Structure-Based Thermodynamic Scale of α-Helix Propensities in Amino Acids[†]

Irene Luque, Obdulio L. Mayorga, and Ernesto Freire*

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

Received June 4, 1996; Revised Manuscript Received August 19, 1996[⊗]

ABSTRACT: A structural parameterization of the folding energetics has been used to predict the effect of single amino acid mutations at exposed locations in α-helices. The results have been used to derive a structure-based thermodynamic scale of α-helix propensities for amino acids. The structure-based thermodynamic analysis was performed for four different systems for which structural and experimental thermodynamic data are available: T4 lysozyme [Blaber et al. (1994) *J. Mol. Biol. 235*, 600−624], barnase [Horovitz et al. (1992) *J. Mol. Biol. 227*, 560−568], a synthetic leucine zipper [O'Neil & Degrado (1990) *Science 250*, 646−651], and a synthetic peptide [Lyu et al. (1990) *Science 250*, 669−673]. These studies have permitted the optimization of the set of solvent-accessible surface areas (ASA) for all amino acids in the unfolded state. It is shown that a single set of structure/thermodynamic parameters accounts well for all the experimental data sets of helix propensities. For T4 lysozyme, the average value of the absolute difference between predicted and experimental ΔG values is 0.09 kcal/mol, for barnase 0.14 kcal/mol, for the synthetic coiled-coil 0.11 kcal/mol, and for the synthetic peptide 0.08 kcal/mol. In addition, this approach predicts well the overall stability of the proteins and rationalizes the differences in α-helix propensities between amino acids. The excellent agreement observed between predicted and experimental ΔG values for all amino acids validates the use of this structural parameterization in free energy calculations for folding or binding.

Molecular recognition, either intramolecular as in protein folding or intermolecular as in ligand binding, hormone/receptor binding, or substrate/inhibitor binding, is a major problem in modern biology. The interest in this area is not only academic since a clear understanding of these processes will facilitate the development of rational design strategies aimed at producing proteins with new functionalities, altered stabilities, pharmaceutical drugs that associate to specific receptors with specific affinities, etc.

The most fundamental quantity necessary to describe the stability of a protein or the affinity between two molecules is the Gibbs energy, ΔG. For this reason, the development of structure-based molecular design protocols requires the ability to predict ΔG from structural considerations. A promising approach is based on the generalization of rules obtained from correlations between experimentally determined thermodynamic parameters and structural parameters (Gomez & Freire, 1995; Murphy et al., 1992; Murphy & Freire, 1992; Spolar et al., 1992; Spolar & Record, 1994). The development of such a structure-based thermodynamic approach has been accelerated by the rapid increase in the number of proteins and protein complexes for which high-resolution structures and high-resolution thermodynamic data are available. While less detailed and explicit than the use of molecular mechanics force fields, these empirical methods effectively incorporate all solvent effects since the structural correlations are made directly with experimental parameters.

The approach devised in this laboratory involves the development of a separate empirical structural parameterization of the three components that define the Gibbs energy: the enthalpy, entropy, and heat capacity changes (DAquino et al., 1996; Gomez et al., 1995; Hilser et al., 1996; Lee et al., 1994; Murphy et al., 1992). This approach requires the availability of precise thermodynamic data for systems for which high-resolution structures are available. The extent and quality of this database ultimately determine the accuracy and predictive ability of the parametrization. The potential of this approach has been demonstrated in several ways, including the prediction of structural features of equilibrium folding intermediates for several proteins (e.g., α-lactalbumin, barnase, staphylococcal nuclease, T4 lysozyme) (Freire, 1995; Hilser et al., 1996; Xie et al., 1994; Xie & Freire, 1994a,b) and the formation of peptide/protein or protein/protein complexes (e.g., angiotensin/antibody, pepstatin/endothiapepsin) (Gomez & Freire, 1995; Murphy et al., 1993). In this paper, we apply this approach to the calculation of α-helix propensities for individual amino acids.

An important structural variable in the structural parameterization of the energetics is the change in the solvent accessibility of different groups during folding or binding. Changes in solvent accessibility are directly related to changes in hydration and indirectly to the changes in internal interactions (van der Waals, hydrogen bonds, etc.) that are established during folding or binding. These changes in solvent accessibility are usually expressed as changes in the solvent-accessible surface area (ASA) (Connolly, 1983; Lee & Richards, 1971; Shrake & Rupley, 1973). It has been shown by different groups that significant portions of the heat capacity, enthalpy, and hydration entropies scale in terms of ASA changes (Gomez & Freire, 1995; Murphy et al., 1992; Murphy & Freire, 1992; Ooi et al., 1987; Spolar et al., 1992; Spolar & Record, 1994). However, the solvent-

[†] Supported by grants from the National Institutes of Health (RR04328 and GM51362). I.L. is a visiting student from the Universidad de Granada, Granada, Spain, partially supported by a fellowship from the Ministerio de Educación y Ciencia (PB93-1163).

