Protein hydration and unfolding – insights from experimental partial specific volumes and unfolded protein models

Lynne Reed Murphy, Nobuyuki Matubayasi*, Vilia A Payne and Ronald M Levy

Background: The partial specific volume of a protein is an experimental quantity containing information about solute–solvent interactions and protein hydration. We use a hydration-shell model to partition the partial specific volume into an intrinsic volume occupied by the protein and a change in the volume occupied by the solvent resulting from the solvent interactions with the protein. We seek to extract microscopic information about protein hydration and unfolding from experimental volume measurements without using computer simulations. We employ the idea that the protein–solvent interaction will be proportional to the surface area of the protein.

Results: A linear relationship is obtained when the difference between the experimental protein partial specific volume and its intrinsic volume is plotted as a function of the protein solvent-accessible surface area. The effect of using different protein volume definitions on the analysis of protein volumetric properties is discussed. Volumetric data are used to test a model for the unfolded state of proteins and to make predictions about the denatured state.

Conclusions: The linear relationship between hydration-shell volume change and accessible surface area reflects the similar surface properties (fractional composition of nonpolar, polar and charged surface) among a diverse set of proteins. This linear relationship is found to be independent of how the solution is partitioned into solute and solvent components. The interpretation of hydration shell versus bulk water properties is found to be very model dependent, however. The maximally exposed unfolded protein model is found to be inconsistent with experimental volume changes of unfolding.

Address: Department of Chemistry, Rutgers, The State University of New Jersey, Piscataway, NJ 08855, USA.

*Present address: Institute for Chemical Research, Kyoto University, Uji, Kyoto 611, Japan.

Correspondence: Ronald M Levy
E-mail: ronlevy@lutece.rutgers.edu

Key words: globular proteins, hydration, surface area, unfolding, volume

Received: **31 December 1997**
Accepted: **05 January 1998**

Published: **23 February 1998**
http://biomednet.com/elecref/1359027800300105

Folding & Design 23 February 1998, **3**:105–118

Introduction

Protein–solvent interactions are crucial to protein stability and biological activity and as such are important in protein folding. The contribution of protein–solvent interactions to thermodynamic parameters of protein folding has been studied extensively. One approach is based on connecting thermodynamic measurements with the change in protein solvent-accessible surface area on going from the native to the unfolded state. Studies using relationships between the thermodynamics of unfolding and the change in solvent-accessible surface area have included examinations of heat capacity, enthalpy, entropy and free energy [1–22]. These studies use a common scheme: the thermodynamic measurement is considered to be composed of additive contributions from constituent atoms or groups (see [23] for a discussion of nonadditive effects). Using free energy as an example, the free energy change of unfolding is calculated as:

$$\Delta G = \sum_i \sigma_i \Delta SA_i \qquad (1)$$

Each σ_i is a contribution to the free energy per unit solvent-accessible surface area for the various types of atoms or groups; the contributions are obtained from thermodynamic measurements on model compounds. The ΔSA_i values are the changes in solvent-accessible surface area for each type of atom or group on going from folded to unfolded protein; ΔSA_i values are calculated from structural data for the native protein and assume a model unfolded structure. Summing the product of the constituent values per unit solvent-accessible surface area and the change in solvent-accessible surface area for each type of atom or group yields the protein thermodynamic parameter.

Studies have also sought to gain information about protein folding by analyzing the solvent-accessible surface area of native proteins compared with a model of the proteins in the unfolded state [24–28]. The unfolded state is often modeled as an extended chain, with the solvent-accessible surface area of the unfolded protein calculated as the sum of residue (X) contributions in tripeptides Gly–X–Gly or Ala–X–Ala. Many interesting conclusions concerning the change in surface area on folding and the amount of nonpolar, polar and charged surface buried upon folding have been obtained with the assumption of a completely unfolded state.

Implicit in schemes that make a connection between thermodynamic properties and solvent-accessible surface area is the concept of the hydration shell — a shell of waters surrounding the protein, which behaves differently from bulk solvent. In such schemes it is assumed that the perturbation of the solvent structure caused by the solute is localized to the region directly surrounding the solute. Introduction of the hydration-shell concept has proven useful in the study of solvation. Recently, the statistical mechanical basis for the hydration-shell model has been analyzed [29,30]. These studies demonstrate that the spatial variation of the solvent contribution to excess thermodynamic quantities is different for different excess solution properties; for example, the solvent contribution to the excess energy is localized to the hydration shell, but the solvent contribution to the excess compressibility is not. The hydration-shell model for the excess volume was found to be qualitatively valid.

Volumetric properties are interesting because they provide insight into solute–solvent interactions and they are sensitive to the features of solute hydration. A recent study by Chothia and coworkers [31] explored the changes in volume upon protein unfolding by calculating amino acid and protein volumes. They calculated the mean volumes of residues buried in protein interiors using the method of Voronoi polyhedra and a set of protein crystal structures. Native protein volumes were then calculated as the sum of the constituent mean residue volumes plus corrections for electrostriction by surface groups. They obtained close agreement between calculated protein volumes and experimental partial specific volumes. The volume of residues in unfolded proteins was obtained from the volume residues occupy in solution, as calculated in solution studies of amino acids. By comparing the volume of residues in the protein interior with the volume of amino acids in solution, they observed that aliphatic groups have smaller volumes in protein interiors than in solution whereas the opposite is true for amide and charged groups; only small changes were found for aromatic, sulfur or hydroxyl groups. Thus, Chothia and coworkers explain the small observed change of volume for proteins upon unfolding as a net cancellation of the positive changes produced when polar groups are buried and the negative changes produced by aliphatic groups. Because this implies that the ratio of different types of buried groups is constant among the proteins, they calculated the chemical compositions of the buried and accessible surfaces and found them to be the same for each protein in the set examined in their study.

Partial specific volume is defined as the change of the solution volume when solute is dissolved in solvent at constant temperature and pressure. It is often approximated as the sum of two contributions; the first is the 'intrinsic' volume of the solute and the second is the change in solvent volume resulting from the perturbation by the solute. In this work we investigate empirical relationships between the experimental partial specific volumes of proteins and their 'intrinsic' volumes calculated from X-ray coordinates. The difference between the experimental partial specific volumes of proteins and their calculated volumes is related to solvent-accessible surface area in the spirit of previous studies that link experimental thermodynamic parameters with protein solvent-accessible surface area. An interesting linear correlation between experimental partial specific volume data and protein accessible surface area is found, which we suggest reflects similar surface properties (fractional composition of nonpolar, polar and ionic groups) among a diverse set of proteins. Alternative procedures for dividing the solution volume into contributions from protein and solvent are discussed and the effects on the interpretation of experimental partial molar volume data are analyzed. In a section concerning the volumetric properties of small nonpolar, polar and ionic solutes we have used the results of computer simulations to determine the partial specific volume of water contained in the hydration shells of these solutes. As discussed below, in order to obtain consistent results between estimates of the specific volume of water in the hydration shell of the proteins and the results for the small solutes, the volume occupied by water in the hydration shells of these solutes must be scaled by a factor proportional to the number of solvent molecules in the hydration shell rather than the solvent-accessible surface area.

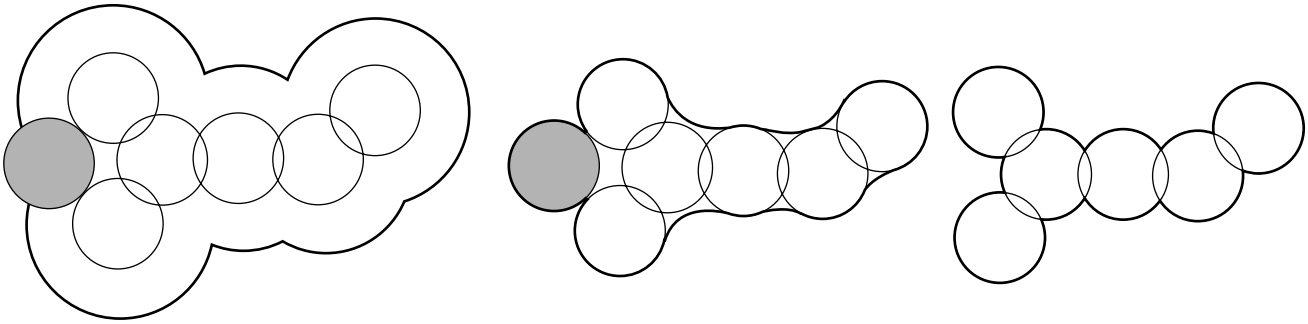
The analysis of protein partial specific volumes can also provide insights into protein unfolding. The measured volume change when a protein unfolds is small relative to the protein partial specific volume, but the exact nature of the unfolded state has not been determined experimentally. In a section concerned with volume changes associated with protein unfolding, the problem is formulated analogously to partial specific volume. There are two terms, corresponding to the change in protein volume and the change in the hydration-shell volume upon unfolding. Assuming a maximally exposed model of the unfolded state and comparing the calculated volume change to the experimental volume change, the performance of the maximally unfolded protein model is evaluated. Finally, we make some predictions about the volume and surface area characteristics of the denatured state of proteins.

