PROTEINS: Structure, Function, and Genetics 26:123–133 (1996)

RESEARCH ARTICLES

The Enthalpy Change in Protein Folding and Binding: Refinement of Parameters for Structure-Based Calculations

Vincent J. Hilser, Javier Gómez, and Ernesto Freire

Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT **Two effects are mainly responsible for the observed enthalpy change in protein unfolding: the disruption of internal interactions within the protein molecule (van der Waals, hydrogen bonds, etc.) and the hydration of the groups that are buried in the native state and become exposed to the solvent on unfolding. In the traditional thermodynamic analysis, the effects of hydration have usually been evaluated using the thermodynamic data for the transfer of small model compounds from the gas phase to water. The contribution of internal interactions, on the other hand, are usually estimated by subtracting the hydration effects from the experimental enthalpy of unfolding. The main drawback of this approach is that the enthalpic contributions of hydration, and those due to the disruption of internal interactions, are more than one order of magnitude larger than the experimental enthalpy value. The enthalpy contributions of hydration and disruption of internal interactions have opposite signs and cancel each other almost completely resulting in a final value that is over 10 times smaller than the individual terms. For this reason, the classical approach cannot be used to accurately predict unfolding enthalpies from structure: any error in the estimation of the hydration enthalpy will be amplified by a factor of 10 or more in the estimation of the unfolding enthalpy. Recently, it has been shown that simple parametric equations that relate the enthalpy change with certain structural parameters, especially changes in solvent accessible surface areas have considerable predictive power. In this paper, we provide a physical foundation to that parametrization and in the process we present a system of equations that explicitly includes the enthalpic effects of the packing density between the different atoms within the protein molecule. Using this approach, the error in the prediction of folding/unfolding enthalpies at 60°C, the median temperature for thermal unfolding, is better than ±3% (standard deviation = 4 kcal/mol).**

Key words: enthalpy, thermodynamics, folding/unfolding

INTRODUCTION

A necessary requirement for the development of successful strategies in structure-based molecular design is the ability to predict accurately the energetics of stabilization in the case of protein folding or association in the case of binding. Since the enthalpy and the entropy changes define the Gibbs energy at any given temperature, it is of fundamental importance to develop accurate algorithms for enthalpy and entropy prediction from structural considerations.

In the past, most algorithms have been aimed at predicting Gibbs energies without taking into consideration the individual contributions of enthalpic and entropic components.[1–4] The estimation of hydrophobic free energies in terms of solvent accessible surface area (ASA) burial is one common example.[5–7] This approach, however, can only provide a first-order approximation to ΔG at a single temperature beyond which further refinement is not possible. In general, conformational entropy changes or long-range polar interactions associated with folding or binding do not scale in terms of ASA changes.[8,9] Furthermore, a separation of enthalpic and entropic contributions provides a better understanding of the origin of the magnitude of ΔG as well as its temperature dependence, hence providing an important tool for molecular design. In most cases, the achievement of more favorable free energies in molecular design implies overcoming ubiquitous enthalpy-entropy compensation effects that tend to minimize any desired changes in ΔG.

Received January 4, 1996; accepted March 29, 1996. Address reprint requests to Ernesto Freire, Department of Biology, The Johns Hopkins University, Baltimore, MD 21218.

Previously, we have shown that a simple empirical equation is able to quantitatively account for the enthalpy of protein folding with a high degree of accuracy:

$$\Delta H(T) = (a(T) \cdot \Delta ASA_{ap} + b(T) \cdot \Delta ASA_{pol}) + \Delta H_{prot} + \Delta H_{other}.$$

This empirical relationship indicates that at any given temperature the enthalpy of protein folding/unfolding can be expressed as a linear combination of the changes in solvent-accessible polar and apolar surface areas, plus the changes due to protonation and other specific effects. The parameters a and b were obtained from a global fit of protein enthalpy data.[10–12] At 60°C, the median temperature for protein thermal unfolding, the enthalpy change can be predicted with an accuracy better than ±6%. The purpose of this paper is to provide a physical foundation to that parametrization and, in the process, propose new strategies for refinement. We show that the simple ΔASA scaling of the enthalpy change is a consequence of the structural similarities existing between different proteins: they have similar atomic compositions, similar distributions of pairwise interactions and similar packing densities. These studies provide an explicit way to incorporate the effect of packing density on the calculation of the folding enthalpy in proteins.

RESULTS AND DISCUSSION

Classic Thermodynamic Cycle

Two effects are mainly responsible for the observed enthalpy change in protein unfolding: the disruption of internal interactions within the protein molecule (van der Waals, hydrogen bonds, electrostatic interactions, etc.) and the hydration of the groups that are buried in the native state and become exposed to the solvent upon unfolding. The traditional thermodynamic approach has been to represent the unfolding process by a cycle similar to that shown in Figure 1 (see, e.g., refs 13–15). In this cycle, the horizontal steps represent the folding/unfolding of the protein in vacuum (top) or in aqueous solution (bottom), and the left and right vertical steps represent the hydration of the native and unfolded states, respectively. In this cycle the only term that can be measured directly is the folding/unfolding reaction in aqueous solution (step 4 in Fig. 1). The hydration steps have usually been approximated by using thermodynamic values measured for small organic molecules. These values are used to estimate the hydration energetics of specific chemical groups found in proteins and are usually expressed on a per mole of Å^2 basis.[14,16] The hydration energetics of the entire protein is then estimated by assuming group additivity, that is, by estimating the total area of a given type that becomes exposed to the solvent and multiplying it by the elementary value obtained from small compounds.[14,16] The disruption of internal interactions is usually estimated by subtracting the calculated hydration effect from the experimental unfolding enthalpy[14,16] or directly from molecular mechanics calculations.[9]

While the cycle in Figure 1 is theoretically correct, its evaluation raises tremendous challenges. In particular, attempts to predict unfolding enthalpies from structures using this approach are not expected to be accurate enough for stability prediction or structure-based molecular design. The main computational difficulty arises from the large difference in magnitude between the experimental unfolding enthalpy, and the calculated hydration contribution and the contribution due to the disruption of internal nonbonded interactions. These two terms are of opposite sign but about one order of magnitude larger than the experimental enthalpies. For hen egg-white lysozyme (HEWL), for example, the experimental unfolding enthalpy at 60°C is 110 kcal/mol, while the calculated hydration term according to Makhatadze and Privalov[14] is close to −2000 kcal/mol. The immediate implication of this large disparity is that any error in the estimation of the hydration enthalpy or the enthalpy due to the disruption of nonbonded interactions will be amplified by a factor of 20 in the evaluation of the unfolding enthalpy. Since a difference of ±2% in the enthalpy change is reflected in a difference of ±6°C in the predicted denaturation temperature, it is clear that successful molecular design strategies would require errors lower than 2% in the enthalpy prediction. The situation is even worse at lower temperatures in which the unfolding enthalpy is smaller in magnitude. In order to achieve a 2% error or less in the unfolding enthalpy, the hydration enthalpy would have to be predicted with an error smaller than 0.1%. The second main difficulty is that the transfer energetics of a protein from the gas phase to water cannot be measured directly, being therefore approximated from the transfer thermodynamics of small organic molecules using some additivity scheme. This is certainly an extremely difficult, if not impossible, task, considering the experimental error in the determination of the hydration enthalpies of model compounds and the fact that the model compounds only approximate the situation existing in proteins. Makhatadze and Privalov[14] estimate their error in hydration enthalpy as ±5%, which translates into an error of ±100 kcal/mol for HEWL, that is, ±100% error in the unfolding enthalpy at 60°C. However, it should be noted that this is only the error in reproducing the experimental values for small molecules and does not include any uncertainties arising from the extrapolation to proteins. In this respect, the validity of simple additivity schemes has recently been questioned, especially for polar interactions.[9] It is evident that approaches similar to those described above are of limited util-

