta_T^{p\,\dagger}$	13.5	9.5	17.5

*β_T^{int} is protein intrinsic compressibility calculated from volume fluctuations (44) for α-lactalbumin and lysozyme.

†β_T^p is the intrinsic compressibility calculated from experimental data including the hydration contribution.

compressibility can be estimated from the hydration water compressibility of small compounds (39,55,56), but the idea was challenged on the basis that the environment of basic groups in proteins is different from that of small molecules in solution or, alternatively, because the charged groups in proteins may not be as exposed to the solvent as those in small molecules ((7) and references therein). The calculation of hydration water compressibility from MD simulations offers the opportunity for a direct evaluation of this proposal.

We determine the solution compressibility of four dipeptides in water, β_T^{sol}, from whole system volume fluctuations and the predicted solution compressibility, β_T^0, based on volume fraction of water and peptide solute. Values for $\langle SAS\rangle^{crg}$, $\langle SAS\rangle^{plr}$, $\langle SAS\rangle^{aplr}$, and total $\langle SAS\rangle$, their respective fractions, as well as peptide and simulation box volumes for small compounds, are presented in Table 4. The intrinsic compressibility is very small for peptides because of the insignificant amount of interstitial space shielded from water, the major source of intrinsic compressibility (44), and the negligible compressibility of the covalent bonds (4). The β_T^{sol} and other values calculated from the simulations are shown in Table 4 for four dipeptides: alanine, glutamic, aspartic, and lysine dipeptides. For all four compounds, the dipeptide volume fraction, Φ^p, is small, ~2% of the total solution volume and Φ^w is 98%. The term $\Phi^p\beta_T^p \approx 0.0$ and the main contributions to the predicted solution compressibility are from water, $\Phi^w\beta_T^w = 0.98\times61\ 10^{-6}\ atm^{-1} = 59.0$. $\Delta\beta_T^{sol} = \Phi^h(\beta_T^h - \beta_T^w)$ for the four small compounds are listed in Table 4.

The $\Delta\beta_T^{h\text{-}type}$ was estimated for the four dipeptides using f_S^{type} and $\Delta\beta_T^{sol}$ shown in Table 4, and $\Phi^h = 0.5$. The optimized values (Table 2) for $\Delta\beta_T^{h\text{-}type}$ from the dipeptides for our small compounds are as follows: $\Delta\beta_T^{h\text{-}crg} = -48.5$, $\Delta\beta_T^{h\text{-}plr} = -6.6$, and $\Delta\beta_T^{h\text{-}aplr} = -3.4$ (in units of 10^{-6} atm^{-1}). After normalization with experimental isothermal compressibility of bulk water, $\Delta\beta_T^{h\text{-}crg} = -35.8$, $\Delta\beta_T^{h\text{-}plr} = -4.9$, and $\Delta\beta_T^{h\text{-}aplr} = -2.5$ (10^{-6} atm^{-1}), and the isothermal compressibilities of water near peptide surfaces are $\beta_T^{h\text{-}crg} = 9.2\ 10^{-6}\ atm^{-1}$, $\beta_T^{h\text{-}plr} = 40.1\ 10^{-6}\ atm^{-1}$, and $\beta_T^{h\text{-}aplr} = 42.5\ 10^{-6}\ atm^{-1}$. For small compounds, the compressibility of the hydration water surrounding all types of atomic groups is lower than the compressibility of bulk water. The ranking among types of hydration water compressibility is the same as that for proteins in that $\beta_T^{h\text{-}crg} < \beta_T^{h\text{-}plr} < \beta_T^{h\text{-}aplr}$. The estimated values for hydration water compressibility around the dipeptides studied here reproduce $\Delta\beta_T^h$ with <5% error (calculated errors are listed in Table 4).