Results

Protein partial specific volume

The intention here is to use experimentally measurable partial specific volume data of proteins to infer microscopic information about protein hydration. Partial specific volume is defined as the change in volume of a solution when a measured amount of solute is added. If the solute and solvent interact ideally, the resulting solution volume would simply be the sum of the original solution volume and the

Figure 1



An illustration of surface area and volume definitions. The shaded circles represent the probe sphere.

volume of the added solute. For a protein dissolved in water, solute–solvent interactions are not ideal and the resulting volume change, the partial specific volume, can be used to gain insight into their interactions. This is accomplished here by using a model to partition the solution volume into solute and solvent components. Thus, the partial specific volume is expressed as the sum of two contributions (the intrinsic volume of the protein and the change in the solvent volume within a hydration shell as a result of its interaction with the protein) according to the following equation (e.g. [32]):

$$\bar{v}_2 = v_2 + \delta_1(v_1 - v_1^o) \quad (2)$$

where $\bar{v}_2$ is the partial specific volume of the protein, v_2 is the 'intrinsic' volume assigned to the protein, δ_1 is the hydration number of the protein, v_1 is the partial specific volume of the hydration-shell water, and v_1^o is the specific volume of bulk water (1 cm^3/g = 30 $Å^3$/molecule). When the volume units are converted from cm^3/g to $Å^3$/molecule, the quantity δ_1 becomes N_h, the number of hydration-shell water molecules per protein molecule. The difference between the protein partial specific volume and the calculated protein volume, $\bar{v}_2 - v_2$, which we define to be ΔV_{sol}, is (by the above equation) equal to $N_h(v_1 - v_1^o)$, which is denoted ΔV_{hs}. Thus, by rearranging Equation 2 we get:

$$\Delta V_{sol} = \Delta V_{hs} \quad (3)$$

where ΔV_{hs} is simply the change in the solvent volume resulting from water in the hydration shell, which has a specific volume that differs from the specific volume of water in the bulk. ΔV_{sol} can be calculated if the partial specific volume of the protein and the intrinsic volume of the protein are both known; thus, ΔV_{hs} can also be determined. Relating ΔV_{sol} to the protein surface area makes the connection between the change in solution volume and protein surface properties.

The model used to partition the solution into solute and solvent contributions depends on the definition used to calculate the volume occupied by the protein. Various volume definitions are illustrated in Figure 1 and discussed in the Materials and methods section. As the choice of dividing surface between protein and solvent is somewhat arbitrary, Equation 2 does not determine a unique decomposition of the solution volume. In this section, results obtained using the protein excluded volume to define v_2 are reported. In a separate section, additional results obtained using alternative protein volume definitions are reported and the effect of the different models on the conclusions concerning protein hydration is discussed. It is also important to note that here the term 'excluded volume' is used as originally defined by Richmond [33] to mean the volume enclosed by the solvent-accessible surface or the volume excluded to the center of a probe sphere. Excluded volume has been used elsewhere to mean the volume excluded to any portion of a probe sphere, which is referred to as the 'molecular volume' here.

The proteins ranging in size from 51 to 307 residues used for this study are listed in Table 1 with the PDB code for their crystal structure coordinates. For each protein, the experimental partial specific volume was obtained and the molecular weight, solvent-accessible surface area and excluded volume were calculated. Table 1 also lists for each protein the difference between the experimental partial molar volume and the intrinsic volume (i.e. the calculated ΔV_{sol}).

A plot of ΔV_{sol} versus accessible surface area is shown in Figure 2. The linear relationship obtained between ΔV_{sol} and the surface area indicates that the partial specific volume of water in the hydration shell averaged over the protein surface is the same for the set of globular proteins

Table 1

Partial specific volume, excluded volume and surface area data with calculated results for 15 globular proteins.

Protein	PDB code	Number of residues	Molecular weight (Å^3/molecule)	Protein psv* (cm^3/g)	Protein psv* (Å^3/molecule)	ASA† (Å^2)	Protein excluded volume‡ (Å^3/molecule)	ΔV_{sol}§ (Å^3/molecule)
Insulin (monomer)	9ins	51	5780	0.735	7057	3421	11142	–4085
Pti	4pti	58	6520	0.718	7776	4087	12650	–4874
α-Lactalbumin	1alc	123	14010	0.735	17105	7162	25760	–8655
Ribonuclease A	3rn3	124	13690	0.703	15987	6884	24862	–8875
Lysozyme	1lzt	129	14310	0.703	16711	6590	25464	–8753
Myoglobin	5mbn	153	17200	0.745	21286	8178	31960	–10674
Adenylate kinase	3adk	195	21680	0.74	26650	10885	40504	–13854
Papain	1ppn	212	23430	0.719	27984	9452	40643	–12659
Bence–Jones REI	1rei	214	23500	0.726	28341	9751	41554	–13213
Concanavalin A (monomer)	2cna	237	25560	0.732	31080	10775	45697	–14617
Elastase	3est	240	25900	0.73	31407	10640	45740	–14333
Carbonic anhydrase B	2cab	261	28800	0.729	34876	10977	48766	–13890
Subtilisin	2sbt	275	27530	0.731	33429	10427	47785	–14356
Rhodanese	1rhd	293	32910	0.742	40563	14173	59171	–18608
Carboxypeptidase A	2ctb	307	34500	0.733	42007	11894	58055	–16048

*Protein psv, protein partial specific volume; †ASA, solvent-accessible surface area. ‡The protein excluded volume and the solvent-accessible surface area are calculated with a 1.4 Å probe radius. §$\Delta V_{sol} = \bar{v}_2 - v_2$.

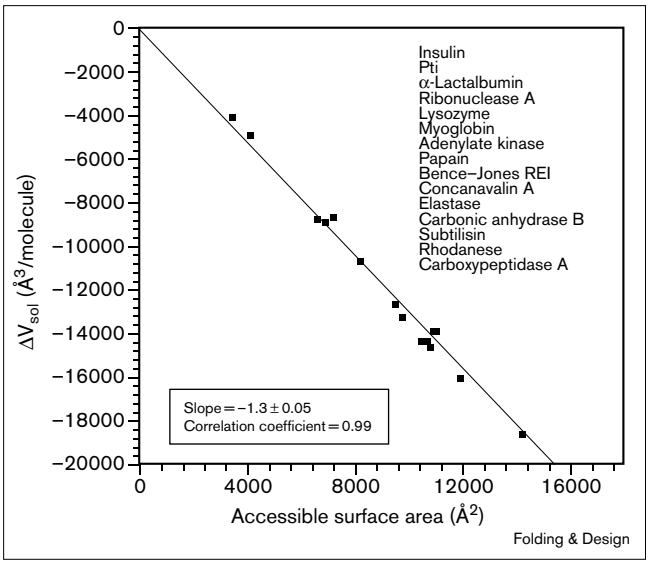
that constitute the database for the present study. The slope of the plot is negative (–1.3 ± 0.05), implying that the partial specific volume of the hydration-shell waters is less than the specific volume of bulk water. As discussed below, however, the slope depends on the choice of dividing surface between protein and solvent.

Because the volumetric properties of hydrating water depend strongly on the polarity of the solute, the linear relationship between ΔV_{sol} and protein surface area observed in Figure 2 suggests that the composition of the surfaces is relatively uniform among these proteins. To examine this, the accessible surface area was decomposed into proportions of nonpolar, polar and charged surface area; the results are listed in Table 2. The distribution of different types of surface area is uniform among this set of proteins, with the majority of the surface being nonpolar in agreement with previous studies [26,28,31,34]. The numbers in parentheses in Table 2 are results from Chothia [28] for proteins common to both studies. Average results from a more recent study by Chothia and coworkers [31] on 11 proteins spanning a larger range of sizes (140–1992 residues) than those used here are also given. There is good agreement between the results of these three studies.

Model solute calculations

The simple linear relationship obtained between ΔV_{sol} and protein surface area motivated us to examine whether the volumetric properties for the proteins could be reconstructed from the corresponding volumetric properties of constituent groups. Model solutes were chosen to study the solvation of different types of surfaces individually. A methane molecule, a water molecule and a chloride ion were used to model nonpolar, polar and charged solvation. For these small, spherical solutes the number of solvent molecules in the hydration shell and their partial molar volume can be calculated directly from the solute–solvent radial distribution function. The radial distribution functions for the model compounds were determined both from experiments and simulations reported in the literature (L.R.M. and R.M.L., unpublished observations; [35–37]). The experimental partial specific volumes are

Figure 2



$\Delta V_{sol} = \bar{v}_2 - v_2$ versus solvent-accessible surface area for 15 globular proteins. Calculations employed the excluded volume with a probe radius of 1.4 Å.