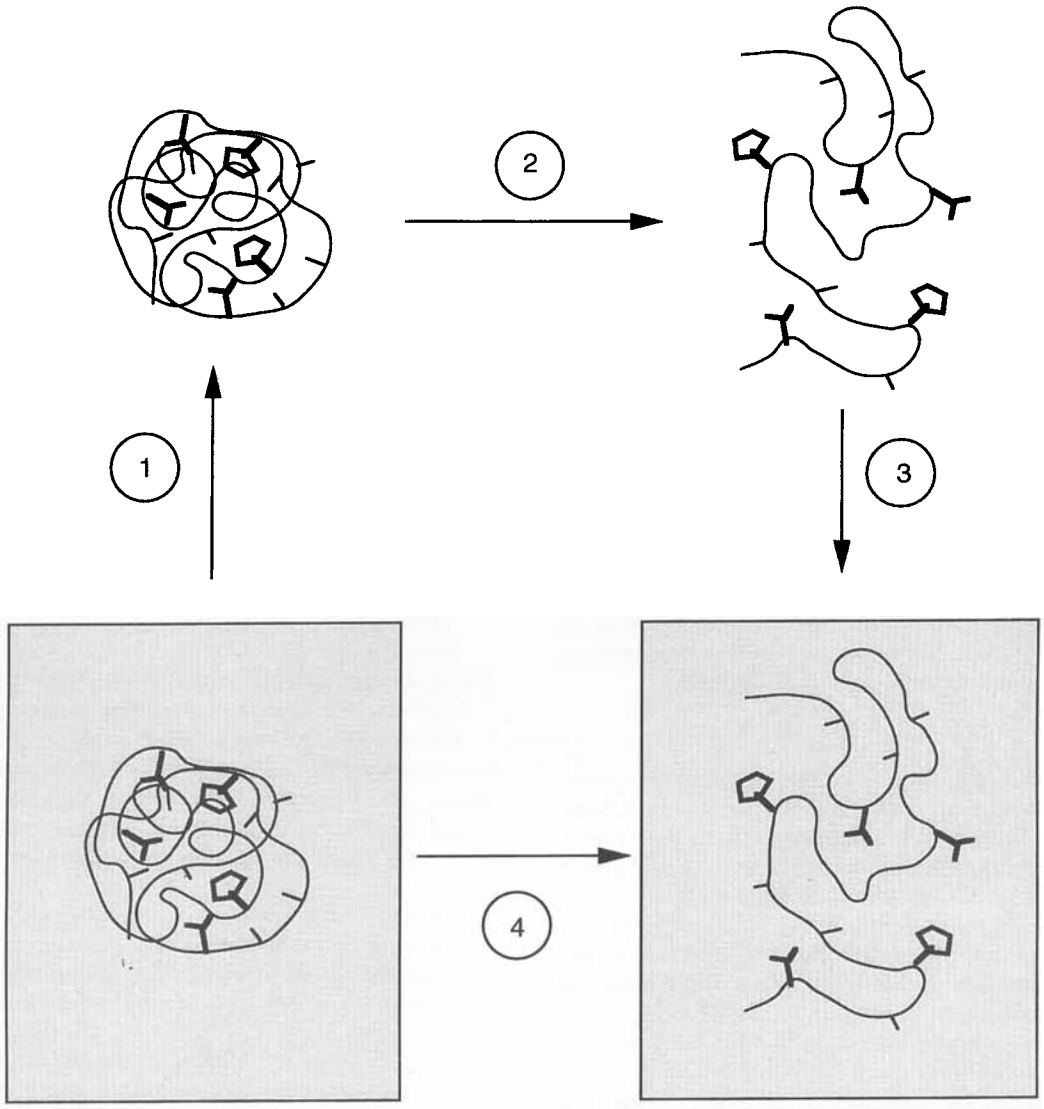

Fig. 1. Classical thermodynamic cycle for the folding/unfolding reaction of proteins. In the figure, step 4 represents the unfolding reaction in water. Steps 1 and 3 represent the hydration/dehydration of the native and unfolded states, and step 2 represents the unfolding reaction in vacuum. Step 4 is the only one that is measured experimentally. Steps 1 and 3 have been traditionally estimated from thermodynamic data for the hydration of small organic molecules.

ity for structure-based enthalpy prediction and that a different approach is warranted.

Structural Parametrization and Group Additivity

There are two general strategies aimed at parametrizing the energetics of protein folding/unfolding in terms of structure. One strategy utilizes data obtained from model compounds normalized in some appropriate way (e.g., ASA) in order to obtain elementary parameters for a given thermodynamic function, and then calculates the expected value for the entire protein by assuming group additivity. The other strategy utilizes the protein data and obtains elementary parameters by relating structural and thermodynamic parameters. The elementary parameters obtained both ways need not be equal. In fact, it can be easily proved that they are only equal if the assumption of additivity holds. Consider the thermodynamic cycle in Figure 1, for an arbitrary thermodynamic quantity ΔX, which describes step 4 (e.g., experimental ΔG, ΔH, ΔS, ΔC_p of protein folding/unfolding) we can write in a general way:

$$\Delta X = \Delta X_{hyd,U} - \Delta X_{hyd,N} + \Delta X_{vac} \quad (1a)$$

$$\Delta X = \Delta\Delta X_{hyd} + \Delta X_{vac} \quad (1b)$$

where $\Delta X_{hyd,U}$ and $\Delta X_{hyd,N}$ are the hydration values for the unfolded and native states, and ΔX_{vac} the difference between the unfolded and native states in vacuum. The differential hydration term between the unfolded and native states of the protein,

$\Delta\Delta X_{hyd}$, can be written as the sum of two terms, $\Delta\Delta X^{add}{}_{hyd}$, the term obtained by adding the contributions of isolated individual groups (i.e., the additive term) and, $\Delta\Delta X^{\#}{}_{hyd}$, the term that accounts for any deviations from additivity (i.e., the nonadditive term). Accordingly,

$$\Delta\Delta X_{hyd} = \Delta\Delta X^{add}{}_{hyd} + \Delta\Delta X^{\#}{}_{hyd}. \quad (2)$$

The additive term is written in terms of elementary parameters obtained from small molecules. For example, if the parametrization is made in terms of ΔASA,

$$\Delta\Delta X_{hyd} = \Sigma\, \alpha'_j \cdot \Delta ASA_j + \Delta\Delta X^{\#}{}_{hyd} \quad (3)$$

where the summation runs over all types of chemical groups and α'_j is the contribution per unit area for group j obtained from the data for small molecules. If the assumption of additivity is correct then $\Delta\Delta X^{\#}{}_{hyd} = 0$ and the approach predicts well the protein value.

