Protein Science (1997), 6:1057–1064. Cambridge University Press. Printed in the USA.

Consistency in structural energetics of protein folding and peptide recognition

CHAO ZHANG,[1] JAMES L. CORNETTE,[2] AND CHARLES DELISI[1]

[1]Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215
[2]Department of Mathematics, Iowa State University, Ames, Iowa 50011

(RECEIVED October 28, 1996; ACCEPTED February 17, 1997)

Abstract

We report a new free energy decomposition that includes structure-derived atomic contact energies for the desolvation component, and show that it applies equally well to the analysis of single-domain protein folding and to the binding of flexible peptides to proteins. Specifically, we selected the 17 single-domain proteins for which the three-dimensional structures and thermodynamic unfolding free energies are available. By calculating all terms except the backbone conformational entropy change and comparing the result to the experimentally measured free energy, we estimated that the mean entropy gain by the backbone chain upon unfolding (ΔS_{bb}) is 5.3 cal/K per mole of residue, and that the average backbone entropy for glycine is 6.7 cal/K. Both numbers are in close agreement with recent estimates made by entirely different methods, suggesting a promising degree of consistency between data obtained from disparate sources. In addition, a quantitative analysis of the folding free energy indicates that the unfavorable backbone entropy for each of the proteins is balanced predominantly by favorable backbone interactions. Finally, because the binding of flexible peptides to receptors is physically similar to folding, the free energy function should, in principle, be equally applicable to flexible docking. By combining atomic contact energies, electrostatics, and sequence-dependent backbone entropy, we calculated a priori the free energy changes associated with the binding of four different peptides to HLA-A2.1 MHC molecule and found agreement with experiment to within 10% without parameter adjustment.

Keywords: atomic contact energies; backbone conformational entropy; protein folding/unfolding transition; protein–peptide interaction; protein stability

One of the most fundamental quests in structural biology is to understand how unstructured polypeptides fold into unique native conformations (Levinthal, 1968; Anfinsen, 1973). Common cellular proteins have free energies of stabilization $\Delta_N^U G$ on the order of 10 kcal/mol (Privalov, 1979), attributable to the cumulative effect of attractive and repulsive interactions at many locations within a given molecule (Dill, 1990). One of the difficulties in calculating $\Delta_N^U G$ accurately is that the separate contributions of the various terms can be very large compared to the net stabilization free energy (Privalov & Gill, 1988; Dill, 1990; Honig & Yang 1995; Lazaridis et al., 1995), and consequently small percent errors in the energy balance can result in a large final error by subtraction.

Semi-empirical methods of varying degrees of applicability have been developed to calculate what is generally agreed to be the major force driving folding, viz the desolvation of nonpolar groups (Kauzmann, 1953; Tanford, 1968; Eisenberg & McLachlan, 1986; Privalov & Gill, 1988; Novotny et al., 1989). Additional stability is provided by favorable intramolecular interactions in the well-packed protein interior. These gains must be sufficient to offset a large unfavorable side-chain and backbone entropy change, and the unfavorable desolvation energy in removing charges or partial charges from solvent.

Direct experimental information on the magnitude of each of these various contributions to the energetics of protein structure is not yet available. However, some insight can be gained by combining computational and experimental data. Over the past several years, a wealth of experimental data has become available concerning the overall thermodynamic properties of proteins, such as heat capacity as a function of temperature, from which enthalpies and certain entropies can be obtained. By correlating the experimental data obtained from such studies with the structural features of proteins, it is possible to interpret the thermodynamic measurements in terms of individual contributions (Privalov, 1979; Baldwin, 1986; Privalov & Gill, 1988; Murphy & Freire, 1992; Oobatake & Ooi, 1992; Freire et al., 1993; Makhatadze & Privalov 1993; Privalov & Makhatadze, 1993). In particular, the magnitude of conformational entropy can be deduced by calculating each of the other terms that contribute to protein stability and subtracting them from the experimental entropy (Freire et al., 1993; Privalov & Makhatadze, 1993).

Reprint requests to: Charles DeLisi, Boston University, College of Engineering, 44 Cummington Street, Boston, Massachusetts 02215; e-mail: delisi@enga.bu.edu.

Such a semi-empirical approach has been used in a number of investigations, which differ from each other mainly in the way they model solvation. Privalov and his colleagues calculate solvation (or hydration) enthalpies and entropies from the accessible surface areas on the protein surface and the contributions of atom groups as derived from solution studies on low molecular weight model compounds (Makhatadze & Privalov, 1993; Privalov & Makhatadze, 1993). In contrast, Freire and colleagues derive surface expansion parameters for the enthalpic contributions due to polar and nonpolar groups from a direct analysis of a protein thermodynamic database, arguing that properties observed in small molecules may not be relevant to the conditions imposed by protein structures (Murphy & Freire, 1992; Freire et al., 1993; D'Aquino et al., 1996). The estimated conformational entropies range from 4 to 12 kcal/K·mol-res, depending on the value used for solvation (Privalov, 1979; Murphy & Freire, 1992; Freire et al., 1993; Privalov & Makhatadze, 1993; Spolar & Record, 1994; Makhatadze & Privalov, 1995; Yang & Honig, 1995; D'Aquino et al., 1996).

An alternative approach, which we present here, is to use a recently developed formulation of atomic contact energies (Zhang et al., 1997) to estimate desolvation effects and side-chain entropy changes, leaving only the backbone entropy change as the main unknown, which can be determined by subtracting ACE and direct electrostatic energies from the experimental unfolding free energy. In our case, the terms in the free energy function are taken from widely disparate sources: atomic contact energies on one hand, which are based on a statistical mechanical analysis of structural data; and calorimetry, on the other. A question of interest is whether terms obtained in this way are sufficiently consistent—as they should be barring unusual errors—to yield a reasonable value of backbone entropy. One of our results is that such an estimate is fully in accord with the most recent estimates of average backbone entropy change per residue, as well as with the average differential change relative to glycine.