TABLE 4 Small compounds solution compressibility, β_T^{sol}, calculated from simulation box volume fluctuations, $(\Delta V^2)^{\frac{1}{2}}_{sol}$, and average, $\langle V\rangle_{sol}$

	Alad	Aspd	Glud	Lysd
β_T^{sol}	57	51	53	54
$\Delta\beta_T^{sol}$*	−2	−8	−6	−5
$\langle SAS\rangle^{crg}$	0	97	98	59
$\langle SAS\rangle^{plr}$	95	88	88	90
$\langle SAS\rangle^{aplr}$	280	230	257	333
$\langle SAS\rangle$	375	416	445	482
f_s^{crg}(%)	0	23	22	12
f_s^{plr}(%)	25	21	20	19
f_s^{aplr}(%)	75	56	58	69
$\langle V_{mol}^{pep}\rangle$ (Å^3)	479	545	600	669
$\langle V_{sol}\rangle$ (Å^3)	32,819	32,650	32,665	32,777
Error $\Delta\beta_T^{sol}$ (%)†	1	1	5	2

Compressibilities are reported as 10^{-6} atm^{-1}.
*$\Delta\beta_T^{sol} = \beta_T^{sol} - \beta_T^0$.
†Errors in the calculation of residual hydration water compressibility from the estimated changes in hydration water compressibility: $\Delta\beta_T^{h\text{-}crg}$, $\Delta\beta_T^{h\text{-}plr}$, and $\Delta\beta_T^{h\text{-}aplr}$.

The $\Delta\beta_T^{h\text{-}type}$ estimates for the dipeptides agree qualitatively with various experimental estimates for small compounds (4,57). It was found that at 25°C the charged amino $-NH_3^+$ and carboxyl-COO^- contribute have large negative partial molar adiabatic compressibility, −34 cm^3 mol^{-1} atm^{-1}. By contrast, for uncharged species, the partial molar adiabatic compressibility could be either positive (2.74 cm^3 mol^{-1} atm^{-1} for glycolamide and 1 cm^3 mol^{-1} atm^{-1} for the peptide −CONH− group) or negative (−2 cm^3 mol^{-1} atm^{-1} for the −OH groups in pentoses and hexoses, and −12.5 cm^3 mol^{-1} atm^{-1} for ribose).

Hydration properties of proteins and peptides

A comparison between the changes in hydration water compressibility around proteins and dipeptides reveals that the charged groups, and to a lesser extent the polar groups, are hydrated similarly in proteins and dipeptides, both exhibiting a decrease in the hydration water compressibility with respect to bulk water. Apolar groups have opposite sign contributions in proteins and dipeptides, with an increase in hydration water compressibility around apolar groups in proteins.

The two sets of $\Delta\beta_T^{h\text{-}type}$ (Table 2) cannot be transferred between proteins and peptides. We have calculated errors in $\Delta\beta_T^h$ for proteins using the values $\Delta\beta_T^{h\text{-}type}$ from dipeptides and, inversely, we attempted to reproduce the $\Delta\beta_T^h$ for dipeptides using the $\Delta\beta_T^{h\text{-}type}$ for proteins. Errors larger than 20% were obtained in all trials, with particularly large errors in the $\Delta\beta_T^h$ of proteins when using $\beta_T^{h\text{-}type}$ from dipeptides. One possible explanation for the differences in the compressibility of hydration contributions of same surface types in proteins and small compounds is the intrinsic differences in the size of the molecules (which translates into a significantly large difference of the curvature of the solute-water interface) and the particular local arrangements of adjacent charged, polar, and apolar surfaces in proteins. Based on the results presented here, we conclude that the hydration properties of native globular proteins cannot be modeled using small molecule data. Therefore, even though the charged, polar, and apolar contributions to hydration compressibility for small compounds are ranked the same way as those in proteins ($\beta_T^{crg} < \beta_T^{plr} < \beta_T^{aplr}$), their magnitude is different for

proteins and small compounds and the results are not interchangeable.