The second strategy utilizes the protein data to derive the elementary parametrization parameters. This approach is based on the equation

$$\Delta\Delta X_{hyd} = \Sigma\, \alpha_j \cdot \Delta ASA_j \quad (4)$$

Since the coefficients α_j in the above equation are evaluated from an analysis of the protein data, they become effective hydration parameters that represent the average contribution per unit area of group j within the context of the protein molecule. *This approach does not assume group additivity.* Rather, it assumes that the protein environment is homogeneous enough such that the average value obtained holds for different proteins. If, on the other hand, the parameter α is evaluated from data for small organic molecules (i.e., α'), the assumptions of group additivity as well as homogeneity are present. In this case, the hydration energetics for the entire protein is estimated from the value obtained for a small molecule, normalized on a per unit area basis, and multiplying it by the total area change in the protein molecule.

If the parameter α obtained from the analysis of protein data and that obtained from small organic molecules are similar, it can be said that the assumption of additivity is correct or approximately correct. If, on the other hand, the values are different the assumption of additivity is not valid. In this case, the data from small molecules cannot be scaled up to estimate the protein data or vice versa.

Packing Density and Enthalpy Change

Previously, it has been shown from a purely phenomenological point of view that the enthalpy and heat capacity changes of protein unfolding scales rather well in terms of the changes in solvent accessible polar (ΔASA_{pol}) and apolar (ΔASA_{ap}) surface areas.[10,11,17] In fact, after correction for protonation effects, the enthalpy change at the median unfolding temperature of 60°C can be predicted within ±6% in terms of ASA changes alone using the following equation[11,12]:

$$\Delta H(60) = a \cdot \Delta ASA_{ap} + b \cdot \Delta ASA_{pol} \quad (5)$$

where the coefficients a and b equal −8.44 and 31.4 cal/mol-Å^2.

The result summarized in Eq. (5) is surprising at first: Why such a simple relationship is able to predict protein unfolding enthalpies better than other more sophisticated approaches? The reason is twofold. First, Eq. (5) does not calculate separately the hydration term and the term due to the disruption of noncovalent interactions. Rather, Eq. (5) is the result of an empirical statistical analysis of the experimental protein unfolding enthalpy in water and, therefore, it does not involve the subtraction of very large numbers. Second, only two parameters (a and b) are enough to account very precisely for the experimental data. It has been argued[18–20] that the reason why proteins show this behavior is that, on average, they are structurally very similar: They exhibit similar atomic compositions, similar pairwise interactions, and similar packing densities. Consequently the enthalpic contribution due to the disruption of internal interactions becomes proportional to changes in accessible surface areas only, and the parameters a and b are similar for all proteins.

The situation discussed above can be evaluated by considering the distribution of separation distances between the different atom pairs found in proteins. For this analysis we considered the following types of atom pairs: C—C, C—O, CO—C, CO—N, CO—O, O—N, O—O, and CS where CO stands for the carbonyl carbon. These types of atom pairs comprise more than 99% of the total attractive pairwise interactions within a protein. Since the energy is given by a standard 6–12 Lennard-Jones potential of the form:

$$\Delta H_{LJ} = -\Sigma(A/R_{i,j}{}^{12} + B/R_{i,j}{}^{6}) \quad (6)$$

the important quantity to consider is not the separation distance per se, but the average inverse separation distance or, more importantly, the average $1/R^6$, since the attractive sixth power term is the one that contributes the most to the enthalpy change of unfolding, ΔH_{LJ}. Table I summarizes the results obtained for fifteen proteins. In these calculations the average separation distance between different atom pairs was calculated using different procedures. In one set of calculations, the average inverse separations, $<1/R>$, were estimated using standard cutoff distances of 7Å and 8Å.[21] In another set of calculations, average $<1/R^6>$ values were estimated using the same cutoff distances. An alternative averaging in which the contribution of each atom pair was

TABLE I. Average Separation Distance Between Different Atom Types for 15 Proteins*

Atom pair	$\langle R_7 \rangle^a$ Å	$\langle R_8 \rangle^a$ Å	$\langle R_7 \rangle^b$ Å	$\langle R_8 \rangle^b$ Å	$\langle R_w \rangle^c$ Å	$R_{min}{}^d$ Å	U^e
C—C	6.138 ± 0.007	6.690 ± 0.025	5.653 ± 0.059	5.968 ± 0.046	5.543 ± 0.07	3.96	0.714
C—O	5.796 ± 0.004	6.471 ± 0.03	5.065 ± 0.025	5.396 ± 0.031	5.071 ± 0.08	3.67	0.724
CO—C	5.848 ± 0.007	6.449 ± 0.061	5.259 ± 0.018	5.563 ± 0.032	5.583 ± 0.13	4.04	0.724
CO—N	5.201 ± 0.022	5.789 ± 0.165	4.057 ± 0.047	4.266 ± 0.069	4.392 ± 0.16	2.95	0.672
CO—O	5.45 ± 0.018	6.087 ± 0.09	4.785 ± 0.064	5.063 ± 0.069	4.677 ± 0.08	3.64	0.778
O—N	5.241 ± 0.016	5.895 ± 0.094	4.036 ± 0.045	4.272 ± 0.054	3.923 ± 0.10	2.91	0.742
O—O	5.798 ± 0.007	6.378 ± 0.088	5.359 ± 0.026	5.651 ± 0.055	5.892 ± 0.08	4.45	0.755
C—S	6.125 ± 0.04	6.727 ± 0.247	5.618 ± 0.393	5.959 ± 0.423	5.461 ± 0.3	3.96	0.725

*The proteins included in the calculations are: Hen egg-white lysozyme, barnase, α-lactalbumin, BPTI, staphylococcal nuclease, ribonuclease A, cytochrome c, carbonic anhydrase, chymotrypsin, papain, pepsinogen, trypsin, ribonuclease T1, GCN4, and protein G third domain.

[a]The average separation distance calculated with the formula $\langle R \rangle = \langle 1/R \rangle^{-1}$. $\langle R_7 \rangle$ and $\langle R_8 \rangle$ were calculated using a cutoff distance of 7 and 8 Å, respectively.

[b]The average separation distance calculated with the formula $\langle R \rangle = (\langle 1/R^6 \rangle)^{-1/6}$. $\langle R_7 \rangle$ and $\langle R_8 \rangle$ were calculated using a cutoff distance of 7 and 8 Å, respectively.

[c]the energy-weighted average separation distance $\langle R_w \rangle = \Sigma R_{i,j} \cdot (E_{i,j}/E_{Tot})$ calculated with the Lennard-Jones parameters of Levitt [1974].

[d]R_{min} values according to Levitt [1974].

[e]The U parameter defined as $(R_{min}/\langle R \rangle)$ calculated with the energy-weighted average separation distance $\langle R_w \rangle$.

weighted according to its energy was also considered. The advantage of this procedure is that it does not require an arbitrary cutoff distance. In this case, for each type of atom pair, x–y, the energy weighted distance average, $\langle R_{x-y,w} \rangle$ is defined as:

$$\langle R_{x-y,w} \rangle = \Sigma R_{i,j} \cdot (E_{i,j}/E_{Tot}) \quad (7)$$

where $R_{i,j}$ is the separation distance between atoms i and j, $E_{i,j}$ the energy of that pair, and E_{Tot} the total energy for all the x–y atom pairs in the protein. In all calculations, only those pairs of atoms not joined by a covalent bond and not belonging to the same amino acid were considered. As seen in Table I, it is evident that, independent of the method used to evaluate the distribution of separation distances, the resulting values show little variation between protein molecules. Figure 2 summarizes the $\langle R_{x-y,w} \rangle$ values obtained for all proteins studied and clearly illustrates that the weighted average separation distances between different atom pairs are clustered together with a small dispersion around their mean values.