Table 2

Surface area decomposition.

Protein	Nonpolar surface area* (%)	Polar surface area* (%)	Positive surface area* (%)	Negative surface area* (%)	Charged surface area* (%)
Insulin (monomer)	61	27	4	7	11
Pti	56(56)	21(22)	18	5	23(21)
α-Lactalbumin	57	22	9	12	22
Ribonuclease A	57	28	10	5	15
Lysozyme	54	28	15	3	19
Myoglobin	62	16	12	10	22
Adenylate kinase	55	18	14	13	26
Papain	55	31	11	3	14
Bence–Jones REI	57(57)	33(34)	6	4	10(9)
Concanavalin A (monomer)	56	28	8	8	15
Elastase	53(50)	36(39)	8	3	11(11)
Carbonic anhydrase B	57	26	9	8	17
Subtilisin	60	33	4	3	7
Rhodanese	59	20	12	10	22
Carboxypeptidase A	56(55)	29(32)	9	6	15(13)

*Solvent-accessible surface areas are calculated with a 1.4 Å probe radius. Numbers in parentheses are from [28]. Average proportions (%) in: this work, 57 ± 3 (nonpolar), 26 ± 6 (polar) and 17 ± 6 (charged); [28], 55 ± 3 (nonpolar), 32 ± 6 (polar) and 14 ± 5 (charged); and [31], 57 ± 4 (nonpolar), 27 ± 4 (polar) and 18 ± 5 (charged).

also available for these solutes. We now consider the relationship between ΔV_{sol} and ΔV_{hs} for the model solutes.

For each of the solutes the hydration shell is defined by the spherical annulus with inner radius r_L, the radius that defines the excluded volume, and outer radius r_U, which is defined by the first minimum in the radial distribution function. With this definition, the intrinsic solute volume is the volume that is excluded to the centers of the solvent molecules. The experimental partial molar volumes are listed in Table 3a. Using these values, we estimate ΔV_{sol} to be –90 Å^3, –61 Å^3 and –106 Å^3 for the nonpolar, polar and charged model compounds, respectively. The number of solvent molecules in the hydration shell of each of the model solutes and their partial molar volumes calculated from the radial distribution functions are also listed in Table 3a. Using these values, ΔV_{hs} is calculated to be –91 Å^3, –59 Å^3 and –105 Å^3 for the nonpolar, polar and charged solutes, respectively. Thus, for the model solutes we find that $\Delta V_{sol} \approx \Delta V_{hs}$ as predicted (Equation 3).

In order to apply the model solute volumetric data to the interpretation of the results for the proteins, ΔV_{sol} must be normalized by the size of the solute. The most straightforward procedure is to normalize by the accessible surface area of the model solutes, as was done for the protein data. Calculation of ΔV_{hs}/SA for each of the model solutes using accessible surface area, however, leads to values that are too small (in absolute value) relative to the value obtained for the proteins (Table 3a) and gives a non-intuitive ordering of the solutes (i.e. using the accessible surface area the polar solute has the least effect on the solvent volume followed by nonpolar and then charged). This is inconsistent with the model solute radial distribution functions and radial distribution functions between protein surface atoms and solvent, which show a high degree of localization of solvent near polar and charged surface but little near nonpolar surface (see low pressure results of Kitchen *et al.* [38]). Alternatively, we can normalize by an effective surface area, which depends on the number of solvent molecules in the hydration shell (see the Materials and methods section); this data is available for the model solutes but not for the proteins. The values ΔV_{hs}/SA using the effective surface area are also listed in Table 3a. When values based on effective surface areas for the nonpolar, polar and charged solutes are combined in the proportions given by the protein surface area decomposition, we obtain a result (–1.1 ± 0.15; Table 3b) in reasonable agreement with the protein data (–1.3 ± 0.05; Figure 2). The procedure used to normalize the model solute data and the protein data are different, however. The effect of using the accessible surface for normalization of the protein data, while using the effective surface to normalize the model compound data, is to reduce the contribution of charged groups to the protein volume change relative to the value calculated for the model charged solute (chloride ion). This is also true to some extent for the polar groups. Qualitatively, this may be explained by nonadditive effects related to electrostriction. The partial molar volume of solvent at the surface of an ion pair, for example, will be smaller than around the model (bare chloride) ion. A quantitative analysis of this point requires a more detailed analysis of the volumetric properties of a larger series of model compounds.

Solution volume partitioning into protein and solvent contributions

Previous volumetric studies have partitioned protein partial specific volume into contributions from the intrinsic

Table 3

Summary of calculated hydration-shell volume changes.

(a) Model solutes

Solute type	Experimental psv* (Å^3/molecule)	ΔV_{sol}† (Å^3/molecule)	N_h‡ (molecules)	RDF psv§ (Å^3/molecule)	ΔV_{hs}¶ (Å^3/molecule)	ASA# (Å^2)	ΔV_{hs}/SA using ASA (Å^3/Å^2)	ESA** (Å^2)	ΔV_{hs}/SA using ESA (Å^3/Å^2)
Nonpolar	61	−90	20	60	−91	137	−0.65	156	−0.57
(r_U = 5.4 Å; r_L = 3.3 Å)		(+32)							
Polar	30	−61	4.4	33	−59	99	−0.60	34	−1.7
(r_U = 3.4 Å; r_L = 2.8 Å)		(+19)							
Charged	30.9	−106	7.2	32	−105	129	−0.81	56	−1.9
(r_U = 3.9 Å; r_L = 3.2 Å)		(+6)							

(b) Proteins

Volume	ΔV_{sol}/SA††
Excluded volume	−1.3 ± 0.05 (0.99)
Molecular volume	0.0073 ± 0.12 (0.0041)
Van der Waals volume	0.78 ± 0.19 (0.94)
Estimate from surface area decomposition and model solutes (psv)	−1.1 ± 0.15

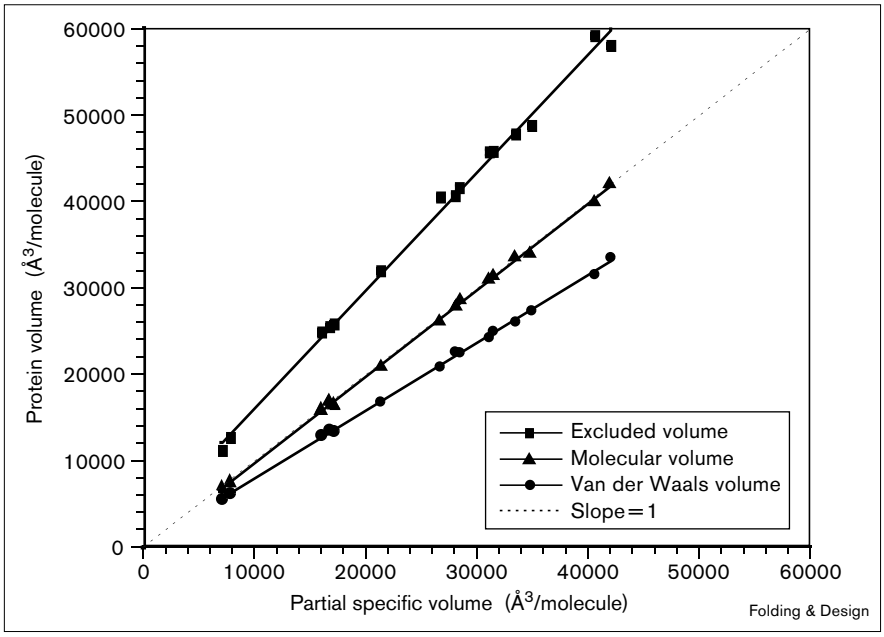
*From [81,82]. †ΔV_{sol} = partial specific volume (psv) − intrinsic volume = psv − (4/3)(πr_L^3). Values in parentheses are for r_L = van der Waals volume. ‡Number of hydration-shell waters, obtained by integrating the first peak in the solute–solvent radial distribution function (RDF). §RDF psv = (4/3)πr_L^3 + [(4/3)πr_U^3 − (4/3)πr_L^3] − $N_h v_1^o$ = (4/3)πr_U^3 − $N_h v_1^o$. ¶ΔV_{hs} = [(4/3)πr_U^3 − (4/3)πr_L^3] − N_h v_1^o. #ASA (solvent-accessible surface area) = $4\pi r_L^2$. **ESA (effective surface area) = N_h × (effective area per water molecule). ††Slope of the plot of ΔV_{sol} versus ASA. Numbers in parentheses are the correlation coefficients of the linear least squares fits.