RDFs of water molecules around proteins and small dipeptides

The perturbation of the water structure around small molecules is compared to that around proteins by consideration of the RDFs in Fig. 2. While several common features exist, significant differences are also observed. One characteristic aspect of the small molecule RDFs is that the structuring of water extends to 10–12 Å from each given atom in the dipeptide. For proteins, there is no noticeable water structuring extending beyond 8–10 Å from the protein surface. One possible explanation is that for proteins, long-range structuring of water is averaged out by the overlapping hydration shells originating in the adjacent charged, polar, and apolar surfaces.

The main differences between the RDFs from apolar surfaces (Fig. 2 *A*) in proteins (*maroon curve*) and dipeptides (*violet curve*) are the two shoulders at short (2.7 Å) and intermediate (4.8 Å) distance in the RDF of the apolar surface of proteins. The scaled (by 1:2) RDF from the charged protein surfaces, the red (*dotted*) curve in panel *A*, is also shown in Fig. 2, to indicate that the positions of the shoulders in the RDFs from the protein apolar surface coincide with the positions of the peaks in the RDFs from the charged surface. The RDFs of water surrounding apolar surfaces are modulated by contributions from water molecules in proximity of charged surfaces in proteins, a direct effect of adjacent and intertwining charged and apolar protein surface patches on the protein solvent-accessible surface. The net result is an overlap in the hydration shells of adjacent apolar and charged solvent-accessible surface areas.

Fig. 2 *B* represents the RDFs for charged surfaces in proteins (*red curve*) and peptides (*orange curve*). The blue curve represents the oxygen-oxygen RDF for TIP3 water. The location of the first hydration shells is displaced by .1 Å in proteins with respect to peptides, i.e., the first water oxygen is in average closer to the peptide surface than to the protein surface, a possible consequence of the different curvature at the solute-solvent interface (with a smaller curvature for the protein surface). The second peaks are located at roughly the same distance, 4.7 Å, but the shape and the width of the peaks are different.

Differences in the specific charged, polar, and apolar hydration in proteins and small peptides originate from both chemistry and the local topology and geometry of the solvent-accessible surface. The overall curvature of the solvent-accessible surface for a 150-residue protein is $\sim$1:15 Å^{-1} and much larger, 1:3.5 Å^{-1}, for dipeptides. The local geometry at the protein-water interface is either convex or concave, corresponding to peaks and valleys on the protein solvent-accessible surface. As a general trend, the charged groups will tend to be on the peaks and the apolar groups on the valleys. Because of their smaller size, this tendency is less strictly followed for small peptides. Global and local differences in the geometry of the solvent-accessible surfaces between proteins and peptides are partially responsible for the fact that data from small compounds cannot be used to estimate the compressibility of water surrounding the protein.

DISCUSSION

The role of hydration and intrinsic protein flexibility in modulating protein stability, conformational dynamics, and protein-ligand binding can be better assessed when contributions from hydration water are well understood and easy to estimate. In this study we found that the compressibility of protein hydration shell is a function of the complex chemical nature of the protein-water interface: the water surrounding charged atomic groups in proteins has lower compressibility than bulk water and on the same order of magnitude as protein intrinsic compressibility; the compressibility of water surrounding polar atomic groups is about the same as bulk water compressibility; and the compressibility of water surrounding apolar atomic groups is higher than that of bulk water, $\beta_T^{prot} \approx \beta_T^{crg} < \beta_T^{w} \approx \beta_T^{plr} < \beta_T^{aplr}$.

From investigations of hydration characteristics of atomic groups at the surface of similar groups in small molecules, we find that the compressibility of protein hydration water can only be assessed from protein-based additive contributions from charged, polar, and apolar surfaces. These contributions are specific to proteins and cannot be modeled based on similar data obtained for small molecules.