For each type of atom pair, the small observed variation in the average separation distance will be reflected in a correspondingly small variation in the average van der Waals energy. Accordingly, the energy associated with the disruption of internal interactions will be essentially the same per unit area of a given type, and becomes a function of accessible surface areas only. As far as the distribution of separation distances is similar among proteins, the scaling coefficients a and b are also similar (i.e., the enthalpies associated with the disruption and hydration of one $Å^2$ of apolar or polar surfaces are essentially the same for all proteins). Deviations from this behavior should be expected in those cases in which the distribution of separation distances differ from the average, that is, for those situations in which the packing density deviates from that of the average native state. Also, deviations should be expected if the compositional distribution is different than the average (e.g., the proportion of polar–polar, apolar–apolar and polar–apolar pairs is different) and for residues at the protein surface. In fact, it has been shown that the a and b coefficients that quantitatively account for the unfolding enthalpy of the native state, overestimate the enthalpy change associated with the unfolding of protein conformations characterized by a more relaxed packing density.[18] Also, in a comparative study of c type lysozymes,[20] it was concluded that the observed enthalpic variations between lysozyme species could be attributed to differences in packing density. It was pointed out before, that the distribution of both the packing density and the location of polar and apolar groups is not homogeneous and that these variations might give rise to situations in which partly folded states are stabilized or situations in which the cooperative folding behavior is disrupted.[11,12,18]

Nonbonded Interactions

According to the thermodynamic cycle in Figure 1, the enthalpy change for protein unfolding in aqueous solution can be represented as the sum of two terms:

$$\Delta H = \Delta H_{vac} + \Delta\Delta H_{hyd}. \quad (8)$$

The enthalpy change due to the disruption of internal interactions (ΔH_{vac}) is of the same order of magnitude but of opposite sign from the hydration term. This term includes the disruption of generic interactions common to all proteins like van der

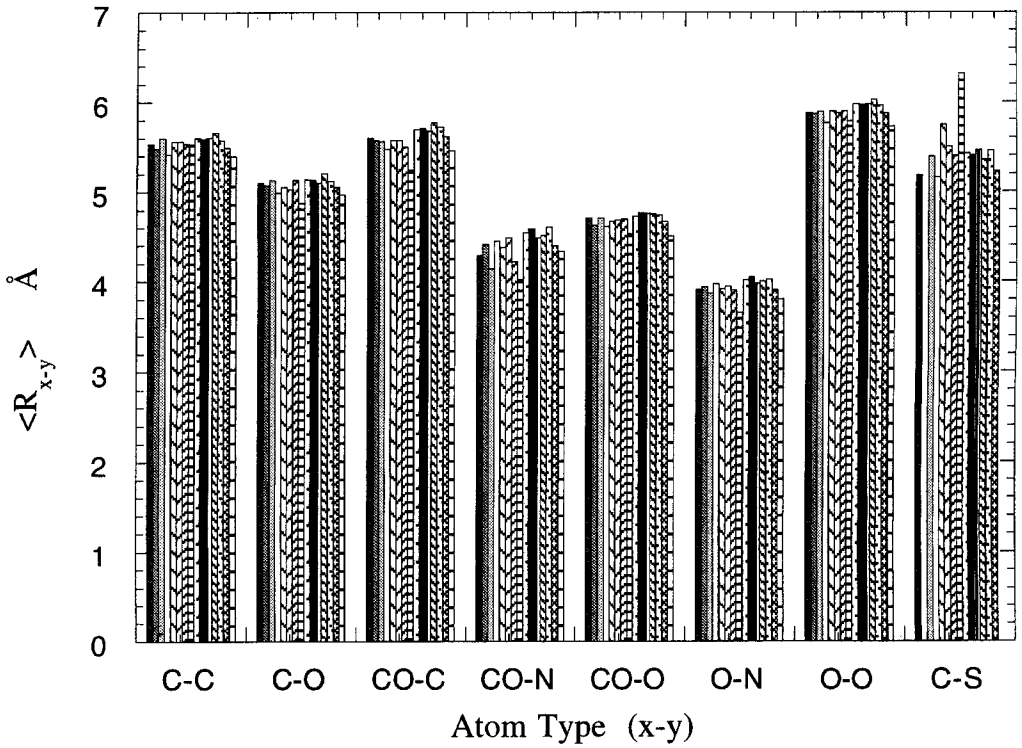

Fig. 2. The energy-weighted average separation for different atom pairs for 15 proteins. This figure illustrates that for the different types of pairwise interactions, the separation distances between atom pairs are clustered together with a small dispersion around their mean values. From left to right the bars correspond to hen egg-white lysozyme (1LZT), barnase (1RNB), α-lactalbumin (1ALC), BPT1 (5PTI), staphylococcal nuclease (1SNC), ribonuclease A (7RSA), cytochrome c,[30] GCN4 (2ZTA), carbonic anhydrase (2CAB), chymotrypsin (4CHA), papain (9PAP), pepsinogen (2PSG), trypsin (1TLD), ribonuclease T1, (Martinez-Oyanedel et al. (1991)), and protein G third domain (1IGD).

Waals interactions, hydrogen bonds, electrostatic interactions. As stated earlier, other terms like the protonation of specific groups or the presence of cofactors or metal ions must be considered on an individual basis, since they require case specific parameters. Here, we will be concerned with generic van der Waals interactions described by a Lennard-Jones potential. The enthalpy change associated with the disruption of these interactions can be written as

$$\Delta H_{LJ} = -\Sigma(A' \cdot (R_{min}/R_{i,j})^{12} + B' \cdot (R_{min}/R_{i,j})^{6}) \quad (9)$$

where the coefficients A′, B′, and Rmin are characteristic for each type of pairwise interaction. Since the ratio $(R_{min}/R_{i,j})$ is in the range between 0.7 and 0.8 for all interactions (Table I), the 12th power term becomes negligible as shown in Figure 3, and Eq. (9) is well approximated by only the 6th power term:

$$\Delta H_{LJ} = -\Sigma B' \cdot (R_{min}/R_{i,j})^{6} . \quad (10)$$

As shown in Figure 3, Eq. (10) is accurate for $(R_{min}/R_{i,j})$ values smaller than 0.85. The use of Eq. (10) is further justified because the starting point in our calculations is a high-resolution structure. The goal here is not to refine a structure, for which a repulsive term is necessary, but to estimate enthalpy changes. Statistically, the number of contacts for which $(R_{min}/R_{i,j})$ is larger than 0.85 is less than 2% of the number of contacts within a cutoff distance of 12 Å.