Furthermore, an application of the backbone entropy estimate leads to a general statement concerning the main source of stability. In particular, we find that, in each of the 17 single-domain structures, backbone interactions provide 90% of the stability, with side-chain–side-chain interactions largely providing specificity.

Finally, because the binding of flexible peptides to receptors involves a disorder/order transition for the peptide fragment (D'Aquino et al., 1996), an event formally similar to a step in protein folding, the same free energy function, now including backbone entropy explicitly, should be applicable to that process (Rosenfeld et al., 1995). We find that the calculated free energy change attending binding of flexible peptides to the HLA-A2 is within 10% without adjustment of parameters.

Free energy decomposition

Atomic contact energies and desolvation

Atomic contact energies (ACE)[3] were estimated by a statistical analysis of atom pairing frequencies in a data set of inhomogeneous protein structures using a procedure formally analogous to the method of Miyazawa and Jernigan (1985). The atomic contact energy, e_{ij}, represents the desolvation energy of forming an atom–atom (i–j) pair from two atom–solvent pairs (i–0 and j–0). In the treatment of Zhang et al. (1997), 18 protein atom types were distinguished according to the chemical identities and their positions in amino acid side chains. The statistical averages of the observed numbers of pairwise contacts between atom types were related to contact energies using the so-called quasi-chemical approximation in which contact pair formation was analogous to a pairwise chemical reaction (Miyazawa & Jernigan, 1985; Zhang et al., 1997).

Contact energies have been applied to processes associated directly with solvation energy changes, such as the transfer of an amino acid from a protein interior to water, and site-specific mutations of hydrophobic amino acids. More generally, the ACE-based desolvation term can be used as part of a free energy function, which also includes pairwise electrostatic energies and the overall translational/rotational entropy loss, to predict the free energy change accompanying the formation of binary protein complexes. The explicit inclusion of direct electrostatics is necessary because few such interactions are present in the interior of a protein and they therefore are not accounted for by contact energies. We found that the calculated binding free energies of nine protease–inhibitor complexes were typically within 10% of the experimentally measured values. In a separate study (Vasmatzis et al., 1997), an ACE-based free energy function was incorporated into a conformation search procedure to predict the joint conformations of several interacting side chains. The results of these studies indicate that the determination of ACE provides an accurate and extremely rapid method for calculating desolvation energies.

Energetics of protein folding/unfolding transition

The unfolding of a small, single-domain protein is usually well approximated by a two-state transition between native and denatured states (Privalov, 1979). This model implies that the native conformation of such a protein breaks down cooperatively as a whole during the folding/unfolding transition. For a number of such proteins, the thermodynamics of unfolding has been characterized reliably by direct calorimetric measurements. This thermodynamic information can be combined with structural information to develop a fuller understanding of the contributions to protein stability.

Protein stability, a measure of the work required to disrupt the native structure, is defined as the Gibbs free energy difference between the native and unfolded state:

$$\Delta_N^U G = \Delta_N^U H - T\Delta_N^U S = \Delta_U G - \Delta_N G, \quad (1)$$

where N and U denote native and unfolded states. We can decompose the Gibbs energy change into three terms:

$$\Delta_N^U G = \Delta_N^U E_C + \Delta_N^U E_{El} - T\Delta S_{bb}^{prot}, \quad (2)$$

where $\Delta_N^U E_C$ and $\Delta_N^U E_{El}$ are the differences in atomic contact energies and electrostatic energies, respectively, between native and denatured states. Equation 2 assumes that any change in the rigid-body translational and rotational degrees of freedom can be neglected compared to other free energy changes accompanying the unimolecular transition. As indicated below, such is not the case for bimolecular reactions.

The combination of contact energies and a solvent-modified electrostatic energy provides a complete description of the energy

[3] Unless otherwise indicated, we will simply refer to these as contact energies, with the understanding that we are referring to atoms (Zhang et al., 1997) rather than amino acid residues (Miyazawa & Jernigan, 1985).

change of displacing water and forming noncovalent packing (Zhang et al., 1997). Specifically, in addition to the traditional definition of solvation effects, we expect that side-chain entropies are also contained in our contact energies because the reduction in the number of possible conformations of side chains must play a role in driving the partition of atoms between solvent and protein environment. Contact energies also contain a differential van der Waals energy (solute–solvent van der Waals replaced by solute–solute van der Waals), which generally favors the highly packed environment of a protein's interior. However, they do not contain contributions from main-chain conformational entropy, because the energies were estimated with polypeptide chain constraints explicitly removed, and our reference state has the same general configuration as the crystal structure. This entropy term is represented here by ΔS_{bb}^{prot}, reflecting the positive entropy change upon protein unfolding. Its value depends on the length and composition of the protein sequence. The average entropy change per residue, ΔS_{bb}, is considered less variable across proteins. We will estimate the ΔS_{bb}^{prot} by subtracting contact energies and electrostatic energies from the measured unfolding free energies.

To calculate $\Delta_N^U E_C$ and $\Delta_N^U E_{El}$, we need to known the corresponding values in both the native and unfolded state. For the native state, the total contact energies can be estimated by

$$\Delta_N E_C = \Sigma\Sigma e_{ij} n_{ij}, \tag{3}$$

where e_{ij} and n_{ij} are, respectively, the effective atomic contact energy and the total number of contacts between an atom of type i and an atom of type j (Zhang et al., 1997). The numbers of contacts were obtained from the solved crystal structures, which were taken from the Protein Data Bank (Bernstein et al., 1977).