Using the detailed information from MD simulations, we have outlined a protocol for separating intrinsic protein and hydration water contributions to experimental compressibility obtained from sound velocity measurements. The formulation in Eq. 14 allows for the intrinsic protein compressibility to be calculated using the chemically specific $\Delta\beta_T^{h\text{-}type}$ and the fractions of charged, polar, and apolar SASA for any protein of known three-dimensional structure.

Our analysis of hydration contribution to protein intrinsic compressibility is similar in spirit to that of Breslauer and co-workers (13). These authors have determined hydration contributions from cross-correlating thermodynamic and structural data for globular proteins. Our formulation combines information from protein volume fluctuations and the radial distribution functions of water around protein and small molecules to arrive at a simple and easy to implement formula for the intrinsic protein compressibility, while the changes in the compressibility of hydration water are explicitly included.

Interpretation of experimental data on compressibility change due to protein conformational transitions and protein-ligand binding (6) can benefit from the formulation presented here. It is a matter of estimating the change in each type of accessible surface area. As an example, in a detailed experimental investigation of the change in the partial specific adiabatic compressibility of globular proteins, Chalikian and

Breslauer (5) show that the partial specific adiabatic compressibility of globular proteins decreases upon unfolding. A decrease in the compressibility is also observed in the native to partially unfolded transition, while the transition to a compact intermediate state results in an increase in compressibility. The behavior may be rationalized from our model of hydration compressibility. When the protein unfolds, a large fraction of the previously buried apolar surface is exposed to water. The newly created hydration water layer surrounding the freshly exposed apolar groups has a higher compressibility than bulk water. The effect is a net increase in the compressibility of the hydration water and a decrease in the negative component of hydration contribution to the total compressibility change. Because the intrinsic contribution to the compressibility of the protein unfolded state is practically negligible, the sole contribution to solution compressibility is the hydration water compressibility. The $\Delta\beta_T^{\text{h-type}}$ and the unfolded state f_S^{type} allow for an estimate of the compressibility of the protein in the unfolded state for any protein of known sequence.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

This work was partially supported by grants from the U.S. National Institutes of Health (AI39639 to C.B.P.) and the Sloan Foundation and Department of Energy (to V.M.D.).