Equation 10 can be simplified further by grouping together interactions between similar types of atoms and using the average $U \equiv (R_{min}/R)$ for each type of interaction:

$$\Delta H_{LJ} \cong \Sigma B_i' \cdot N_i \cdot (R_{min}/R)_i^{6} \quad (11a)$$

$$\cong \Sigma B_i' \cdot N_i \cdot U_i^{6} \quad (11b)$$

where the summation runs over the different types of pairwise atomic interactions existing in the protein (see Table I). In our calculations we have used U values defined in terms of the weighted average separation distance and the R_{min} values calculated with the parameters published by Levitt.[22] Other definitions give similar results even though the numeric values of the resulting fitting parameters will be different.

The Enthalpy of Hydration

The enthalpy of hydration associated with protein unfolding is parametrized in terms of the surface area that becomes exposed to the sol-

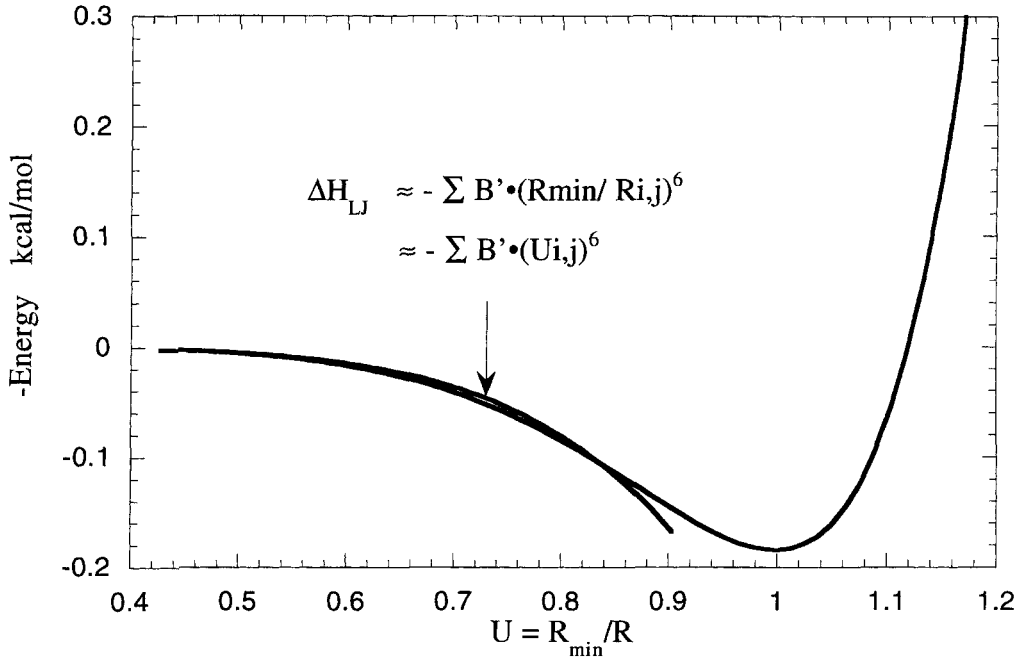

Fig. 3. The Lennard-Jones potential for the C—C interaction according to Levitt[22] parameters as a function of R_{min}/R. This figure illustrates that for R_{min}/R values smaller than 0.85 the potential is well-approximated by only a sixth power term.

vent.[10,14,17,23–25] According to Eq. (4), the hydration enthalpy can be written as

$$\Delta\Delta H_{hyd} = \Sigma\, \alpha_j \cdot \Delta ASA_j \qquad (12)$$

where the summation runs over atom classes that exhibit different hydration behavior. For example, a commonly used classification is to separate the protein atoms into polar and apolar groups,[10–12,26] or to further subdivide the apolar groups into aromatic and aliphatic.[14] The parameters α in Eq. (12) are evaluated from protein data, and as such they represent average elementary values that may or may not be equal to those obtained from small molecules.

A Parametric Equation for the Enthalpy of Protein Unfolding

According to the discussion above, after correction for protonation or other specific effects, the enthalpy change for protein unfolding can be estimated by adding Eq. (11) and (12):

$$\Delta H = \Sigma\, B_i' \cdot N_i \cdot U_i^6 + \Sigma\, \alpha_i \cdot \Delta ASA_i \qquad (13)$$

where the summation runs over the different types of pairwise interactions existing in the protein. In Eq. (13), however, the terms representing the contributions of hydration and those representing the disruption of nonbonded interactions are still separated, making difficult a joint optimization. In some cases, however, the number of internal pairwise interactions that are disrupted on unfolding are proportional to the change in accessible surface area of the groups involved in the interactions. This is shown in Figure 4 for apolar, polar, and mixed (polar–apolar) interactions. It is clear in this figure that, from purely empirical considerations, the number of van der Waals pairs is directly proportional to ΔASA. If this is the case, the contribution of those interactions to the total enthalpy can be parametrized in terms of ΔASA_i and equation 13 reduces to

$$\begin{aligned}\Delta H = {} & (\alpha_{ap} + \beta_{ap} \cdot U_{ap}^6) \cdot \Delta ASA_{ap} \\ & + (\alpha_{pol} + \beta_{pol} \cdot U_{pol}^6) \cdot \Delta ASA_{pol} \\ & + \beta_{mix} \cdot U_{mix}^6 \cdot \Delta ASA_{Total} \end{aligned} \qquad (14)$$

where the coefficients α_i and β_i are effective parameters that can be obtained from an analysis of the protein thermodynamic database. The term corresponding to mixed interactions does not include an α parameter because that contribution is already taken into account in the polar and apolar terms. An immediate application of Eq. (14) is that it allows to explicitly account for the effect of atomic packing in the calculation of the enthalpy change. A nonlinear least squares fit of the protein data set yields $\alpha_{ap} = -12.96$, $\beta_{ap} = 25.34$, $\alpha_{pol} = 24.38$, $\beta_{pol} = 16.57$, and $\beta_{mix} = 16.42$. The accuracy of the fit is shown in Figure 5. Even though Eq. (14) provides a statistically significant improvement over Eq. (5) (3% versus 6% error), its major advantage is that it allows consideration of a variety of situations, such as solvent-exposed helices, molten globules, and other partly folded states in which the packing density deviates from the average found in the native state of proteins. Equations (11) to (14) indicate that the enthalpy change is proportional to the packing den-