We assume that, in the unfolded state, the residual noncovalent interactions between chemical groups are energetically insignificant, and model this state by placing the polypeptide in an extended conformation that corresponds to a β-stand. Such extended chains were constructed with default ϕ, ψ angles (−140°, 135°) using the CHARMM program (Brooks et al., 1983).

A number of analyses have been based on a polypeptide chain with neutral side chains (Makhatadze & Privalov, 1993; Privalov & Makhatadze, 1993; Lazaridis et al., 1995). The general idea is that, in the folded state, there are few fully charged side chains in the interior of proteins, and their interaction energies are largely canceled by the large free energy cost associated with removing charged groups from water, whereas in the unfolded state, their interactions with one another are assumed to be shielded by counterions. Consequently, they do not contribute significantly to the overall energy change. In keeping with these ideas, we used neutral side chains for all ionizable amino acids. Electrostatic interactions then come entirely from partial charges, whose locations and magnitudes are based on the polar hydrogen model of the CHARMM program (Brooks et al., 1983).

To avoid artifacts due to bad contacts, both native and extended structures were gently minimized by 300 steepest descent steps, as suggested by Lazaridis et al. (1995). Finally, following Makhatadze and Privalov (1995), all the prosthetic groups, ions, and cofactors were assumed to dissociate from the protein so energies were not included for them.

Flexible peptide binding free energies

For the association of two folded proteins where no conformational change occurs upon the formation of a complex (i.e., "rigid body docking"), the binding free energy can be estimated by (Zhang et al., 1997):

$$\Delta G = \Delta E_C + \Delta E_{El} - T\Delta S_{trv}, \tag{4}$$

where ΔE_C and ΔE_{El} are, respectively, the total contact energies and electrostatic energies between the two interacting molecules and ΔS_{trv} is the entropy change of making one molecule out of two. Specifically, ΔS_{trv} includes entropy loss associated with the six degrees of translational/rotational freedom and the compensating vibrational entropy change associated with the new low frequency modes that appear in the complex. The statistical mechanical estimate of ΔS_{trv} is −50 cal/K·mol at 25 °C (Finkelstein & Janin, 1989; Janin, 1995).

The association of peptide and protein involves a disorder/order transition in the peptide fragment. The bound segment has a unique conformation, but the unbound fragment has multiple conformations. An additional term representing the entropy change associated with this transition therefore must be included explicitly in the binding free energy calculation. Equation 4 is then modified to

$$\Delta G = \Delta E_C + \Delta E_{El} - T\Delta S_{trv} - T(-\Delta S_{bb}^{pep}), \tag{5}$$

where ΔS_{bb}^{pep} is the entropy change of the peptide backbone from a unique conformation to an unstructured state.

In principle, a complete binding free energy function should also contain a term reflecting the internal energy change associated with the disorder-to-order transition of the peptide. However, for consistency, we must take the unfolded state to be highly extended, and, because the bound state is also extended for peptides that bind class I MHC molecules, the internal free energy difference will be small compared to other terms in the expression.

Results and discussion

Backbone conformational entropy change in protein unfolding

In this study, we selected the 17 single-domain proteins in Makhatadze and Privalov (1995) for which denaturation is reversible and is approximated well by a two-state transition. For these proteins, calorimetric measurements of the Gibbs free energies, and of the enthalpies and entropies of transition between the native and denatured states, are available (see Makhatadze & Privalov, 1995 and references therein).

The entropy of protein unfolding results from a change in the conformational freedom of the polypeptide, and from hydration of the groups exposed to water. Because atomic contact energies contain contributions from side-chain entropy changes and solvation free energies, the difference between the measured free energy and the sum of electrostatic and atomic contact energies should approximate ΔS_{bb}^{prot}, the polypeptide backbone conformational entropy (Equation 2). In this way, we calculated ΔS_{bb}^{prot} for each of the 17 proteins (Table 1) and found that the mean backbone conformational entropy change per residue averaged over the 17 proteins is $\Delta S_{bb} = 5.3 \pm 0.6$ cal/K·mol-res.

The backbone entropy change of protein unfolding is a function of the peptide chain length and composition, as well as the presence of disulfide bridges or other covalent bonds in the backbone. In the absence of covalent links, the change in backbone entropy is primarily a function of the steric hindrances imposed by the side

Table 1. *Energetics of protein folding/unfolding transition*[a]

Protein	PDB[b]	N_r	$\Delta_N^U G^{exp}$	$\Delta_N^U H^{exp}$	$\Delta_N^U S^{exp}$	$\Delta_N^U E_C$	$\Delta_N^U E_{El}$	$\Delta\Delta S_i$	ΔS_{bb}	ΔS_{Gly}
SH3	1shg	57	3.47	12.44	30.14	73.42	37.32	87.50	6.32	7.85
BPTI	5pti	58	10.57	31.10	68.90	66.89	27.68	81.79	4.86	6.27
CI-2	2ci2	65	6.63	32.30	86.12	59.78	28.38	110.02	4.21	5.90
Eglin c	1acb	63	8.83	27.51	62.68	73.95	35.34	92.58	5.35	6.82
G Protein	1pgx	70	5.70	16.03	34.69	67.55	35.71	104.18	4.68	6.17
Tendamistat	1hoe	74	8.97	16.75	26.08	78.22	38.68	104.78	4.89	6.31
Ubiquitin	1ubq	76	9.59	6.46	−10.53	86.83	43.73	113.94	5.34	6.84
RNase T1	8rnt	104	8.97	67.22	195.45	118.69	51.02	128.37	5.19	6.42
Cytochrome *c*	5cyt	103	8.88	21.29	41.63	107.42	40.88	137.86	4.54	5.88
Barnase	1rnb	109	11.70	73.45	207.18	100.68	49.36	145.88	4.26	5.60
RNase A	7rsa	124	6.46	70.33	214.35	150.99	76.60	175.00	5.98	7.40
Lysozyme	1lz1	130	13.83	57.89	147.85	166.46	71.96	183.47	5.80	7.21
Interleukin-1β	1i1b	151	7.54	36.12	95.93	178.50	55.19	221.04	5.03	6.49
Myoglobin	1mbo	153	9.71	1.44	−27.75	206.34	105.21	238.61	6.62	8.18
T4 Lysozyme	3lzm	164	16.36	57.42	137.80	206.20	100.38	241.51	5.94	7.41
Papain	9pap	212	22.42	39.23	56.46	238.67	110.54	289.44	5.17	6.54
Pepsinogen	2psg	370	18.59	17.22	−4.55	514.36	178.15	505.84	6.11	7.48
Mean									5.31	6.73