protein volume and protein–water interactions [21,39–46]. But these studies did not address the different ways in which this partitioning can be accomplished with respect to conclusions that can be drawn about volumetric properties. When the volume of the solute is calculated, the space occupied by the solution is effectively divided into non-overlapping regions occupied by the solute and the solvent. But the partition of the solution space into solute volume and solvent volume is not unique. Protein volume calculations have primarily involved the use of three definitions of volume: excluded, molecular and van der Waals. Each of these definitions yields different values for the protein volume because the dividing surface between solute and solvent is different. The smallest protein volume is obtained using the van der Waals surface of the solute as the dividing surface, which traces the surface of van der Waals spheres of the constituent atoms. The van der Waals volume depends only on the van der Waals radii of the atoms and is independent of the size of the solvent. With this definition, the entire interfacial volume between the solute and solvent is assigned to the solvent regardless of whether an interfacial region is large enough to contain a solvent molecule. The largest value for the intrinsic volume of the protein is obtained using the excluded volume, in which the entire interfacial region between the solute and solvent that is inaccessible to the solvent atomic center is assigned to the solute volume. Intermediate between the van der Waals and excluded volumes is the molecular volume, which is composed of the protein atom volumes plus the interfacial region that is inaccessible to the surface of the solvent sphere. Both the excluded and molecular volumes depend on the size of the solvent molecules. Figure 3 shows the excluded, molecular and van der Waals volumes calculated for the 15 proteins versus their experimental partial specific volumes. A linear relationship is obtained with all three volume definitions; the difference in the three volume definitions is reflected in the slopes. The molecular volumes are approximately equal to the partial specific volumes yielding a slope close to one, whereas the excluded volume slope is greater than one and the van der Waals volume slope is less than one. The variation in slope corresponds conceptually to movement of the dividing plane between the protein and solvent.

Another approach to studying protein volume uses Voronoi polyhedra [40,47]. The method of Voronoi polyhedra involves the calculation of the volume of an atom in a molecule by building a polyhedron around the atom from planes placed on the interatomic vectors to neighboring atoms. The method is well-defined only for buried atoms. To calculate the volume of surface atoms, molecular dynamics simulations have been used [48,49] to provide reasonable placement of water molecules around the solute, which is necessary to locate the dividing planes between surface atoms and water. This method was not explored here, but a similar issue to that above is encountered when deciding where to place the plane between surface atoms and water; for example, bisecting the interatomic vector and placement based on atomic radii. Different placements of the planes change the dividing surface between the protein and solvent.

Figure 3

Calculated protein volumes versus the experimental partial specific volume using a 1.4 Å probe radius.



The choice of dividing surface between the protein and solvent affects the determination of the difference between the partial specific volume and the intrinsic volume of the protein. As shown above, calculation of ΔV_{sol} using the excluded volume definition as the intrinsic volume of the protein gives negative values of ΔV_{sol}; if the protein molecular volume or van der Waals volume were to be used to define the protein intrinsic volume, ΔV_{sol} would be close to zero or positive, respectively. Calculation of ΔV_{sol} for the model solutes is consistent with the protein findings with respect to the sign of the hydration-shell contribution (see Table 3).

The difference between the partial specific volume and the calculated protein volume can be plotted versus accessible surface area using the molecular and van der Waals volumes as the intrinsic volume of the protein. Figure 4 shows the plots of ΔV_{sol} versus surface area for the excluded volume (from Figure 2) along with that of the molecular and van der Waals volumes; a summary of the calculated protein volumes is given in Table 4. The slopes of the ΔV_{sol} versus surface area plots along with their correlation coefficients are collected in Table 3b. (The correlation coefficient is poor in the molecular volume case because the experimental and calculated volumes are very close in value and their difference is small compared to the volumes themselves.) A linear relationship with a high correlation coefficient is obtained for the plots based on the protein excluded volume and the van der Waals volume. Linearity indicates that the hydration-shell waters have the same average volume for all the proteins; Figure 4 shows that this property is not dependent on how the solution is partitioned into protein and solvent components. The slope of the plots, however, vary greatly depending on which protein volume definition is used. The sign of the slope of ΔV_{sol} versus surface area is model dependent; this reflects the fact that the excluded volume of a protein is larger than its partial molar volume whereas the van der Waals volume is smaller.

Volume change on protein unfolding

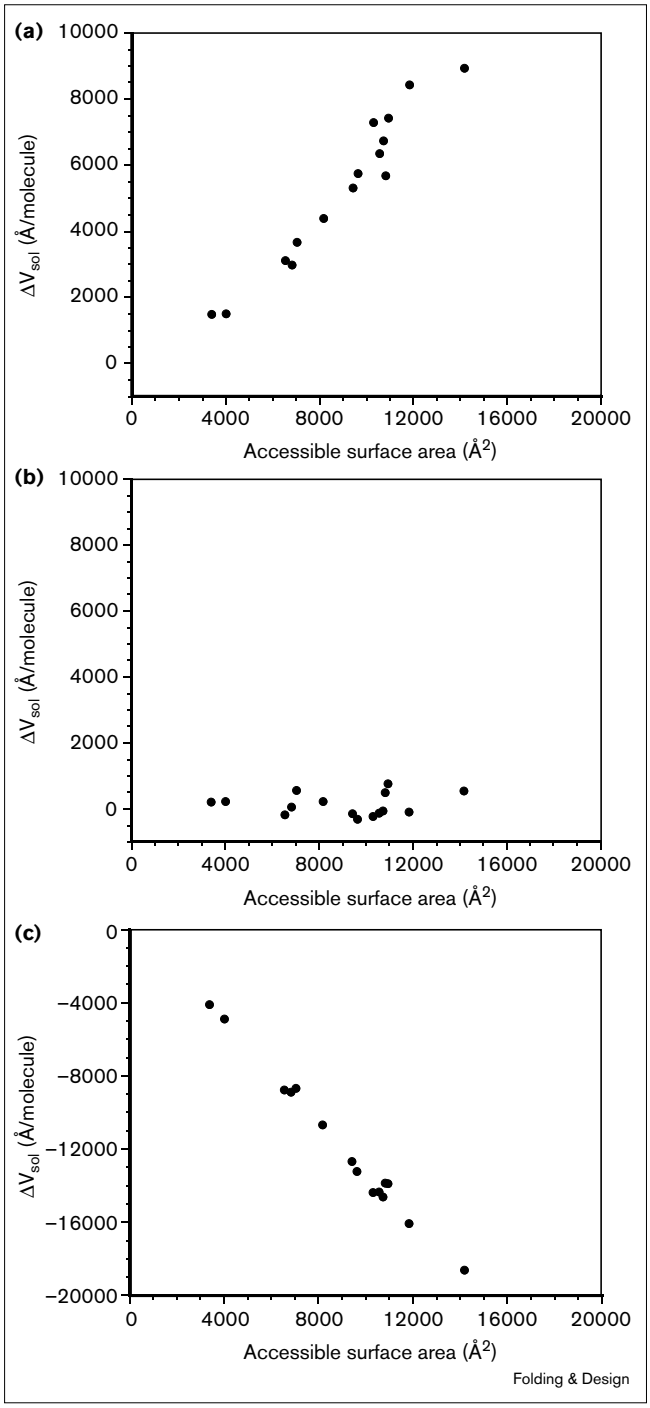
The partial specific volume of a protein is composed of two terms — a protein volume term and a term reflecting the change in molar density in the hydration shell. When the excluded volume is used to define the intrinsic protein volume, the hydration-shell term contributes negatively to the partial specific volume. The change in volume when the protein unfolds, ΔV_{unfold}, will be formulated analogously to partial specific volume as being composed of two terms, the change in the excluded volume of the protein and the change in the hydration-shell volume upon unfolding:

$$\Delta V_{unfold} = \Delta V_{prot} + \Delta\Delta V_{hs} \qquad (4)$$

where $\Delta V_{prot} = V^D - V^N = v_2^D - v_2^N$, $\Delta\Delta V_{hs} = \Delta V_{hs}^D - \Delta V_{hs}^N$, $\Delta V_{hs} = N_h(v_1 - v_1^o)$; the D denotes the unfolded state and the N denotes the native state. Using the relationship between the partial specific volume and the hydration-shell volume explored above for the native proteins and expressing $\Delta\Delta V_{hs}$ in terms of the surface areas of the unfolded and native proteins gives:

$$\Delta\Delta V_{hs} = \Delta V_{sol}^D - \Delta V_{sol}^N = \frac{\Delta V_{sol}^D}{SA^D} SA^D - \frac{\Delta V_{sol}^N}{SA^N} SA^N \qquad (5)$$

Figure 4



$\Delta V_{sol} = \bar{v}_2 - v_2$ versus solvent-accessible surface area for 15 globular proteins using **(a)** the van der Waals volume, **(b)** the molecular volume, and **(c)** the excluded volume calculated with a 1.4 Å probe radius.