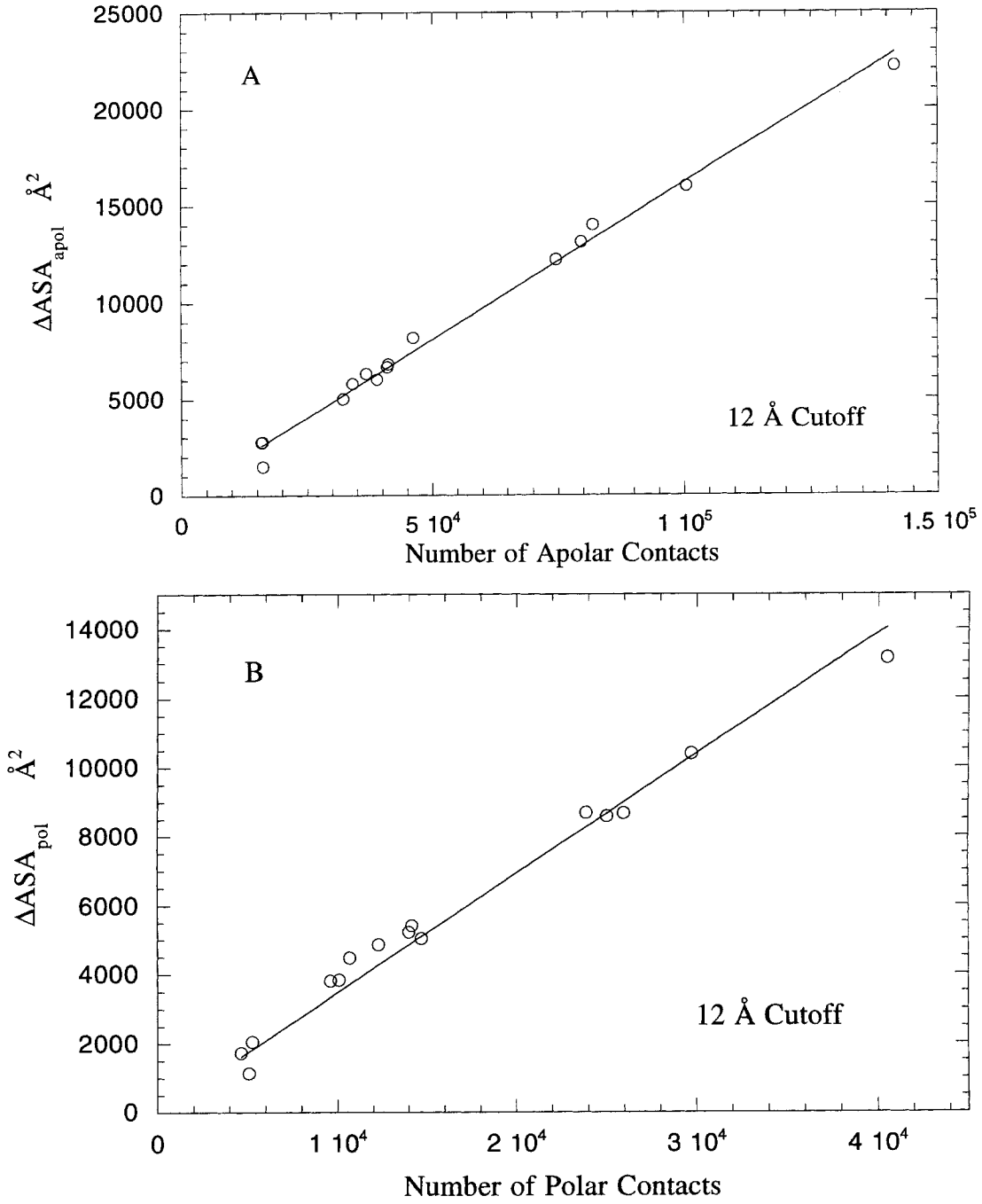

Fig. 4. The relationship between the change in accessible surface area and number of atomic contacts for apolar (**A**), polar (**B**) and mixed (**C**) interactions. The observed linearity permits simplification of the equations describing the enthalpy change. Shown in the graphs are the results obtained using a cutoff distance of 12 Å. Similar linearity is obtained with smaller cutoff distances.

sity through the dimensionless packing parameter U. If U = 1, ΔH is maximal and, conversely, if U = 0, ΔH is minimal and essentially equal to the hydration enthalpy. It must be noted also that if the average U parameters found in proteins are substituted into Eq. (14), it immediately reduces to Eq. (5). Thus, Eq. (14) accounts for the small differences in atomic separation distances found between proteins.

The Meaning of the α and β Coefficients

The coefficients α_i and β_i are parametric terms that account for different contributions averaged over the ensemble of proteins in the database. Only in the case in which the contribution of a specific interaction exhibits additivity are these parameters similar to those obtained from small molecules. For

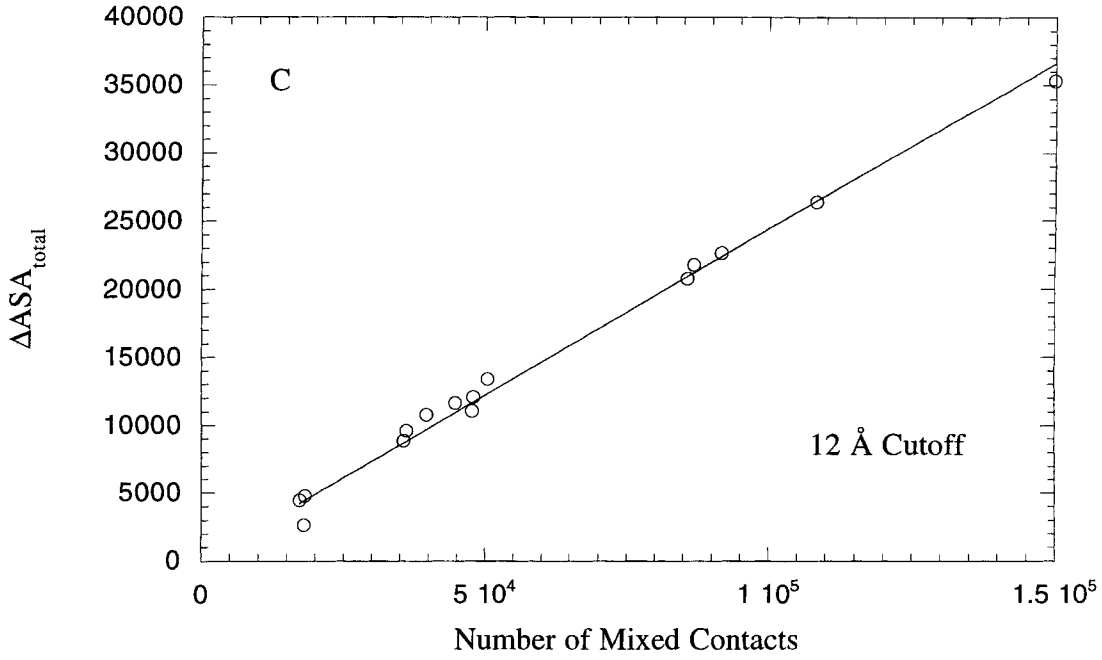

Fig. 4 (continued).

example, if additivity holds, the coefficient α_{ap} must be close to the enthalpy of transfering one Å^2 of apolar area from the gas phase to water at 60°C. In fact, inspection of the data tabulated by Makhatadze and Privalov[14] for the hydration enthalpies of simple organic compounds reveals a value of −12.88 cal/mol-Å^2 of aliphatic surface, extremely close to the one obtained from the statistical analysis of the protein database. This result indicates that the assumption of additivity is probably correct for apolar hydration. The coefficient β_{ap}, on the other hand, must be close to the enthalpy of sublimation of a closely packed apolar organic molecule (U = 1) at 60°C. For comparison, the enthalpy of sublimation of benzene (with a density of ρ = 0.835 cm^3/g) is 40.91 cal/mol-Å^2 at 60°C and the enthalpy of sublimation of cyclohexane (with a density of ρ = 0.741 cm^3/g) is 28.95 cal/mol-Å^2 at 60°C. The equality should not hold perfectly in this case because apolar groups in proteins are not only packed against apolar groups but also against polar groups. If the coefficient β_{mix} is also included, the total apolar term becomes 41.8 cal/mol-Å^2, which is close to the value obtained for model compounds. Thus, these results suggest that group additivity is probably a good approximation for the thermodynamic parameters of apolar groups.