[a]Units for $\Delta_N^U G^{exp}$, $\Delta_N^U H^{exp}$, $\Delta_N^U E_C$, and $\Delta_N^U E_{El}$ are kcal/mol; units for $\Delta_N^U S^{exp}$ and $\Sigma\Delta\Delta S_i$ are cal/K·mol; units for ΔS_{bb} and ΔS_{Gly} are cal/K·mol-res.
[b]Coordinate files were taken from Protein Data Bank (Bernstein et al., 1977).

chains on the rotational degrees of freedom of the peptide chain about backbone bonds (Nemethy et al., 1966). This entropy is maximal for glycine; replacement of glycine by other side chains alters the range and distribution of ϕ, ψ angles of the polypeptide backbone and stabilizes the native state relative to the denatured states (Stites & Pranata, 1995).

If the influence of side chains on the backbone conformational entropy is considered explicitly, we can estimate the backbone entropy for glycine, ΔS_{Gly}. Following Murphy et al. (1993), the mean conformational entropy change per residue upon protein unfolding, ΔS_{bb}, can be written, to a first approximation, in terms of the conformational entropy change of each amino acid type, ΔS_i, weighted by its occurrence, P_i:

$$\Delta S_{bb} = \sum_{i=1}^{20} P_i \Delta S_i. \tag{6}$$

Connection with experiment is made most readily if ΔS_i is written as a differential entropy with glycine as reference:

$$\Delta S_{bb} = \sum_{i=1}^{20} P_i(\Delta S_{Gly} - \Delta\Delta S_i), \tag{7}$$

where $\Delta\Delta S_i$ is the difference in backbone entropies between amino acid type i and glycine. Using the values of $\Delta\Delta S_i$ suggested by Stites and Pranata (1995), the average value of ΔS_{Gly} for the 17 proteins is 6.7 cal/K·mol-res (Table 1). We therefore estimate the difference between the entropy of glycine and the average entropy of all residues as 1.4 cal/K·mol-res.

There are a number of theoretical ways to estimate the mean entropy gain by the backbone chain upon unfolding (Schellman, 1955; Nemethy & Scheraga, 1965; Spolar & Record, 1994; Yang & Honig, 1995; D'Aquino et al., 1996; Wang & Purisima, 1996). Schellman (1955) first estimated that the loss of conformational entropy of the peptide backbone upon α-helix formation must lie between 3.0 and 7.18 cal/K·mol-res, with a mean of about 5.0 cal/K·mol-res. Recently, Spolar and Record (1994) deduced a value of 5.6 cal/K·mol-res based on experimental measurements of the heat capacity change for protein folding and their model for the contribution of polar and nonpolar group burial to the measured heat capacity. Honig and colleagues concluded from a separate calculation of conformational entropies and hydration effects that the backbone conformation entropy of protein unfolding is 6.7 cal/K·mol-res (Yang & Honig, 1995). More recently, Wang and Purisima (1996) reexamined the mean backbone entropy loss in a helical matrix and their estimate, 5.0 cal/K·mol-res, is in agreement with the previous result of Schellman (1955).

Nemethy and Scheraga (1965) estimated 21 conformations for glycine dipeptides or, equivalently, a conformational entropy of 6.1 cal/K·mol-res. A recent estimate by Freire and colleagues (D'Aquino et al., 1996) suggests that, with chain length-dependent excluded volume effects considered explicitly, ΔG_{Gly} has a value of 6.4 cal/K·mol-res. Thus, according to the most recent estimates, the difference between the average residue entropy and the entropy of glycine is (6.4 − 5.0 =) 1.4 cal/K·mol-res. This coincides with our own estimate for the differential entropy. If we take this number seriously, we would conclude that the average backbone conformational entropy is 5.0 cal/K·mol-res, slightly below our estimate of 5.3 cal/K·mol-res. In fact, it seems likely that 5.3 cal/K·mol-res is a slight overestimate because the use of the extended conformation for the unfolded state may underestimate intramolecular energies. Considering the magnitude of error in experimental and calculated values, however, these distinctions are likely to be marginal at best.

The effect of disulfide bonds is not considered explicitly in our analysis. According to the statistical theory of a random coil, a single disulfide crosslinking two atoms separated by N bonds in a Gaussian chain reduces the conformational entropy by $-(2.1 + 3/2R \ln N)$ cal·K^{-1}·mol^{-1} (DeLisi & Crothers, 1971; Pace et al.,

1988). For large chains, the fractional error in neglecting such disulfide loops goes as (ln *N*)/*N*, and the error drops below 10% for as few as 20–30 bonds. We therefore do not expect our conclusions to be affected appreciably.