Assuming that the surface composition of D is approximately the same as the surface composition of N, the term $\Delta V_{sol}/SA$ will be similar for native and denatured proteins and can be obtained from the slope of the ΔV_{sol} versus SA plot for the native proteins. ΔV_{unfold} can then be expressed in terms of the change in surface area between D and N (where $\Delta SA = SA^D - SA^N$):

$$\Delta V_{unfold} = \left(\frac{\Delta V_{prot}}{\Delta SA} + \frac{\Delta V_{sol}}{SA} \right) \Delta SA \qquad (6)$$

The volume change of unfolding is expressed as the change in surface area multiplied by the sum of protein and hydration coefficients. Given a model for the structure of the unfolded state of the protein, the excluded volume and accessible surface area can be calculated for the unfolded proteins as was done previously for the native protein crystal structures. This formulation depends on the assumption that protein unfolding is a two-state process; in other words, a transition from the native state to a single predominant unfolded state. Testing the two-state approximation is beyond the scope of this investigation. Another assumption is that the surface of the unfolded state has a similar composition as that in the native state. Here, the unfolded state will be modeled as maximally exposed (see below). Calculation of the surface area decomposition for these unfolded model proteins and comparison to the native protein decomposition (see Table 5) supports this. The similarity in composition of the extended chain to that of native proteins has also been observed previously [25,50].

Both the excluded volume and the hydration-shell contributions to ΔV_{unfold} are highly dependent on the model used for the unfolded state. Knowing that for globular proteins the volume change of unfolding is small and negative, the quality of an unfolded model can be judged by using Equation 6 to predict the unfolding volume ΔV_{unfold}; the comparison of calculated ΔV_{unfold} with experimental ΔV_{unfold} provides a way of testing models of the denatured state of proteins. The unfolded model proteins have been

Table 4

Summary of calculated protein volumes (Å^3/molecule).

Protein	Excluded (1.4 Å probe)	Molecular (1.4 Å probe)	Van der Waals
Insulin (monomer)	11142	6848	5564
Pti	12650	7540	6266
α-Lactalbumin	25760	16526	13435
Ribonuclease A	24862	15927	12989
Lysozyme	25464	16875	13595
Myoglobin	31960	21056	16880
Adenylate kinase	40504	26144	20964
Papain	40643	28116	22670
Bence–Jones REI	41554	28646	22592
Concanavalin A (monomer)	45697	31127	24341
Elastase	45740	31528	25047
Carbonic anhydrase B	48766	34107	27443
Subtilisin	47785	33649	26124
Rhodanese	59171	40012	31620
Carboxypeptidase A	58055	42098	33571

Table 5

Surface area decomposition of unfolded proteins compared to native proteins.

Surface area*	Native (%)	Unfolded model (%)
Nonpolar	57 ± 3	60 ± 2
Polar	26 ± 6	31 ± 4
Charged	17 ± 6	9 ± 3

*Solvent-accessible surface area calculated with a 1.4 Å probe radius.

generated here assuming a maximally exposed unfolded state. Values of accessible surface area and excluded volume were calculated for the 20 amino acid residues X in tripeptides Gly–X–Gly constructed as described in [28]. Protein surface area and excluded volume were calculated by summing over the constituent residue values. The resulting surface areas and excluded volumes are given in Table 6. The unfolded state quantities were calculated as the sum of constituent amino acid values; this is analogous to the experimental practice of obtaining protein thermodynamic measurements from model small molecule or peptide values (e.g. [51]).

Modeling the unfolded proteins from tripeptides produces an increase in the excluded volume and accessible surface area relative to the native proteins. With this model, 35 ± 5% of the unfolded surface area is accessible in the native state. This is in agreement with Richards' [52] estimate that on folding a completely extended chain the surface area will decrease to about one third of its value. The protein itself occupies more (excluded) volume in the unfolded state, but because there is more surface to interact with more solvent is perturbed in the hydration shell.

Protein unfolding is accompanied by a small volume change at normal pressures, which becomes negative at high pressures [31,45,53–56]. For proteins, the volume change on denaturation has been found experimentally to range from –50 cm^3/mol to –300 cm^3/mol [46,52], which is < 1% of the protein volume. The absolute value of the volume change on unfolding is very small compared to either the protein volume or the partial specific volume. Calculating ΔV_{unfold} therefore involves computing differences of large numbers to obtain small numbers, which will be sensitive to errors in the volume measurements. Thus, in the following analysis of ΔV_{unfold} the results will be judged not on the basis of quantitative agreement with individual experimental values but on the basis of yielding volume changes for protein unfolding of reasonable absolute value.

The results of the calculation of ΔV_{unfold} are presented in Table 7. The absolute values of ΔV_{unfold} show an overall trend of increasing with protein size and the predicted volume changes for unfolding are much larger than

Table 6

Solvent-accessible surface area and excluded volume of unfolded model proteins.*

Protein	Surface area (Å^2)	Excluded volume (Å^3/molecule)
Insulin (monomer)	8608	15827
Pti	9899	18041
α-Lactalbumin	21048	38674
Ribonuclease A	20676	37736
Lysozyme	21615	39522
Myoglobin	26427	48415
Adenylate kinase	33222	60745
Papain	35319	64764
Bence–Jones REI	35256	64720
Concanavalin A (monomer)	38390	70554
Elastase	38889	71459
Carbonic anhydrase B	42670	78341
Subtilisin	41665	76277
Rhodanese	49661	91125
Carboxypeptidase A	51895	95422

*Calculated with a 1.4 Å probe radius.

observed in experiments. These results suggest that the maximally exposed unfolded state is not a good model for the denatured state of proteins.

We can consider variations in volume and surface area of the unfolded model that are necessary to bring predicted protein volume changes upon unfolding based on Equation 6 into accord with experiment. According to Equation 6, the change in volume upon unfolding depends on two factors: the change in surface area upon unfolding, and the ratio of volume to surface area for the unfolded state. If the change in surface area upon unfolding is decreased while holding the ratio of volume to surface area of the unfolded state constant, the predicted volume changes deviate even further from experiment. In order to obtain predicted volume changes with unfolding, which are small (~1% of the native volume), it is necessary to reduce the surface area of the unfolded model while increasing the ratio of volume to surface area. Table 7 shows the effect of reductions of SA^D by 10% to 24% on ΔV_{unfold}. All reductions of SA^D cause the absolute value of ΔV_{unfold} to be reduced relative to the maximally exposed model values, and reducing SA^D by 23% reduces the calculated ΔV_{unfold} to within the experimentally determined range. This demonstrates that decreasing the surface area and increasing the volume to surface area ratio of the unfolded model proteins is sufficient modification of the maximally unfolded model to produce reasonable unfolding volume changes. This also suggests an explanation for why the maximally exposed model is not a good description of the denatured state. An improved model of the denatured state has a higher volume to surface area ratio than the maximally unfolded model. Increasing the volume to surface area ratio of the unfolded model effectively makes it more native-like. When the surface area of the maximally exposed

Table 7

ΔV_{unfold} for unfolded model proteins.*

Protein	ΔV_{unfold} (Å^3/molecule)	ΔV_{unfold} with SA^D reduced by 10%	20%	21%	22%	23%	24%
Insulin (monomer)	−2023	−1082	−243	−169	−96	−32	62
Pti	−2267	−1206	−268	−187	−106	−35	69
α-Lactalbumin	−5416	−2945	−677	−473	−271	−90	177
Ribonuclease A	−5379	−2931	−676	−473	−271	−90	177
Lysozyme	−5860	−3216	−749	−524	−302	−101	197
Myoglobin	−7117	−3902	−907	−635	−365	−122	238
Adenylate kinase	−8711	−4754	−1098	−768	−441	-−147	287
Papain	−10088	−5584	−1316	−923	−532	−177	348
Bence–Jones REI	−9947	−5495	−1292	−905	−522	−174	341
Concanavalin A (monomer)	−10770	−5944	−1396	−978	−564	−188	368
Elastase	−11017	−6090	−1433	−1004	−579	−193	378
Carbonic anhydrase B	−12360	−6857	−1621	−1137	−656	−219	429
Subtilisin	−12183	−6768	−1603	−1124	−650	−217	425
Rhodanese	−13840	−7630	−1789	−1253	−722	−241	471
Carboxypeptidase A	−15600	−8703	−2074	−1455	−842	−281	551

*ΔV_{unfold} is calculated using the excluded volume and the solvent-accessible surface area calculated with a 1.4 Å probe radius.

unfolded model is reduced by 23%, the volume to surface area ratio is increased to 2.39; this is slightly less than midway between the maximally exposed model value of 1.83 and the ratio for native proteins, 3.57.

In summary, the above analysis of the volume change of protein unfolding has shown that an increase in the volume to surface area ratio with respect to a maximally unfolded model is needed to obtain results in agreement with the experimental finding that the volume change of the solution for protein unfolding is approximately zero. Based on volumetric data, denatured proteins are predicted to have ~80% of the surface area of the fully exposed model. Because the surface area of native proteins is ~35% that of the fully unfolded model, denatured proteins are predicted to have a solvent-accessible surface area that is slightly more than twice that of the native proteins. The volume to surface area ratio of denatured proteins is predicted to be about two thirds that of native proteins.