The situation is different for polar interactions. In this case, the values of the coefficients α_{pol} and β_{pol} are not similar to coefficients obtained from model compounds, suggesting that polar interactions do not obey the additivity assumption. The overall polar term is a composite term including different contributions such as hydrogen bonding, electrostatic interactions, and polar van der Waals interactions. As such, the coefficients α_{pol} and β_{pol} cannot be related directly to a specific interaction or a specific process. The coefficients α_{pol} and β_{pol} should be treated as phenomenological fitting parameters only.

According to the thermodynamic cycle in Figure 1, the polar enthalpy change, ΔH_{pol}, can be written as

$$\Delta H_{pol} = \Delta\Delta H^{add}{}_{pol,hyd} + \Delta\Delta H^{\#}{}_{pol,hyd} + \Delta H_{pol,vac} \quad (15)$$

While the term $\Delta\Delta H^{add}{}_{pol,hyd}$, which is the one obtained from small molecules, contains only the differential hydration enthalpy of isolated groups between the unfolded and native states, the nonadditive term, $\Delta\Delta H^{\#}{}_{pol,hyd}$, includes not only all the interaction terms but also the enthalpy difference in internal interactions associated with the hydration of the native and unfolded states. For example, the change in electrostatic interactions associated with the transfer of the protein from a medium of dielectric constant of one (vacuum) to a medium of dielectric constant of 78 (water). The nonadditive term for polar interactions is significant for proteins as demonstrated recently by Lazaridis et al.[9] The analysis of the energetics of transfer (ΔG, ΔH, ΔS) from the gas phase to aqueous solution of small organic compounds in terms of group contributions leads to the same conclusion: while the thermodynamic properties associated with the transfer of molecules containing at most one functional polar group (e.g., monoalcohols, monoamines, etc.) can be accounted for in terms of the addition of individual contributions, large deviations are observed when the same approach is used to calculate the thermodynamic

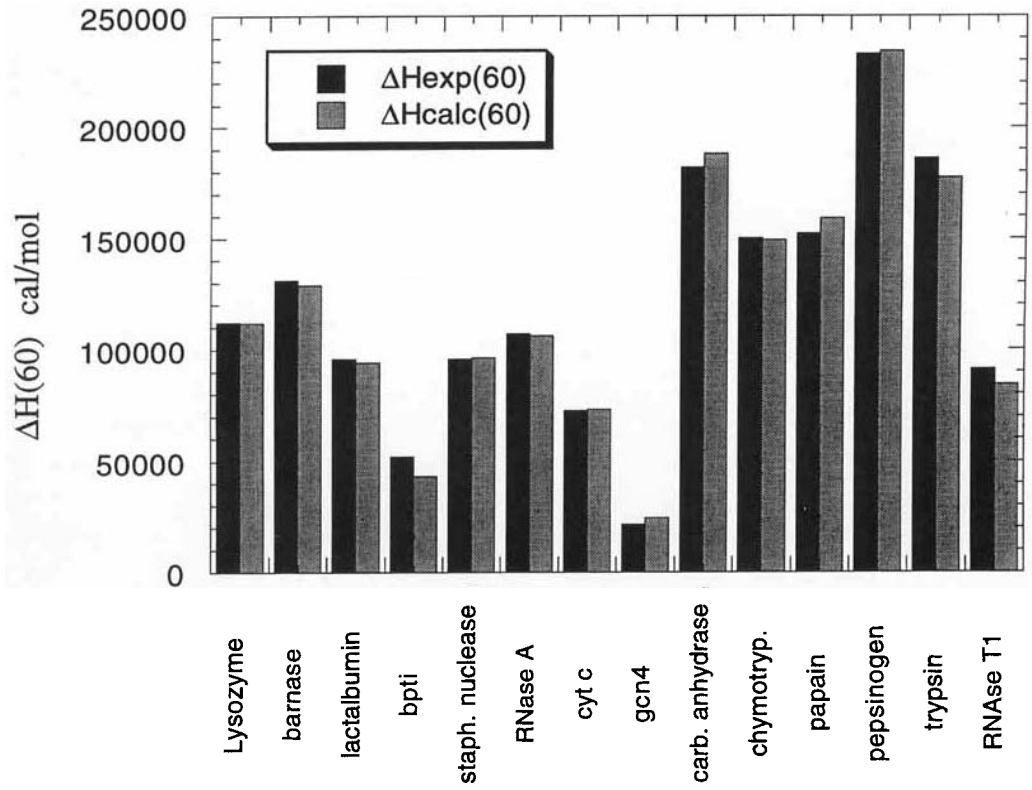

Fig. 5. The experimental and calculated enthalpy changes at 60°C. As shown in the figure, the parametrization developed in this paper accurately accounts for the experimental enthalpy with a standard deviation of ±4 kcal/mol and a standard error of 1 kcal/mol. The experimental data are from the following references: hen egg-white lysozyme,[31] barnase,[32] lactalbumin,[33] staphylococcal nuclease,[18] RNase A,[34] cytochrome c,[35] gcn4,[25] carbonic anhydrase,[36] chymotrypsin,[37] papain,[38] pepsinogen (unpublished results from this laboratory), trypsin,[39] RNase T1.[40] The calculated value for barnase includes the contribution from buried water molecules. In all calculations, ASA values were evaluated using the Lee and Richards algorithm as described.[10] ASA values for the unfolded state were optimized for each amino acid as described in ref. 41.

functions of polyfunctional polar compounds.[27] In particular, the enthalpy and entropy changes of polar hydration are smaller in magnitude than those expected from simple group additivity.[27] In other words, the use of data obtained for model compounds assuming additivity will overestimate the enthalpy and entropy of polar hydration in proteins.

The Effect of Buried Water

In many cases water molecules are found in the protein interior or in protein–protein complexes. These water molecules contribute to the atomic packing and as such contribute to the folding/unfolding enthalpy or the enthalpy of complex formation. We have found that for those cases in which water molecules have been clearly identified in the high-resolution structure, it is necessary to consider them explicitly. Perhaps, the situation is more apparent in protein–protein associations in which the enthalpy changes are not as large as in protein folding and the presence of water molecules make a significant difference. An important example for which both, high–resolution structural and thermodynamic data are available is the complex formation between lysozyme and specific antibodies. Consider, for example, the association between lysozyme and the D44.1 antibody for which it is known that 20 water molecules are buried at the binding interface with a solvent accessibility smaller than 50%.[28] If the enthalpy change is calculated without taking into consideration the water molecules, the predicted value is −0.6 kcal/mol at 25°C compared to the experimental value of −10.7 kcal/mol at the same temperature. The situation is similar for D1.3. This analysis indicates that, on the average, each buried water molecule contributes around −0.5 kcal/mol to the binding enthalpy of the lysozyme antibody complexes. Similar contributions are calculated if the thermodynamic and structural data for the complex between barnase and barstar are used in the analysis.[29]

CONCLUSIONS

The results presented here rationalize the observation that the enthalpy change of protein folding/unfolding scales in terms of changes in accessible surface area, and provide a way to incorporate explicitly the effects of packing density in the struc-

ture based prediction of enthalpy changes. Three types of interactions are considered in the parametrization: apolar, polar, and mixed interactions. A first set of parameters have been derived with the data at hand. While this parametrization predicts protein unfolding enthalpies with an error close to ±3%, it is expected that in the future the value of the parameters will be refined as more data, covering situations in which specific interactions can be better defined, becomes available.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (RR04328) and the National Science Foundation (MCB-9118687).