REFERENCES

1. Raschke, T. M. 2006. Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.* 16:152–159.
2. Eden, D., J. B. Matthew, J. J. Rosa, and F. M. Richards. 1982. Increase in apparent compressibility of cytochrome c upon oxidation. *Proc. Natl. Acad. Sci. USA.* 79:815–819.
3. Phelps, D. K., and C. B. Post. 1995. A novel basis for capsid stabilization by antiviral compounds. *J. Mol. Biol.* 254:544–551.
4. Chalikian, T. V., A. P. Sarvazyan, and K. J. Breslauer. 1994. Hydration and partial compressibility of biological compounds. *Biophys. Chem.* 51:89–109.
5. Chalikian, T. V., and K. J. Breslauer. 1996. Compressibility as a means to detect and characterize globular protein states. *Proc. Natl. Acad. Sci. USA.* 93:1012–1014.
6. Gekko, K. 2002. Compressibility gives new insight into protein dynamics and enzyme function. *Biochim. Biophys. Acta.* 1595:382–386.
7. Chalikian, T. V. 2003. Volumetric properties of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 32:207–235.
8. Gekko, K., A. Kimoto, and T. Kamiyama. 2003. Effects of disulfide bonds on compactness of protein molecules revealed by volume, compressibility, and expansibility changes during reduction. *Biochemistry.* 42:13746–13753.
9. Taulier, N., I. V. Beletskaya, and T. V. Chalikian. 2005. Compressibility changes accompanying conformational transitions of apomyoglobin. *Biopolymers.* 79:218–229.
10. Dubins, D. N., R. Filfil, R. B. Macgregor, and T. V. Chalikian. 2003. Volume and compressibility changes accompanying thermally induced native-to-unfolded and molten globule-to-unfolded transitions of cytochrome c: A high pressure study. *Biochemistry.* 42:8671–8678.
11. Chalikian, T. V., and R. Filfil. 2003. The thermodynamics of protein-protein recognition as characterized by a combination of volumetric and calorimetric techniques: the binding of turkey ovomucoid third domain to alpha-chymotrypsin. *J. Mol. Biol.* 326:1271–1288.
12. Jung, C. 2002. Cytochrome p-450-co and substrates: lessons from ligand binding under high pressure. *Biochim. Biophys. Acta.* 1595: 309–328.
13. Chalikian, T. V., M. Totrov, R. Abagyan, and K. J. Breslauer. 1996. The hydration of globular proteins as derived from volume and compressibility measurements: cross correlating thermodynamic and structural data. *J. Mol. Biol.* 260:588–603.
14. Dadarlat, V. M., and C. B. Post. 2003. Adhesive-cohesive model for protein compressibility: an alternative perspective on stability. *Proc. Natl. Acad. Sci. USA.* 100:14778–14783.
15. Steinbach, P., and B. Brooks. 1993. Protein hydration elucidated by molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 90: 9135–9139.
16. Schroder, C., T. Rudas, S. Boresch, and O. Steinhauser. 2006. Simulation studies of the protein-water interface. I. Properties at the molecular resolution. *J. Chem. Phys.* 124:234907.
17. Rudas, T., C. Schroder, S. Boresch, and O. Steinhauser. 2006. Simulation studies of the protein-water interface. II. Properties at the mesoscopic resolution. *J. Chem. Phys.* 124:234908.
18. Vitkup, D., D. Ringe, G. A. Petsko, and M. Karplus. 2000. Solvent mobility and the protein ''glass'' transition. *Nat. Struct. Biol.* 7:34–38.
19. Rupley, J. A., and G. Careri. 1991. Protein hydration and function. *Adv. Protein Chem.* 41:37–172.
20. Pollack, G. H. 2001. Is the cell a gel—and why does it matter? *Jpn. J. Physiol.* 51:649–660.
21. Pollack, G. H. 2003. The role of aqueous interfaces in the cell. *Adv. Colloid Interface Sci.* 103:173–196.
22. Brunne, R. M., E. Liepinsh, G. Otting, K. Wuthrich, and W. F. van Gunsteren. 1993. Hydration of proteins: a comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. *J. Mol. Biol.* 231: 1040–1048.
23. Badger, J. 1993. Multiple hydration layers in cubic insulin crystals. *Biophys. J.* 65:1656–1659.
24. Svergun, D. I., S. M. Richard, H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. USA.* 95: 2267–2272.
25. Venu, K., L. A. Svensson, and B. Halle. 1999. Orientational order and dynamics of hydration water in a single crystal of bovine pancreatic trypsin inhibitor. *Biophys. J.* 77:1074–1085.
26. Russo, D., G. Hura, and T. Head-Gordon. 2005. Hydration dynamics near a model protein surface. *Biophys. J.* 86:1852–1862.
27. Levitt, M., and R. Sharon. 1988. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA.* 85:7557–7561.
28. Makarov, V. A., K. A. Andrews, and B. M. Pettitt. 1998a. Protein hydration density: theory, simulation and crystallography. *Curr. Opin. Struct. Biol.* 8:218–221.
29. Makarov, V. A., K. A. Andrews, and B. M. Pettitt. 1998b. Reconstructing the protein-water interface. *Biopolymers.* 45:469–478.
30. Makarov, V. A., B. K. Andrews, P. E. Smith, and B. M. Pettitt. 2000. Residence times of water molecules in the hydration sites of myoglobin. *Biophys. J.* 79:2966–2974.
31. Henchman, R. H., and J. A. McCammon. 2002. Structural and dynamic properties of water around acetylcholinesterase. *Protein Sci.* 11:2080–2090.
32. Halle, B., and M. Davidovic. 2003. Biomolecular hydration: from water dynamics to hydrodynamics. *Proc. Natl. Acad. Sci. USA.* 100:12135–12140.
33. Pratt, L. R., and D. Chandler. 1977. Theory of the hydrophobic effect. *J. Chem. Phys.* 67:3683–3704.