Discussion

When a solute is added to solution, there are three contributions to the measured volume change: the volume change resulting from the solute, the volume change resulting from the perturbation of the solvent molecules closest to the solute (the first hydration shell), and the remaining volume change resulting from the perturbation of more distant solvent molecules. All these contributions are taken into account in the statistical mechanical formulation of excess volume [30]:

$$\Delta V = \int_v d\vec{r}\,\rho(\vec{r})\left(\frac{1}{\rho(\vec{r})} - \frac{1}{\rho(\infty)}\right) + o(1) \qquad (7)$$

where $\rho(\vec{r})$ is the one-particle distribution of the solvent at the point $\vec{r}$, and $\rho(\infty)$ is the asymptotic value of $\rho(\vec{r})$. Introducing a cutoff in the above integral is equivalent to localizing the solvent perturbation within a particular distance from the solute; for example, within the first hydration shell. With this cutoff, and replacing $\rho(\infty)$ with the bulk density and neglecting o(1) terms in the thermodynamic limit, the formulation of the partial specific volume is as in Equation 2. Thus, according to the strict statistical mechanical definition of excess or partial specific volume, Equation 2 is an approximation brought about by assuming that the solvent perturbation is localized to the hydration shell. The successful correlation between the change in volume ΔV_{sol} and the surface area implies that for the proteins studied approximately the same proportion of the total volume change resulting from hydration is localized mainly in the first hydration shell.

Early in the study of proteins, Kauzmann [57] reasoned that because nonpolar atoms are hydrophobic they should be preferentially located in the protein interior, whereas the surface should be enriched in hydrophilic groups. Despite studies to the contrary, the common conception of proteins is still that hydrophobic residues primarily cluster in the interior of proteins and polar and charged residues lie on the surface. This work and previous studies [24,25,28,31,34] demonstrate that a high proportion of the protein surface is nonpolar, in fact > 50% of the total surface. Rose *et al.* [58] studied the distribution of buried surface area for each residue type. They found that as expected the charged residues tend to be on the protein surface, but only four nonpolar residues (phenylalanine, leucine, isoleucine and methionine) tend to be fully buried. The remaining nonpolar residues were found to be distributed throughout the proteins. Using the native protein and the extended unfolded model protein surface areas calculated here, the proportion of surface area buried

on going from the native to the unfolded state can be calculated. If nonpolar surface is compared to the sum of polar and charged surface, the same proportion of nonpolar surface is buried as charged plus polar surface ($67 \pm 6\%$ nonpolar, $63 \pm 4\%$ polar + charged), in agreement with previous studies [24,25,27]). Considering polar and charged surface separately, equal amounts of nonpolar surface and polar surface are buried, which is more than the charged surface buried ($67 \pm 6\%$ nonpolar, $71 \pm 3\%$ polar and $39 \pm 12\%$ charged). The amount of charged surface buried increases with protein size, whereas the nonpolar and polar surface areas buried show little size dependence [59]. Thus, in the analysis of protein properties it should be kept in mind that a large number of nonpolar groups remain on the surface; this greatly affects a protein's interaction with its environment. In fact, Richards [52] cautioned as early as 1977 that "the 'grease' is by no means all 'buried'" and that the situation requires a "more careful definition than is implied by the common feeling that inside equals nonpolar and outside equals polar". As discussed by Kitchen *et al.* [38], the relatively large nonpolar surface area will affect pressure-induced denaturation.

When protein is added to solution, the solution volume changes by an amount measured as the partial specific volume; part of this volume change results from the volume occupied by the protein itself, whereas part results from the change in the volume of water interacting with the protein. The choice of dividing surface to separate protein from solvent affects how the measured volume change is formally partitioned among the components of the solution; there is an unavoidable ambiguity in this partitioning. The linear relationship that we observe between ΔV_{sol} and solvent-accessible surface area, however, suggests that the average specific volume of water in the protein hydration shell is approximately the same for different proteins regardless of the manner in which the solution is partitioned into protein and hydration-shell components. The slope of the plot of ΔV_{sol} versus surface area, however, does depend in both magnitude and sign on the definition used for the protein volume. This implies that the density calculated for solvent in the hydration shell relative to the density of bulk water is strongly dependent on the definition of the dividing surface between protein and solvent. We therefore disagree that the sign of the solvent contribution to the partial molar volume of a protein must be positive, as is sometimes suggested [46].

Levitt and coworkers [48] recognized the significance of the allocation of space around the protein in their calculation of the volume of pancreatic trypsin inhibitor. Using molecular dynamics simulations to assign the positions of water molecules around the protein surface, they used the Voronoi polyhedron method to calculate the volume of interior and surface atoms. In the Voronoi procedure, each atom is surrounded by a polyhedron whose faces are formed from dividing planes perpendicular to the interatomic vectors. Levitt and coworkers explain the larger volume of some protein surface atoms relative to the interior in terms of the packing of water around the surface atoms; if water is not packed tightly, the resulting cavity makes the surface atom appear larger when the space is allocated to the protein. They proposed using different radii for water molecules situated around nonpolar, polar and carboxyl oxygen atoms because the 1.4 Å value usually employed reflects the hydrogen bonding in pure water. The deviation of the distributions of interatomic distances from the molecular dynamics simulation compared to the expected interatomic distance from the sum of van der Waals radii was used to assign new water radii (nonpolar 1.96 Å, polar 1.23 Å and carboxyl oxygen 1.08 Å). The adjusted radii are qualitatively comparable to the results reported here for estimating ΔV_{hs} based on the use of radial distribution functions for model compounds.

Because the change in volume on protein unfolding is very small it is difficult to observe experimentally by direct volumetric methods [51]. The best way to probe unfolding volume changes is through the use of pressure. Recently, the effect of pressure on the hydrophobic interaction in a simple model system was examined computationally, demonstrating that the trend of a hydrophobic dimer to dissociate at pressures of several kbar is consistent with a hydrophobicity-driven mechanism of pressure-induced protein denaturation [60]. This effect depends on the sign of the excess compressibility of water in the hydration shell, rather than the density [30]. Other than the work described in [60], there have been few computational studies on the effects of pressure on hydrophobic interactions. Experimental studies of the effect of pressure on protein unfolding include the work of Markley and coworkers [61], who measured the relaxation kinetics of the folding and unfolding of staphylococcal nuclease and the activation volumes for the transitions between the native, molten globule and denatured states using pressure-jump experiments. They explain pressure-induced protein unfolding as resulting from the effects of increased solvation and decreased molecular volume of the protein (which they refer to as excluded volume). Both these effects serve to reduce the volume of the unfolded system relative to the native system because the application of pressure will drive the system towards the point of least volume. In their analysis, the increased hydration upon unfolding contributes positively to the unfolding volume change and is compensated for by a negative protein volume term to yield a small net unfolding volume change. In the analysis of protein unfolding presented here, the reverse is true — the change in protein volume makes a positive contribution to ΔV_{unfold} whereas the change in hydration makes a negative contribution to the unfolding volume change; again the net effect is to

produce a small volume change. Chalikian and Breslauer [21], using yet another model to partition the volume of the protein solution, conclude that the small negative volume of unfolding observed for proteins arises from the opposition of a positive "thermal volume change" with negative contributions from the change in void volume and from a term arising from solute–solvent interactions.

The different partitioning schemes used to divide the solution volume into the solute and solvent occupied volumes used in these three studies demonstrate how the choice affects the interpretation of the origin of the small volume change. The solute and solvent volumes are not, however, separate thermodynamic observables and the partition of the solution volume into solute and solvent components is not unique. The use of a variety of partitioning schemes is consistent with the experimental data, even though the meaning of the individual component terms is largely phenomenological.

We have found that the model for the unfolded state, which is maximally exposed to solvent, is not consistent with experimental volume changes; a totally extended unfolded state leads to a predicted unfolding volume change much larger than that observed. The totally extended model is estimated to have a surface area almost 20% too large. Based on the modeling of volumetric properties reported here, the denatured state is predicted to have a solvent-exposed surface area slightly more than twice that of the native protein. Similar estimates of exposure in the denatured state have been obtained from compressibility [62] and calorimetric [63,64] measurements. The conclusion that the denatured state of a protein is not maximally unfolded, but retains a significant degree of structure, is thus supported by a strong body of evidence from both experimental and computational studies. Because a more compact denatured state will affect both the kinetic and thermodynamic aspects of protein folding and unfolding there is clearly a need for more direct probes of the molecular features of denatured proteins, through such techniques as NMR spectroscopy and computer simulation.