REFERENCES

1. Novotny, J., Bruccoleri, R.E., Saul, F.A. On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. Biochemistry 28:4735–4749, 1989.
2. Horton, N., Mitchell, L. Calculation of the free energy of association for protein complexes. Protein Sci. 1:169–181, 1992.
3. Miyamoto, S., Kollman, P.A. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics free energy perturbation approaches. Proteins 16:226–245, 1993.
4. Kuntz, D.I., Meng, E.C., Shoichet, B.K. Structure-based molecular design. Accounts Chem. Res. 27:117–123, 1994.
5. Eisenberg, D., McLachlan, A.D. Solvation energy in protein folding and binding. Nature 319:199–203, 1986.
6. Sharp, K.A., Nicholls, A., Fine, R.F., Honig, B. Extracting hydrophobic free energies from experimental data: Relationship to protein folding and theoretical models. Biochemistry 30:9686–9697, 1991.
7. Sharp, K.A. The hydrophobic effect. Curr. Opin. Struct. Biol. 1:171–174, 1991.
8. Lee, K.H., Xie, D., Freire, E., Amzel, L.M. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. Proteins Struct. Func. Genet. 20:68–84, 1994.
9. Lazaridis, T., Archontis, G., Karplus, M. Enthalpic contributions to protein stability: atom-based calculations and statistical mechanics. Adv. Protein Chem. 47:231–306, 1995.
10. Murphy, K.P., Bhakuni, V., Xie, D., Freire, E. The molecular basis of co-operativity in protein folding III: Identification of cooperative folding units and folding intermediates. J. Mol. Biol. 227:293–306, 1992.
11. Xie, D., Freire, E. Structure based prediction of protein folding intermediates. J. Mol. Biol. 242:62–80, 1994.
12. Xie, D., Freire, E. Molecular basis of cooperativity in protein folding. V. thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins Struct. Func. Genet. 19:291–301, 1994.
13. Oobatake, M., Ooi, T. Hydration and heat stability effects on protein unfolding. Prog. Biophys. Mol. Biol. 59:237–284, 1992.
14. Makhatadze, G.I., Privalov, P.L. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. J. Mol. Biol. 232:639–659, 1993.
15. Makhatadze, G.I., Privalov, P.L. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. J. Mol. Biol. 232:660–679, 1993.
16. Makhatadze, G.I., Privalov, P.L. Energetics of protein structure. Adv. Protein Chem. 47:307–425, 1995.
17. Gomez, J., Freire, E. Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. J. Mol. Biol. 252:337–350, 1995.
18. Xie, D., Fox, R., Freire, E. Thermodynamic characterization of a staphylococcal nuclease folding intermediate. Protein Sci. 3:2175–2184, 1994.
19. Freire, E. Thermodynamics of partly folded intermediates in proteins. Ann. rev. of Biophys. Biomol. Struct. 24:141–165, 1995.
20. Griko, Y.V., Freire, E., Privalov, G., Van Dael, H., Privalov, P.L. The unfolding thermodynamics of c-type lysozymes: A calorimetric study of the heat denaturation of equine lysozyme. J. Mol. Biol. 252:447–459, 1995.
21. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. J. Computat. Chem. 4:187–217, 1983.
22. Levitt, M. Energy refinement of hen egg-white lysozyme. J. Mol. Biol. 82:393–420, 1974.
23. Murphy, K.P., Gill, S.J. Solid model compounds and the thermodynamics of protein unfolding. J. Mol. Biol. 222: 699–709, 1991.
24. Murphy, K.P. Hydration and convergence temperatures: On the use and interpretation of correlation plots. Biophys. Chem. 51:311–326, 1994.
25. DAquino, J.A., Gómez, J., Hilser, V.J., Lee, K.H., Amzel, L.M., Freire, E. The magnitude of the backbone conformational entropy change in protein folding. Proteins. In Press: 1995.
26. Spolar, R.S., Livingstone, J.R., Record, M.T. Jr. 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, 1992.
27. Cabani, S., Gianni, P., Mollica, V., Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. J. Solution Chem. 10:563–595, 1981.
28. Schwarz, F.P., Tello, D., Goldbaum, F.A., Mariuzza, R.A., Poljak, R.J. Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies. Eur. J. Biochem. 228:388–394, 1995.
29. Buckle, A.M., Schreiber, G., Fersht, A.R. Protein-protein recognition: Crystal structure of a barnase-barstar complex at 2.0 Å resolution. Biochemistry 33:8878–8889, 1994.
30. Bushnell, G.W., Louie, G.V., Brayer, G.D. High-resolution three-dimensional structure of horse heart cytochrome *c*. J. Mol. Biol. 214:585–595, 1990.
31. Khechinashvili, N.N., Privalov, P.L., Tiktopulo, E.I. Calorimetric investigation of lysozyme thermal denaturation. FEBS Lett. 30:57–60, 1973.
32. Martinez, J.C., Harrous, M.E., Filimonov, V.V., Mateo, P.I., and Fersht, A.R. A calorimetric study of the thermal stability of barnase and its interaction with 3'GMP. Biochemistry 33:3919–3926, 1994.
33. Griko, Y., Freire, E. and Privalov, P.L. Energetics of the alpha-lactalbumin states: A calorimetric and statistical thermodynamic study. Biochemistry 33:1889–1899, 1994.
34. Privalov, P.L., Tiktopulo, E.I., Khechinashvili, N.N. Calorimetric investigation of ribonuclease thermal denaturation. Int. J. Pept. Protein Res. 5:229–237, 1973.
35. Fu, L., Freire, E. On the origin of enthalpy and entropy convergence temperatures in protein folding. Proc. Natl. Acad. Sci. U.S.A. 89:9335–9338, 1992.
36. Tatunashvili, L.V., Privalov, P.L. Microcalorimetric study of heat denaturation of carbonic anhydrase B. Biofizika 31:578–581, 1986.
37. Tischenko, V.M., Tiktopulo, E.I., Privalov, P.L. Calorimetric investigation of the conformational conversions of alpha-chymotrypsin. Biofizika 19:400–404, 1974.
38. Tiktopulo, E.I., Privalov, P.L. Papain denaturation is not a two-state transition. FEBS Lett. 91:57–58, 1978.
39. Tischenko, V.M., Gorodnov, B.G. Calorimetric investigation of thermal denaturation of beta-trypsin. Biofizika. 24: 334–335, 1979.
40. Plaza del Pino, I.M., Pace, C.N., Freire, E. Temperature and guanidine hydrochloride dependence of the structural stability of ribonuclease T_1. Biochemistry 31:11196–11202, 1992.
41. Luque, I., Freire, E. Structure based thermodynamic scale of α-helix propensity. Submitted, 1996.
42. Martinez-Oyandel, J., Choe, H-W. Heinemann, U., and Saenger, W. Ribonuclease T_1 with free recognition and catalytic site: Crystal structure analysis at 1·5Å resolution. J. Mol. Biol. 222:335–352, 1991.