Forces contributing to protein stability

Although atomic contact energies include solvation, side-chain entropies, and differential van der Waals energies in a way that cannot be uncoupled easily, we can gain insight into the dominant driving factors in protein folding by parsing the sum of the contact energies and electrostatic energies (i.e., total energy) into effective interacting energies between backbone–backbone ($\Delta E_{b\text{-}b}$), backbone–side-chain ($\Delta E_{b\text{-}s}$), and side-chain–side-chain ($\Delta E_{s\text{-}s}$) (Table 2). The net sum of these terms drives folding, and must be sufficient to compensate for the highly unfavorable backbone entropy loss.

The most striking observation emerging from this dissection is that the interactions associated with the backbone provide the dominant driving force. On average, backbone–backbone interactions contribute 64% of the total energy; backbone–side-chain interactions, 26%; and side-chain–side-chain interactions, 10%. For the smaller proteins, the sum of $\Delta E_{b\text{-}b}$ and $\Delta E_{b\text{-}s}$ almost entirely compensates for the unfavorable backbone conformational entropy change. The electrostatic component of backbone interactions stem predominantly from partial charges that contribute to the peptide dipole. These electrostatic interactions make a substantial contribution to the backbone–backbone interaction energy, having a typical magnitude approximately 2/3 that of the atomic contact energies.

It is generally agreed that the hydrophobic effect (i.e., the favorable free energy change accompanying desolvation of nonpolar groups) is the major contributor to stabilization of the folded structure of globular proteins, whereas hydrogen bonding and steric hindrances constitute structural constraints on the internal architecture (Dill, 1990; Yang & Honig, 1995). The results presented here lend some precision to that concept by showing that, energetically, folding is driven predominantly by backbone interactions that have both electrostatic and solvation components. The hydrophobic interactions of side chains provide a marginal gain in the overall stability, but both electrostatic and hydrophobic interactions of side chains are responsible for specifying the unique native structure of a protein. To the extent that backbone interactions favor folding, they are probably not very selective for one compact conformation relative to others.

Although the contributions of backbone interactions have often been overlooked by theoretical analyses, there is accumulating experimental evidence that the chemical nature of the polypeptide backbone itself plays a crucial role in protein folding. One of the most convincing examples is the spontaneous formation of α-helix in short alanine-based peptides in solution (Marqusee et al., 1989; Chakrabatty & Baldwin, 1995). Because the side chain of alanine can interact only minimally with other side chains, helix formation by alanine is stabilized predominantly by the backbone. The implication for protein folding is that the backbone interactions can largely offset the unfavorable entropy change caused by fixing the polypeptide backbone. Favorable side-chain interactions would provide the marginal gain in protein stability, whereas unfavorable side-chain interactions would break down the the balance and favor unfolding. It has been found that short peptides that do not contain significant amounts of alanine will not form stable helices in water unless they are otherwise stabilized by side-chain–side-chain interactions (Chakrabatty & Baldwin, 1995). In fact, most isolated helical segments of proteins are unstable in water. Recently, based on directed mutagenesis and high-resolution structure analysis of the bacteriophage T4 lysozyme, it was found that the replacement of a large number of amino acids (50% or more) with alanine has little effect on the successful folding of the protein (Matthews, 1996). Such "polyalanine" approach clearly demonstrates that, energetically, side-chain interactions are not the dominant driving force, but, rather, are responsible for the internal architecture of a protein through the specific distribution of "essential" residues.

Table 2. *Effective interacting energies (kcal/mol-res) between backbone–backbone($\Delta E_{b\text{-}b}$), backbone–side-chain ($\Delta E_{b\text{-}s}$), and side-chain–side-chain ($\Delta E_{s\text{-}s}$)*

	$\Delta E_{b\text{-}b}$		$\Delta E_{b\text{-}s}$		$\Delta E_{s\text{-}s}$		
Protein	ACE	Elect.	ACE	Elect.	ACE	Elect.	$T\Delta S_{bb}$
SH3	0.68	0.42	0.49	0.16	0.11	0.07	1.88
BPTI	0.56	0.24	0.52	0.14	0.08	0.10	1.45
CI-2	0.61	0.36	0.26	0.03	0.05	0.05	1.25
Eglin c	0.62	0.44	0.40	0.02	0.16	0.10	1.59
G Protein	0.62	0.42	0.27	0.07	0.07	0.02	1.39
Tendamistat	0.66	0.37	0.35	0.13	0.05	0.02	1.46
Ubiquitin	0.66	0.47	0.32	0.06	0.16	0.04	1.59
RNase T1	0.59	0.34	0.43	0.11	0.12	0.05	1.55
Cytochrome *c*	0.70	0.36	0.26	0.02	0.09	0.01	1.35
Barnase	0.63	0.29	0.28	0.12	0.02	0.04	1.27
RNase A	0.69	0.37	0.46	0.18	0.07	0.06	1.78
Lysozyme	0.73	0.38	0.35	0.10	0.20	0.08	1.73
Interleukin-1β	0.67	0.34	0.42	−0.03	0.09	0.05	1.50
Myoglobin	0.82	0.66	0.36	0.00	0.17	0.03	1.97
T4 Lysozyme	0.76	0.53	0.35	0.01	0.14	0.07	1.77
Papain	0.69	0.34	0.37	0.11	0.07	0.08	1.54
Pepsinogen	0.70	0.39	0.43	0.05	0.25	0.04	1.82

The empirical analyses of protein folding/unfolding thermodynamic data in general have difficulty in separating the conformational entropies ($\Delta S^{conf} = \Delta S_{bb} + \Delta S_{sc}$) from the entropies of solvation (Makhatadze & Privalov, 1995; D'Aquino et al., 1996). This has prevented a firm grasp of the role of entropic contributions to the folding process. The energy function developed here permits us to separate ΔS_{bb} from the rest of the contributions to protein stability and to demonstrate that the theoretically estimated value, 5 cal/K·mol-res, is consistent with the thermodynamic data. However, our analysis does not provide a direct estimate of the magnitude of the total conformational entropy of protein unfolding. Using the mean backbone entropy gain $\Delta S_{bb} = -5$ cal/K·mol-res and the mean side-chain entropy change $\Delta S_{sc} = 3.66$ cal/K·mol-res (Pickett & Sternberg, 1993; Doig & Sternberg, 1995), ΔS^{conf} comes up to ~9 cal/K·mol-res, which is somewhat in between the values estimated by two recent studies: 6 cal/K·mol-res by Freire and colleagues (D'Aquino et al., 1996) and 12 cal/K·mol-res found by Makhatadze and Privalov (1995).