* Correspondence should be addressed to this author. Voice: (410) 516-7743. Fax: (410) 516-6469. E-mail: bcc@biocal2.bio.jhu.edu.

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0006-2960(96)01319-0 CCC: $12.00

accessible surface area, especially that of unfolded conformations, is not a precisely defined quantity and therefore limits the precision with which thermodynamic parameters can be calculated from structure. In this paper, we present a refined procedure to estimate ASA values in the unfolded state and consequently ASA changes associated with folding/unfolding. Together with the recent evaluation of backbone conformational entropies (DAquino et al., 1996), all the basic elements in the structural parametrization are present for the first time and have allowed us to assess the effects of individual amino acids. In this paper, the structural parameterization is used to develop a α-helix propensity scale for 19 amino acids.

RESULTS AND DISCUSSION

Structural Parameterization of Folding Energetics. The most important quantity in the thermodynamic description of folding or binding is the Gibbs energy (ΔG) which is completely specified if the enthalpy (ΔH), entropy (ΔS), and heat capacity (ΔC_p) changes are known at some reference temperature (T_R):

$$\Delta G = \Delta H(T) - T\cdot\Delta S(T) \quad (1)$$

$$\Delta H(T) = \Delta H(T_R) + \int_{T_R}^{T} \Delta C_p \, dT \quad (2)$$

$$\Delta S(T) = \Delta S(T_R) + \int_{T_R}^{T} \Delta C_p \, d \ln T \quad (3)$$

Different studies have shown that some of the contributions that define the Gibbs energy of folding and binding can be expressed in terms of changes in solvent-accessible surface areas (ΔASA). In particular, it has been shown by several groups that the heat capacity change can be accurately written as a linear combination of the changes in apolar (ΔASA_{ap}) and polar (ΔASA_{pol}) solvent-accessible surface areas (Gomez et al., 1995; Murphy et al., 1992; Spolar et al., 1992; Spolar & Record 1994). Similarly, the enthalpic contributions due to the disruption of internal van der Waals interactions and subsequent hydration of the participating groups also scale in terms of ΔASA changes (Hilser et al., 1996; Xie & Freire, 1994a,b). Finally, of the different contributions that define the entropy change, the hydration term is also proportional to changes in ΔASA_{ap} and ΔASA_{pol} (DAquino et al., 1996; Lee et al., 1994).

The structural parameterization of the free energy involves calculation of the relative heat capacity (ΔC_p), enthalpy (ΔH), and entropy (ΔS) at the desired temperature. The heat capacity change is a weak function of temperature and has been parameterized in terms of changes in solvent-accessible surface areas (ΔASA) since it originates mainly from changes in hydration (Gomez et al., 1995; Murphy et al., 1992):

$$\Delta C_p = \Delta C_{p,ap} + \Delta C_{p,pol} \quad (4a)$$

$$\Delta C_p = a_C(T)\cdot\Delta ASA_{ap} + b_C(T)\cdot\Delta ASA_{pol} + c_C(T)\cdot\Delta ASA_{OH} \quad (4b)$$

where the coefficients $a_C(T) = 0.45 + (2.63 \times 10^{-4})(T - 25) - (4.2 \times 10^{-5})(T - 25)^2$ and $b_C(T) = -0.26 + (2.85 \times 10^{-4})(T - 25) + (4.31 \times 10^{-5})(T - 25)^2$. The hydration of the hydroxyl group in aliphatic hydroxyl side chains (Ser and Thr) appears to contribute positively and not negatively to ΔC_p [0.17 cal/(K·mol) Å^2 at 25 °C] (Gomez & Freire, 1995; Habermann & Murphy, 1996). In the equation above, ΔASA changes are in Å^2 and the heat capacity in cal/(K·mol). In general, for low-temperature calculations ($T <$ 80 °C), the temperature-independent coefficients are sufficient. Other contributions to ΔC_p are significantly smaller in magnitude and amount to less than 10% of the total change upon unfolding (Gomez et al., 1995).

The bulk of the enthalpy change also scales in terms of ΔASA changes, and at the reference temperature of 60 °C, it can be written as

$$\Delta H_{gen}(60) = a_H(60)\cdot\Delta ASA_{ap} + b_H(60)\cdot\Delta ASA_{pol} \quad (5)$$

where $a_H(60) = -8.44$ and $b_H(60) = 31.4$ (Hilser et al., 1996; Xie & Freire, 1994a,b). At 60 °C (the median denaturation temperature of proteins), ΔH_{gen} contributes over 90% of the total enthalpy change. The remaining contributions include protonation and other electrostatic effects, specific ligands or other interactions that need to be calculated on an individual basis (Hilser et al., 1996). It must be emphasized that the numerical values of the coefficients in eqs 4 and 5 depend on the method selected to determine ASA values. In our work, we have employed the Lee and Richards algorithm (Lee & Richards, 1971) as implemented by S. Presnell (University of California, San Francisco) using a probe radius of 1.4 Å and a slice width of 0.25 Å.