34. Lazaridis, T., and M. E. Paulaitis. 1992. Entropy of hydrophobic hydration: a new statistical mechanical formulation. *J. Chem. Phys.* 96:3847.

35. Lazaridis, T., and M. E. Paulaitis. 1994. Simulation studies of the hydration entropy of simple, hydrophobic solutes. *J. Phys. Chem.* 98:635.

36. Hummer, G., S. Garde, A. E. García, and L. R. Pratt. 2000. New perspectives on hydrophobic effects. *Chem. Phys.* 258:349–370.

37. Lee, B. 1983. Calculation of volume fluctuation for globular protein models. *Proc. Natl. Acad. Sci. USA.* 80:622–626.

38. Gekko, K., and Y. Hasegawa. 1986. Compressibility structure relationship of globular-proteins. *Biochemistry.* 25:6563–6571.

39. Kharakoz, D. P., and A. P. Sarvazyan. 1993. Hydrational and intrinsic compressibilities of globular proteins. *Biopolymers.* 33:11–26.

40. Chalikian, T. V., and K. J. Breslauer. 1998. Thermodynamic analysis of biomolecules: a volumetric approach. *Curr. Opin. Struct. Biol.* 8:657–664.

41. Chalikian, T. V. 2001. Structural thermodynamics of hydration. *J. Phys. Chem. B.* 105:12566–12578.

42. Marchi, M. 2003. Compressibility of cavities and biological water from Voronoi volumes in hydrated proteins. *J. Phys. Chem. B.* 107:6598–6602.

43. Burling, F. T., W. I. Weis, K. M. Flaherty, and A. T. Brünger. 1996. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science.* 271:72–77.

44. Dadarlat, V. M., and C. B. Post. 2001. Insights into protein compressibility from molecular dynamics simulations. *J. Phys. Chem. B.* 105:715–724.

45. Gavish, B., E. Gratton, and C. J. Hardy. 1983. Adiabatic compressibility of globular proteins. *Proc. Natl. Acad. Sci. USA.* 80:750–754.

46. Paci, E., and M. Marchi. 1996. Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulations at high pressure. *Proc. Natl. Acad. Sci. USA.* 93:11609–11614.

47. Darden, T. A., D. M. York, and L. G. Pedersen. 1993. Particle mesh Ewald: An $n\log(n)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

48. Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4: 187–217.

49. Dadarlat, V. M. 2005. Potentials of mean force (PMFs) for the interaction of blocked alanine dipeptide molecules in water and vacuum from MD simulations. *Biophys. J.* 89:1433–1445.

50. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

51. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

52. Gerstein, M., and C. Chothia. 1996. Packing at the protein-water interface. *Proc. Natl. Acad. Sci. USA.* 93:10167–10172.

53. Weast, R. C. 1989. Handbook of Chemistry and Physics, 70th Ed. CRC Press, Boca Raton, FL.

54. Lioutas, T. S., I. C. Baianu, and M. P. Steinberg. 1986. Oxygen-17 and deuterium nuclear magnetic resonance studies of lysozyme hydration. *Arch. Biochem. Biophys.* 247:68–75.

55. Iqbal, M., and R. E. Verrall. 1987. Volumetric properties of aqueous solutions of bovine serum albumin, human serum albumin, and human hemoglobin. *J. Phys. Chem.* 91:1935–1941.

56. Lockwood, D. M., P. J. Rossky, and R. M. Levy. 2000. Functional group contributions to partial molar compressibilities of alcohols in water. *J. Phys. Chem. B.* 104:4210–4217.

57. Taulier, N., and T. V. Chalikian. 2002. Compressibility of protein transitions. *Biochim. Biophys. Acta.* 1595:48–70.