All these values should be interpreted cautiously because the actual contribution of side-chain entropy (ΔS_{sc}) to the unfolding reactions remains elusive. A number of highly correlated scales for the entropy changes of transferring side chains that are buried in the interior of the protein to a "free" state have been introduced (Finkelstein & Janin, 1989; Pickett & Sternberg, 1993; Abagyan & Totrov, 1994; Koehl & Delarue, 1994; Creamer & Rose, 1994; Lee et al., 1994). The consensus reached for the free energy gain of relaxing side-chain motion is ~0.5 kcal/mol per rotamer at 25 °C (Doig & Sternberg, 1995). However, as pointed out by Makhatadze and Privalov (1995), most of these scales were based on rotamer distributions observed or calculated in the presence of solvent. There is a possibility that the calculated side-chain entropies are biased by various solvation effects, and thus are smaller than the "actual" values (Makhatadze & Privalov, 1995, 1996). On the other hand, not all the side chains in the native structure of a protein are buried completely and, therefore, the side-chain entropy scales cannot be used directly unless the proportion of the side chains that are truly restricted in the native state is taken into account. Freire and colleagues separated side-chain conformational entropies into $\Delta S^{bu\text{-}ex}$ (associated with the transfer of a buried side chain to the surface: 0.82 cal/K·mol-res) and $\Delta S_{sc}^{ex\text{-}u}$ (associated with the unfolding of a solvent-exposed side chain: 2.62 cal/K·mol-res) and estimated the average conformational entropy contribution upon unfolding due to side chain to be ~1.4 kcal/K·mol-res (Lee et al., 1994; D'Aquino et al., 1996).

The main focus of this study is to develop a free energy function based on structural data, which can represent consistently the energetics of folding and binding. Of the three terms in our unfolding energy function—contact energy, electrostatic energy, and backbone conformational entropy—only the backbone conformational entropy is a free parameter, and our analysis indicates that the optimal ΔS_{bb} value for reproducing thermodynamic data is consistent with theoretical estimates. In the next section, we further justify $\Delta S_{bb} = -5$ cal/K·mol-res by applying it to the calculation of protein/peptide binding free energies where ΔS_{bb} and solvation entropies are clearly uncoupled.

MHC–peptide interaction

We estimated the binding free energies of HLA-A2.1 molecule complexed with four nonamers based on the solved crystal structures (Madden et al., 1993). The four peptides were derived from HIV-1 gp120 (GP), Influenza A matrix (MT), HIV-1 RTase (RT), and HTLV-1 TAX (TX), respectively. Sette and colleagues (1994) have measured the binding of these four peptides to HLA-A2.1 using competition experiments. The IC_{50} values (i.e., the concentration that causes 50% inhibition) are: GP, 294 nM (Altuvia et al., 1995); MT, 6 nM (Sette et al., 1994); RT, 909 nM (Ruppert et al., 1993); and TX, 11 nM (Altuvia et al., 1995). The IC_{50} value can be related to the equilibrium dissociation constant K_i of the test peptide ligand by $K_i = IC_{50}/(1 + [L]/K_d)$ (Cheng & Prusoff, 1973), where K_d is dissociation constant of the radiolabeled ligand and $[L]$ the concentration of the radiolabeled ligand in the free (unbound) state. Under the assay conditions for binding of the four peptides, $[L]/K_d \ll 1$; thus, $K_i \approx IC_{50}$ and $\Delta G^{exp} \approx RT\ln(IC_{50})$ (A. Sette, pers. comm.).

Our goal here is to use the experimentally derived data as criteria in evaluating the reliability of the proposed binding free energy function, Equation 5. Because the Protein Data Bank files (Bernstein et al., 1977) do not provide coordinates for polar hydrogens, which are essential for the calculation of electrostatic energies, polar hydrogens were introduced into the crystallographic structures by using the program CHARMM (Brooks et al., 1983). The structures were then energy minimized for 200 steps of an ABNR procedure with a nonbond cutoff distance of 20 Å. Two sets of binding free energies, ΔG^{theo}, were calculated using different estimates of ΔS_{bb}^{pep}. The first set takes into account the influence of side chains on backbone entropy; ΔS_{bb}^{pep} was calculated by

$$\Delta S_{bb}^{pep} = \sum_{i=1}^{9} (\Delta S_{Gly} - \Delta\Delta S_i), \tag{8}$$

where ΔS_{Gly} is 6.4 cal/K·mol-res and the $\Delta\Delta S_i$ values were taken from Stites and Pranata (1995). For the second set, the mean backbone entropy change ΔS_{bb} (5.0 cal/K·mol-res) is used for all amino acids, and thus