Materials and methods

Surface area and volume definitions

The van der Waals volume is the volume occupied by the atoms as represented by hard spheres with assigned radii. The molecular and excluded volumes and surface areas use a probe sphere (which represents a solvent molecule) rolling on the outside of the van der Waals envelope of the solute and maintaining contact with the surface. The solvent-accessible surface area is the area of the surface generated by the center of the probe sphere. The excluded volume is the volume enclosed by the solvent-accessible surface. The molecular surface is the surface traced out by the inward-facing part of the probe sphere and is not displaced from the van der Waals surface. The molecular volume is the volume enclosed by the molecular surface and thus is the volume inaccessible to any part of the spherical probe. Both the excluded volume and the molecular volume are dependent on the size of the probe sphere (i.e. they are solvent dependent) whereas the van der Waals volume is solvent independent. The solvent-accessible surface and molecular surface approach the van der Waals surface as the probe radius approaches zero. For the special case of a spherical molecule, the molecular surface is the same as the van der Waals surface. The various quantities defined above are illustrated in Figure 1.

Protein calculations

A set of 15 small to moderate sized proteins were chosen on the basis of the availability of experimentally determined partial specific volumes and of X-ray crystal structures from the Brookhaven PDB [65,66]. Most of the crystal structures were of better than 2.0 Å resolution. The set of proteins contains representatives from all classes of proteins – $\alpha + \beta$ (five proteins), α/β (four proteins), all α (two proteins), all β (four proteins) – and many different protein folds [67]. The partial specific volumes, measured at 20°C, were taken from compilations by Smith, Creighton and Hinz, as reported in units of cm^3/g [68–70]. The crystal structures were used to calculate the protein surface areas and volumes.

Volumes and surface areas were calculated using a program written by Jay W. Ponder which uses the algorithms from the AMS/VAM programs of Connolly [71–73] and implements Richards' molecular surface definition [52]. This analytical method constructs the protein surface as a collection of polygons for which the surface area and volume can then be computed. To obtain the solvent-accessible surface for excluded volume and solvent-accessible surface area the radius of each solute atom was taken as the sum of the van der Waals radius and the effective radius of the solvent (i.e. probe radius) and the volume and surface area calculated using a probe radius of zero. Molecular and van der Waals volumes were calculated using the van der Waals radius for the radius of each solute atom and using a probe radius of 1.4 Å (the effective radius of the solvent) or 0 Å, respectively. Solvent-accessible surface areas were verified by comparison with those from the molecular simulation program IMPACT [74] and the surface area decomposition was also performed using IMPACT. Van der Waals radii were from McCammon *et al.* [75] for extended atoms. Comparison of surface area calculated using these radii and the set employed in IMPACT [76] show no significant difference. The calculated volumes were verified by comparison with literature values where available [39–41,72,77–80].

Molecular weights were calculated from the protein sequences and were used to convert the partial specific volume units from cm^3/g to $Å^3$/molecule. Surface area was decomposed into percentages of nonpolar, polar and charged using atom-based definitions similar to Chothia [28]: carbons were considered nonpolar; (terminal) sidechain nitrogens of arginine, lysine and N-terminal capping residues, and (terminal) sidechain oxygens of aspartic acid, glutamic acid and C-terminal capping residues were considered charged; and the remaining nitrogen, oxygen and sulfur atoms were considered polar.

Model solute calculations

In order to model nonpolar, charged and polar solvation, calculations have been performed for the solutes methane, chloride ion and water, respectively, in water solvent.

The experimental partial specific volumes of the solutes and the calculated solute volumes can be used to calculate ΔV_{sol} analogously to the proteins:

$$\Delta V_{sol} = \bar{v}_2 - v_2 \tag{8}$$

where $\bar{v}_2$ is the partial specific volume. The excluded volume for these spherical solutes was calculated from the definition of the volume of a sphere:

$$v_2 = \left(\frac{4\pi}{3}\right) r^3 \tag{9}$$

For these small, spherical solutes the total volume of the hydration shell was calculated from the radial distribution function between the solute and the water oxygens:

$$V_{hs} = \left(\frac{4\pi}{3}\right) r_U^3 - \left(\frac{4\pi}{3}\right) r_L^3 \tag{10}$$

where r_U and r_L are the upper and lower bounds of the first peak in the solute–solvent radial distribution function. Note that r_U defines how far out from the solute the hydration shell extends, while r_L defines the solute volume. The hydration-shell volume change was calculated as the difference between V_{hs} and the volume occupied by an equivalent number of bulk water molecules:

$$\Delta V_{hs} = V_{hs} - N_h v_1^0 \quad (11)$$

ΔV_{hs} was normalized by surface area for comparison with protein results. The solvent-accessible surface area (ASA) of the model solutes was obtained from the formula for the surface area of a sphere:

$$ASA = 4\pi r_L^2 \quad (12)$$

An effective surface area (ESA) was defined as the total area occupied by the number of hydration-shell water molecules given by N_h. The effective surface area was obtained by formulating a conversion factor, an area of a slice through the center of the water. This represents the amount of area the water molecule will occupy on the dividing surface between solute and solvent. Actually, this surface would be a curved circular piece of a spherical surface, but we approximate the area as the area of a flat square with sides 2R (where R is the radius of the water molecule). With a water radius of 1.4 Å the conversion factor is 7.8 Å^2. The effective surface area is then calculated as:

$$ESA = 7.8N_h \quad (13)$$

Acknowledgements

We thank Jay W. Ponder for providing us with his program to calculate protein volumes. This work has been supported in part by a grant from the National Institutes of Health (GM-52210) and by the Center for Advanced Food Technology at Rutgers University (CAFT). L.R.M. acknowledges a National Research Service Award from the National Institutes of Health.

References

1. Eisenberg, D. & McLachlan, A.D. (1986). Solvation energy in protein folding and binding. *Nature* **319**, 199-203.
2. Ooi, T., Oobatake, M., Nemethy, G. & Scheraga, H.A. (1987). Accessible surface area as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl Acad. Sci. USA* **84**, 3086-3090.
3. Ooi, T. & Oobatake, M. (1988). Effects of hydrated water on protein unfolding. *J. Biochem.* **103**, 114-120.
4. Khechinashvili, N.N. (1990). Thermodynamic properties of globular proteins and the principle of stabilization of their native structure. *Biochem. Biophys. Acta* **1040**, 346-354.
5. Makhatadze, G.I. & Privalov, P.L. (1990). Heat capacity of proteins I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. *J. Mol. Biol.* **213**, 375-384.
6. Makhatadze, G.I., & Privalov, P.L. (1993). Contribution of hydration to protein folding thermodynamics I. The enthalpy of hydration. *J. Mol. Biol.* **232**, 639-659.
7. Makhatadze, G.I. & Privalov, P.L. (1994). Hydration effects in protein unfolding. *Biophys. Chem.* **51**, 291-309.
8. Privalov, P.L. & Makhatadze, G.I. (1990). Heat capacity of proteins II. Partial molar heat capacity of unfolded polypeptide chains of proteins: protein unfolding effects. *J. Mol. Biol.* **213**, 385-391.
9. Privalov, P.L. & Makhatadze, G.I. (1992). Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. *J. Mol. Biol.* **224**, 715-723.
10. Privalov, P.L. & Makhatadze, G.I. (1993). Contribution of hydration to protein folding thermodynamics II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **232**, 660-679.
11. Livingstone, J.R., Spolar, R.S. & Record, M.T. (1991). Contributions to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area. *Biochemistry* **30**, 4237-4244.
12. Vila, J., Williams, R.L., Vasquez, M. & Scheraga, H.A. (1991). Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitor. *Proteins* **10**, 199-218.
13. Renner, M., Hinz, H.-J., Scharf, M. & Engels, J.W. (1992). Thermodynamics of unfolding of the α-amylase inhibitor tendamistat. Correlation between accessible surface area and heat capacity. *J. Mol. Biol.* **223**, 769-779.
14. Spolar, R.S., Livingstone, J.R. & Record, M.T. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31**, 3947-3955.
15. Yang, A.-S., Sharp, K.A. & Honig, B. (1992). Analysis of the heat capacity dependence of protein folding. *J. Mol. Biol.* **227**, 889-900.
16. Freire, E. (1993). Structural thermodynamics: prediction of protein stability and protein binding affinities. *Arch. Biochem. Biophys.* **303**, 181-184.
17. Makhatadze, G.I., Kim, K.-S., Woodward, C. & Privalov, P.L. (1993). Thermodynamics of bovine pancreatic trypsin inhibitor folding. *Protein Sci.* **2**, 2028-2036.
18. Koehl, P. & Delarue, M. (1994). Polar and nonpolar atomic environments in the protein core: implications for folding and binding. *Proteins* **20**, 264-278.
19. Khechinashvili, N.N., Janin, J. & Rodier, F. (1995). Thermodynamics of the temperature-induced unfolding of globular proteins. *Protein Sci.* **4**, 1315-1324.
20. Myers, J.K., Pace, C.N. & Schotz, J.M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138-2148.
21. Chalikian, T.V. & Breslauer, K.J. (1996). On volume changes accompanying conformational transitions of biopolymers. *Biopolymers* **39**, 619-626.
22. Hilser, V.J., Gomez, J. & Freire, E. (1996). The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins* **26**, 123-133.
23. Dill, K.A. (1997). Additivity principles in biochemistry. *J. Biol. Chem.* **272**, 701-704.
24. Richards, F.M., Wycoff, H.W., Carlson, W., Allwell, N.M., Lee, B. & Mitsui, Y. (1971). Protein structure, ribonuclease-S and nucleotide interactions. *Cold Spring Harbor Symp. Quant. Biol.* **36**, 35-43.
25. Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.
26. Shrake, A. & Rupley, J.A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351-371.
27. Chothia, C. (1975). Structural invariants in protein folding. *Nature* **254**, 304-308.
28. Chothia, C. (1976). The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**, 1-14.
29. Matubayasi, N., Reed, L.H. & Levy, R.M. (1994). Thermodynamics of the hydration shell. 1. Excess energy of a hydrophobic solute. *J. Phys. Chem.* **98**, 10640-10649.
30. Matubayasi, N. & Levy, R.M. (1996). Thermodynamics of the hydration shell. 2. Excess volume and compressibility of a hydrophobic solute. *J. Phys. Chem.* **100**, 2681-2688.
31. Harpaz, Y., Gerstein, M. & Chothia, C. (1994). Volume changes on protein folding. *Structure* **2**, 641-649.
32. Tinoco, I., Jr, Sauer, K. & Wang, J.C. (1985). Transport Phenomena. In *Physical Chemistry.* pp. 217-274, Prentice-Hall, Inc., NJ, USA.
33. Richmond, T.J. (1984). Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**, 63-89.
34. Lijnzaad, P., Berendsen, H.J.C. & Argos, P. (1996). Hydrophobic patches on the surfaces of protein structures. *Proteins* **25**, 389-397.
35. Palinkas, G., Kalman, E. & Kovacs, P. (1977). Liquid water. II. Experimental atom pair-correlation functions of liquid deuterium oxide. *Mol. Phys.* **34**, 526-531.
36. Narten, A.H., Thiessen, W.E. & Blum, L. (1982). Atom pair distribution functions of liquid water at 25°C from neutron diffraction. *Science* **217**, 1033-1034.
37. Impey, R.W., Madden, P.A. & McDonald, I.R. (1983). Hydration and mobility of ions in solution. *J. Chem. Phys.* **87**, 5071-5083.
38. Kitchen, D.B., Reed, L.H. & Levy, R.M. (1992). Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry* **31**, 10083-10093.
39. Klapper, M.H. (1971). On the nature of the protein interior. *Biochem. Biophys. Acta* **229**, 557-566.
40. Richards, F.M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1-14.
41. Pavlov, M.Y. & Fedorov, B.A. (1983). Improved technique for calculating X-ray scattering intensity of biopolymers in solution: evaluation of the form, volume, and surface of a particle. *Biopolymers* **22**, 1507-1522.
42. Gekko, K. & Hasegawa, Y. (1986). Compressibility-structure relationship of globular proteins. *Biochemistry* **25**, 6563-6571.