Decomposition of Protein Experimental Compressibility into Intrinsic and Hydration Shell Contributions - Supplementary Material

Voichita M. Dadarlat and Carol Beth Post[1]
Department of Medicinal Chemistry and Molecular Pharmacology,
Markey Center for Structural Biology, Purdue Cancer Center
Purdue University, West Lafayette, IN 47907
*Corresponding author
email:cbp@purdue.edu

July 11, 2006

[1]This work was partially supported by grants from the U.S. National Institutes of Health (AI39639 to C. B. P.) and the Sloan Foundation and DOE (to V. M. D.)

1 Average hydration of individual amino acids

Residue specific hydration at the protein–water charged, polar and apolar interfaces is illustrated by a comparison of the RDFs from the amino acid averaged and pure water RDFs. Any influences from the specific local environment and the local topology of the solvent accessible surface are eliminated by averaging the RDFs of 10 to 15 amino acids from the proteinwater interfaces in trypsin, ribonuclease A, HEW lysozyme and α-lactalbumin. The RDFs for selected charged (carboxylate and amino), polar (hydroxyl) and apolar (methyl) groups to water oxygen, are shown in Figure 2 together with $g_{OO}(R)$ for bulk water. The first solvation shells of the charged carboxylate groups (red curve) located at $R = 2.66$ Å and charged amino groups, located at 2.8 Å, are both more highly populated than the corresponding shell for pure water (blue curve). In contrast, the methyl group of the ALA residues (black solid line) is poorly hydrated, with a first peak at 3.7 Å and possibly lower water density in its vicinity. The first hydration shell of the hydroxyl group of SER residues (green curve) is less occupied than the water-water first hydration shell, while it exhibits a weakly formed second hydration shell which is not present in bulk water. The second hydration shells appear at 4.7 Å for carboxylate groups, at 5 Å for the amino groups and at 4.7 Å for hydroxyl groups in SER. A third hydration shell is weakly formed around the charged groups of GLU (at 6.8 Å) and LYS (at 7.3 Å) amino acids. In general, residue specific RDFs show more water structure than the cumulative, protein-based RDFs, (Figures 1 and 2) indicating that some of the features in the residue specific RDFs are averaged out at large distances from the protein surface.