$$\Delta S_{bb}^{pep} \approx 9 \times 5.0 = 45.0 \text{ cal/K·mol.} \tag{9}$$

The calculated free energies ΔG^{theo} are shown in Table 3 (column 7), along with experimental binding data ΔG^{exp} (column 9). The values in parentheses (column 7) are the estimated binding free energies when Equation 9 is used to calculate ΔS_{bb}^{pep}. The calculated energies have the same rank order as the experimental energies and the maximum deviation is 1.8 kcal/mol. The results are noticeably better than those obtained previously (Vajda et al., 1994), which are based on a different energy decomposition. Slightly better results were obtained when Equation 8 was used to calculate ΔS_{bb}^{pep}. In this case, the largest difference between the calculated and measured free energies is about 1.3 kcal/mol, comparable to the accuracy of the calculation for "rigid body" docking (Zhang et al., 1997). These results indicate that the energy terms for protein folding, including ACE, electrostatics, and backbone conformational entropy, appear to be transferable to the energetics of peptide binding. The results also suggest that the free energy decomposition presented here promises to provide a reasonably accurate and computationally viable target function for peptide docking and design (Rognan et al., 1994; Vajda et al., 1994; Rosenfeld et al., 1995; Gulukota et al., 1996; Sezerman et al., 1996).

Conclusion

A new free energy decomposition for the folding/unfolding transition of single-domain proteins is presented here. The empirical

Table 3. *Binding free energies of HLA-A2.1 complexed with four peptides*[a]

Peptide	PDB[b]	Sequence	ΔE_C	ΔE_{EI}	ΔS_{bb}^{pep}	ΔG^{theo} [c]	IC_{50}	ΔG^{exp}
HIV-1 gp120 (GP)	1hhg	TLTSCNTSV	−2.35	−34.56	48.25	−7.80 (−8.50)	294	−8.92
Influenza A (MT)	1hhi	GILGFVFTL	−8.75	−32.71	47.73	−12.50 (−13.05)	6	−11.22
HIV-1 RTase (RT)	1hhj	ILKEPVHGV	4.59	−39.74	43.37	−7.49 (−6.74)	909	−8.25
HTLV-1 TAX (TX)	1hhk	LLFGYPVYV	−7.87	−31.01	44.86	−10.78 (−10.47)	11	−10.87

[a]Units for the energies are kcal/mol, unit for IC_{50} is nM.
[b]Coordinate files were taken from the Protein Data Bank (Bernstein et al., 1977).
[c]Numbers in parentheses present binding free energies when ΔS_{bb}^{pep} is estimated by Equation 9.

approach involves evaluating solvation free energy using ACE for both unfolded state and native conformation. Polypeptide backbone entropy was then calculated by subtracting all other contributions, as estimated by ACE and electrostatic energies, from the calorimetrically measured unfolding free energy. The mean entropic cost associated with fixing backbone dihedral angles, ΔS_{bb}, was found to be 5.3 cal/K·mol-res, in excellent agreement with the value 5.0 cal/K·mol-res obtained by several theoretical approaches (Schellman, 1955; Wang & Purisima, 1996). The slight overestimate in our analysis is attributed to the use of extended chain as a model for the unfolded state. Interestingly, the average backbone entropy for glycine obtained here, ΔS_{Gly} = 6.7 cal/K·mol-res, is also 0.3 cal/K·mol-res higher than the theoretically estimated value (D'Aquino et al., 1996). A consensus on the magnitude of changes in backbone conformational entropy represents a necessary step toward structure-based stability calculations with the accuracy needed for protein design. In addition, our analysis also highlights the important role played by the backbone in the generation of protein structure. A dissection of the free energies indicates that most of the negative backbone conformational entropy change upon folding is overcome by the effective interactions associated with backbone itself.

The values confirmed for backbone entropy were then incorporated into a free energy function for the calculation of binding free energies of protein–peptide association. Calculated free energy values are in good agreement with experimental data. Slightly better results were obtained when the influence of side chain on backbone entropy was considered explicitly so that the inhomogeneity of the amino acid compositions in the peptide was taken into account. The fact that our approach reproduces the observed values of free energy to within about 1 kcal/mol without parameter adjustment suggests that the free energy breakdown involved has been reasonably well described by our calculation.

More generally, it appears from this study that peptide recognition and protein folding are based on the same physicochemical principle. Precise estimations of the conformational entropy changes and solvation effects are essential for the accurate prediction of protein stability and binding affinity. The energy decomposition presented here provides a useful tool that will facilitate the engineering of proteins as well as the rational design of peptide ligands.

Acknowledgments

This work has been sponsored by a grant (AI30535) from the NIH/NIAID. We thank Drs. A. Sette and H. Margalit for informative communication.

References

Abagyan R, Totrov M. 1994. Biased probability Monte-Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol 235*:983–1002.
Altuvia Y, Schueler O, Margalit H. 1995. Ranking potential binding peptides to MHC molecules by a computational threading approach. *J Mol Biol 249*:244–250.
Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science 181*:223–230.
Baldwin RL. 1986. Temperature dependence of the hydrophobic interactions in protein folding. *Proc Natl Acad Sci USA 83*:8069–8072.
Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.
Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem 4*:187–217.
Chakrabatty A, Baldwin RL. 1995. Stability of α-helices. *Adv Protein Chem 46*:141–176.
Cheng YC, Prusoff WH. 1973. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50% inhibition (IC_{50}) of an enzymatic reaction. *Biochem Parmacol 22*:3099–3108.
Creamer TP, Rose GD. 1994. α-Helix-forming propensities in peptides and proteins. *Proteins Struct Funct Genet 19*:85–97.
D'Aquino JA, Comez J, Hilser VJ, Lee KH, Amzel LM, Freire E. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins Struct Funct Genet 25*:143–156.
DeLisi C, Crothers DM. 1971. Electrostatic contributions to oligonucleotide transitions. *Biopolymers 10*:2323–2343.
Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7155.
Doig AJ, Sternberg MJE. 1995. Side-chain conformational entropy in protein folding. *Protein Sci 4*:2247–2251.
Eisenberg D, McLachlan AD. 1986. Solvation energy in protein folding and binding. *Nature 319*:199–203.
Finkelstein A, Janin J. 1989. The price of lost freedom: Entropy of biomolecular complex formation. *Protein Eng 3*:1–3.
Freire E, Haynie DT, Xie D. 1993. Molecular-basis of cooperativity in protein-folding. IV. CORE—A general cooperative folding model. *Proteins Struct Funct Genet 17*:111–123.
Gulukota K, Vajda S, DeLisi C. 1996. Peptide docking using dynamic programming. *J Comput Chem 17*:418–428.
Honig B, Yang AS. 1995. Free energy balance in protein folding. *Adv Protein Chem 46*:27–58.
Janin J. 1995. Elusive affinities. *Proteins Struct Funct Genet 21*:30–39.
Kauzmann W. 1953. Some factors in the interpretation of protein denaturation. *Adv Protein Chem 14*:1–63.
Koehl P, Delarue M. 1994. Application of self-consistent mean field theory to predict side-chain conformation and estimate their conformational entropy. *J Mol Biol 239*:249–275.
Lazaridis T, Archontis G, Karplus M. 1995. Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. *Adv Protein Chem 47*:231–306.
Lee KH, Xie D, Freire E, Amzel LM. 1994. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. *Proteins Struct Funct Genet 20*:68–84.
Levinthal C. 1968. Are there pathways in protein folding? *J Chim Phys 65*:44–45.