43. Kharakoz, D.P. (1992). Partial molar volumes of molecules of arbitrary shape and the effect of hydrogen bonding with water. *J. Sol. Chem.* **21**, 569-595.
44. Chalikian, T.V., Totrov, M., Abagyan, R. & Breslauer, K.J. (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.
45. Mozhaev, V.V., Heremans, K., Frank, J., Masson, P. & Balny, C. (1996). High pressure effects on protein structure and function. *Proteins* **24**, 81-91.
46. Prehoda, K.E. & Markley, J.L. (1996). Use of partial molar volumes of model compounds in the interpretation of high-pressure effects on proteins. In *High-Pressure Effects in Molecular Biophysics and Enzymology.* (Markley, J.L., Northrop, D.B. and Royer, C.A., eds), pp. 33-43, Oxford University Press, New York.
47. Finney, J.L. (1978). Volume occupation, environment and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* **119**, 415-441.
48. Gerstein, M., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**, 955-966.
49. Paci, E. & Velikson, B. (1997). On the volume of macromolecules. *Biopolymers* **41**, 785-797.
50. Miller, S., Janin, J., Lesk, A.M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641-656.
51. Makhatadze, G.I., Medvedkin, V.N. & Privalov, P.L. (1990). Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range. *Biopolymers* **30**, 1001-1010.
52. Richards, F.M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.
53. Brandts, J.F., Oliveira, R.J. & Westort, C. (1970). Thermodynamics of protein denaturation. Effect of pressure on the denaturation of ribonuclease A. *Biochemistry* **9**, 1038-1047.
54. Hawley, S.A. (1971). Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **10**, 2436-2442.
55. Zip, A. & Kauzmann, W. (1973). Pressure denaturation of metmyoglobin. *Biochemistry* **12**, 4217-4228.
56. Hinz, H.-J., Vogl, T. & Meyer, R. (1994). An alternative interpretation of the heat capacity changes associated with protein unfolding. *Biophys. Chem.* **52**, 275-285.
57. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14,** 1-64.
58. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. & Zehfus, M.H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229,** 834-838.
59. Helin, S., Kahn, P.C., Lakshmi Guha, L., Mallows, D.J., Steitz, T.A. & Goldman, A. (1995). The refined X-ray structure of muconate lactonizing enzyme from *Pseudomonas putida* at 1.85 Å resolution. *J. Mol. Biol.* **254**, 918-941.
60. Payne, V.A., Matubayashi, N., Murphy, L.R. & Levy, R.M. (1997). Monte Carlo study of the effect of pressure on hydrophobic association. *J. Phys. Chem.* **101**, 2054-2060.
61. Vidugiris, G.J.A., Markley, J.L. & Royer, C.A. (1995). Evidence for a molten globule-like transition state in protein folding from determination of activation volumes. *Biochemistry* **34**, 4909-4912.
62. Chalikian, T.V., Gindikin, V.S. & Breslauer, K.J. (1996). Spectroscopic and volumetric investigation of cytochrome *c* unfolding at alkaline pH: characterization of the base-induced unfolded state. *FASEB J.* **10**, 164-170.
63. Ahmad, F. & Bigelow, C.C. (1986). Estimation of the stability of globular proteins. *Biopolymers* **25**, 1623-1633.
64. Lee, B. (1991). Isoenthalpic and isoentropic temperatures and the thermodynamics of protein denaturation. *Proc. Natl Acad. Sci. USA* **88**, 5154-5158.
65. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
66. Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. & Weng, J. (1987). In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications.* (Allen, F.H., Bergerhoff, G. and Sievers, R., eds), pp. 107-132, Data Commission of the International Union of Crystallography, Bonn.
67. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
68. Smith, M.H. (1970). Molecular weights of proteins and some other materials Including sedimentation, diffusion and frictional coefficients and Partial Specific Volumes. In *Handbook of Biochemistry.* (Sober, H.A., ed.), pp. C3-C35, The Chemical Rubber Co., Ohio.
69. Creighton, T. (1983). *Proteins.* W.H. Freeman and Co., New York.
70. Hinz, H.-J. (1986). *Thermodynamic Data for Biochemistry and Biotechnology.* Springer-Verlag, New York.
71. Connolly, M.L. (1983). Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16**, 548-558.
72. Connolly, M.L. (1985). Computation of molecular volume. *J. Am. Chem. Soc.* **107**, 1118-1124.
73. Connolly, M.L. (1996). Molecular surfaces: a review. *NetSci* **2**. http://www.netsci.org/Science/Compchem/feature14.html.
74. Kitchen, D.B., Hirata, F., Westbrook, J.D., Levy, R.M. & Yarmush, M. (1990). Conserving energy during molecular dynamics simulations of water, proteins and proteins in water. *J. Comp. Chem.* **11**, 1169-1180.
75. McCammon, J.A., Wolynes, P.G. & Karplus, M. (1979). Picosecond dynamics of tyrosine sidechains in proteins. *Biochemistry* **18**, 927-942.
76. Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* **7**, 230-252.
77. Finney, J.L. (1975). Volume occupation, environment and accessibility in proteins. The problem of the protein surface. *J. Mol. Biol.* **96**, 721-732.
78. Gellatly, B.J. & Finney, J.L. (1982). Calculation of protein volumes: an alternative to the Voronoi procedure. *J. Mol. Biol.* **161**, 305-322.
79. Connolly, M.L. (1985). Molecular surface triangulation. *J. Appl. Crystallogr.* **18**, 499-505.
80. Kundrot, C.E., Ponder, J.W. & Richards, F.M. (1991). Algorithms for calculating excluded volume and its derivatives as a function of molecular conformation and their use in energy minimization. *J. Comp. Chem.* **12**, 402-409.
81. Ben-Naim, A. (1974). *Water and Aqueous Solutions.* Plenum Press, New York.
82. Padova, J. (1963). Ion-solvent interaction. II. Partial molar volume and electrostriction: a thermodynamic approach. *J. Chem. Phys.* **39**, 1552-1557.

Because *Folding & Design* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/fad – for further information, see the explanation on the contents pages.