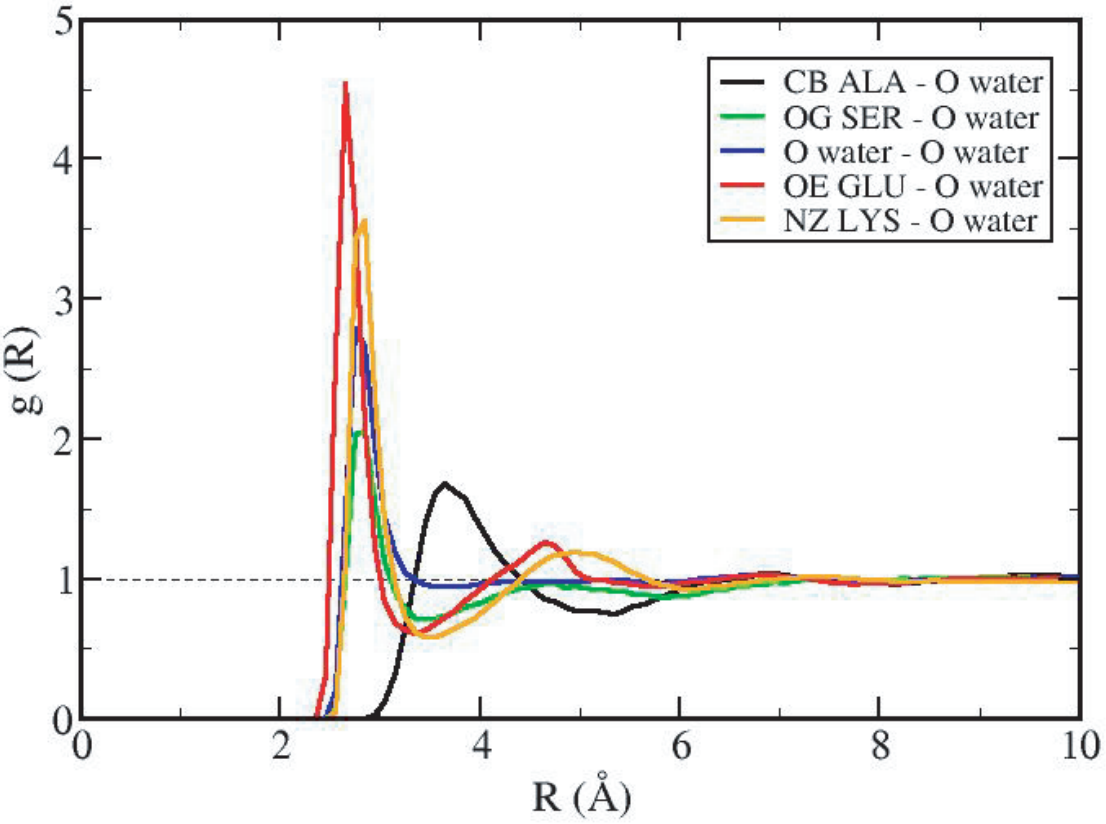

Figure 1: Averaged charged, polar and apolar, amino acid based RDFs.

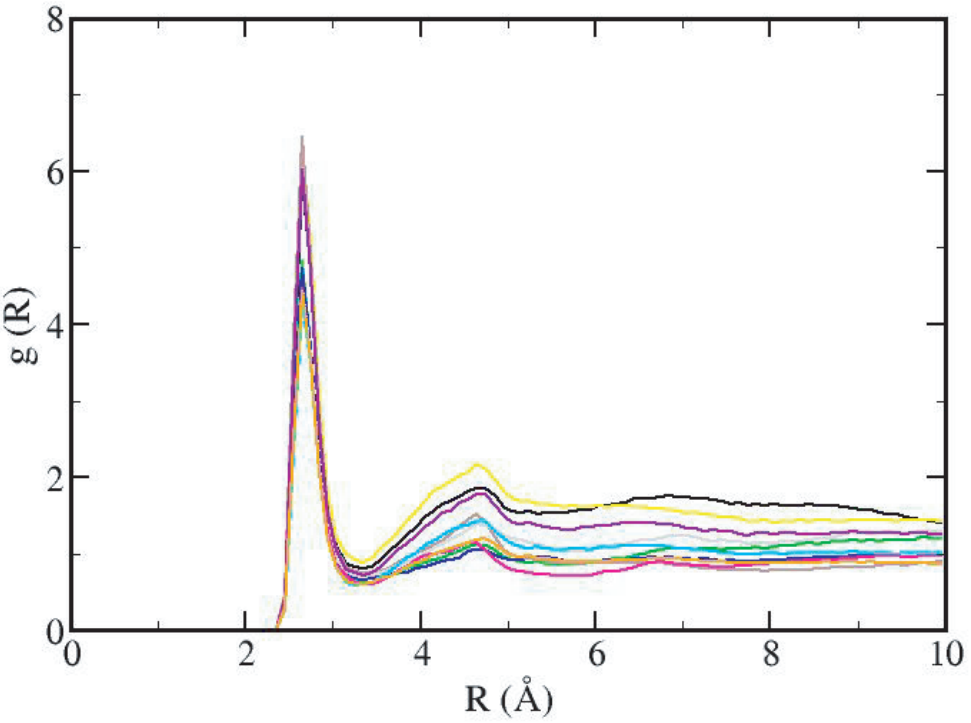

Figure 2: Influence of the local environment on the RDFs of charged Oxygen atoms from 12 solvent exposed GLU amino acids.