Madden DR, Garboczi DN, Wiley DC. 1993. The antigenic identity of peptide-MHC complexes: A comparison of the conformations of five viral peptides presented by HLA-A2. *Cell 75*:693–708.

Makhatadze GI, Privalov PL. 1993. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J Mol Biol 232*:639–659.

Makhatadze GI, Privalov PL. 1995. Energetics of protein structure. *Adv Protein Chem 47*:307–425.

Makhatadze GI, Privalov PL. 1996. On the entropy of protein folding. *Protein Sci 5*:507–510.

Marqusee S, Robbins V, Baldwin RL. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc Natl Acad Sci USA 86*:5286–5290.

Matthews BW. 1996. Structural and genetic analysis of the folding and function of T4 lysozyme. *FASEB J 10*:35–41.

Miyazawa S, Jernigan RL. 1985. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules 18*:534–552.

Murphy K, Freire E. 1992. Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv Protein Chem 43*:313–361.

Murphy K, Xie D, Garcia K, Amzel L, Freire E. 1993. Structural energetics of peptide recognition: Angiotenin II/antibody binding. *Proteins Struct Funct Genet 15*:113–120.

Nemethy G, Leach S, Scheraga H. 1966. The influence of amino acid side chains on the free energy of helix-coil transitions. *J Phys Chem 70*:998–1004.

Nemethy G, Scheraga HA. 1965. Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. *Biopolymers 3*:155–184.

Novotny J, Bruccoleri RE, Saul FA. 1989. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry 28*:4735–4749.

Oobatake M, Ooi T. 1992. Hydration and heat stability effects on protein unfolding. *Prog Biophys Mol Biol 59*:237–284.

Pace C, Grimsley G, Thomson J, Barnett B. 1988. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem 263*:1820–1825.

Pickett SD, Sternberg MJE. 1993. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol 231*:825–839.

Privalov PL. 1979. Stability of protein. Small globular proteins. *Adv Protein Chem 33*:167–241.

Privalov PL, Gill SJ. 1988. Stability of protein structure and hydrophobic interaction. *Adv Protein Chem 39*:191–234.

Privalov PL, Makhatadze GI. 1993. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J Mol Biol 232*:660–679.

Rognan D, Scapozza L, Folkers G, Daser A. 1994. Molecular dynamics simulation of MHC-peptide complexes as a tool for predicting potential T cell epitopes. *Biochemistry 33*:11476–11485.

Rosenfeld R, Vajda S, DeLisi C. 1995. Flexible docking and design. *Annu Rev Biophys Biomol Struct 24*:677–700.

Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A. 1993. Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. *Cell 74*:929–937.

Schellman J. 1955. The stability of hydrogen-bonded peptide structures in aqueous solution. *Trav Lab Carlsberg Ser Chim 29*:230–259.

Sette A, Sidney J, del Guercio MF, Southwood S, Ruppert J, Dahlberg C, Grey HM, Kubo RT. 1994. Peptide binding to the most frequent HLA-A class I alleles measured by quantitative molecular binding assays. *Mol Immunol 31*:813–822.

Sezerman OU, Vajda S, DeLisi C. 1996. Free energy mapping of class I MHC molecules and structural determination of bound peptides. *Protein Sci 5*:1272–1281.

Spolar R, Record M. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*:777–784.

Stites WE, Pranata J. 1995. Empirical evaluation of the influence of side chains on the conformational entropy of the polypeptide backbone. *Proteins Struct Funct Genet 22*:132–140.

Tanford CH. 1968. Protein denaturation. *Adv Protein Chem 23*:121–242.

Vajda S, Weng Z, Rosenfeld R, DeLisi C. 1994. Effect of conformational flexibility and solvation on receptor–ligand binding free energies. *Biochemistry 33*:13977–14988.

Vasmatzis G, Zhang C, Cornette JL, DeLisi C. 1997. Computational determination of side chain specificity of pockets in class I MHC molecules. Forthcoming.

Wang J, Purisima EO. 1996. Analysis of thermodynamic determinants in helix propensities of non-polar amino acids through a novel free energy calculation. Forthcoming.

Yang AS, Honig B. 1995. Free energy determinants of secondary structure formation. I. α-Helices. *J Mol Biol 252*:351–365.

Zhang C, Vasmatzis G, Cornette JL, DeLisi C. 1997. Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol 267*:707–726.