In the calculation of the entropy change, two primary contributions are included: one due to changes in solvation and the other due to changes in conformational degrees of freedom ($\Delta S = \Delta S_{solv} + \Delta S_{conf}$). The entropy of solvation can be written in terms of the heat capacity if the temperatures at which the apolar and polar hydration entropies are zero ($T^*_{S,ap}$ and $T^*_{S,pol}$) are used as reference temperatures:

$$\Delta S_{solv} = \Delta S_{solv,ap} + \Delta S_{solv,pol} \quad (6a)$$

$$\Delta S_{solv} = \Delta C_{p,ap} \ln (T/T^*_{S,ap}) + \Delta C_{p,pol} \ln (T/T^*_{S,pol}) \quad (6b)$$

$T^*_{S,ap}$ has been known to be equal to 385.15 K for some time (Baldwin, 1986), and $T^*_{S,pol}$ has been recently estimated to be close to 335.15 K for proteins (DAquino et al., 1996). In the above parameterization, the coefficients have been optimized by fitting the protein database rather than by scaling up thermodynamic data obtained from small organic molecules. In that sense, the main assumption is not group additivity but structural regularity between proteins [see Hilser et al. (1996)].

Conformational entropies are evaluated by explicitly considering the following three contributions for each amino acid: (1) $\Delta S_{bu\rightarrow ex}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; (2) $\Delta S_{ex\rightarrow u}$, the entropy change gained by a surface-exposed side chain when the peptide backbone unfolds; and (3) ΔS_{bb}, the entropy change gained by the backbone itself upon unfolding. The magnitude of these terms for each amino acid has been estimated by computational analysis of the probability of different conformers as a function of the dihedral and torsional angles (DAquino et al., 1996; Lee et al., 1994). Table 1 summarizes the conformational entropy values published by D'Aquino et al. (1996).

Table 1: Conformational Entropies for Amino Acids[a]

amino acid	$\Delta S_{bu\rightarrow ex}$ [cal/(K·mol)]	$\Delta S_{ex\rightarrow u}$ [cal/(K·mol)]	ΔS_{bb} [cal/(K·mol)]
Ala	0.00	0.00	4.1
Arg	7.11	−0.84	3.4
Asn	3.29	2.24	3.4
Asp	2.00	2.16	3.4
Cys	3.55	0.61	3.4
Gln	5.02	2.12	3.4
Glu	3.53	2.27	3.4
Gly	0.00	0.00	6.5
His	3.44	0.79	3.4
Ile	1.74	0.67	2.18
Leu	1.63	0.25	3.4
Lys	5.86	1.02	3.4
Met	4.55	0.58	3.4
Phe	1.40	2.89	3.4
Ser	3.68	0.55	3.4
Thr	3.31	0.48	3.4
Trp	2.74	1.15	3.4
Tyr	2.78	3.12	3.4
Val	0.12	1.29	2.18

[a] Reproduced from D'Aquino et al. (1996). $\Delta S_{bu\rightarrow ex}$ and $\Delta S_{ex\rightarrow u}$ are from Lee et al. (1994). For the purpose of approximating ΔS_{bb}, the amino acids have been divided into four groups: (1) no β-carbon (Gly); (2) β-carbon only (Ala); (3) β branching (Val and Ile); and (4) γ-carbon (remaining amino acids).

Average Solvent-Accessible Surface Areas. The above results underscore the importance of accurate calculations of changes in solvent accessibilities since many of the thermodynamic parameters are expressed in terms of changes in apolar and polar solvent-accessible surface areas. Usually, solvent accessibilities are estimated by using crystallographic or NMR structures for native conformations, and computer-built extended peptides for the unfolded state. One common characteristic of these computational methods is that they utilize a single static structure. In contrast, the solvent accessibility of a protein in solution is expected to vary with the rotation of side chains, and in unfolded conformations with backbone rotations also. These variations in solvent accessibility need to be assessed for a quantitative estimation of their effects on protein energetics.

Under equilibrium conditions, the solvent-accessible surface area of an amino acid in a polypeptide chain is not a constant but an average quantity to which all accessible conformations contribute according to their probability. In general, the probability of any side chain conformation, $P(\phi,\psi,\chi)$, is a function of the side chain dihedrals (χ) and backbone torsional angles (ϕ,ψ) and is given by the equation:

$$P(\phi,\psi,\chi) = \frac{\exp[-E(\phi,\psi,\chi)/kT]}{\sum\exp[-E(\phi,\psi,\chi)/kT]} \qquad (7)$$

where k is the Boltzmann constant and T the absolute temperature. The summation in the denominator in eq 7 defines the conformational partition function. The Boltzmann average accessible surface area, $\langle$ASA$\rangle$, is equal to

$$\langle \mathrm{ASA} \rangle = \frac{\sum \mathrm{ASA}(\phi,\psi,\chi)\cdot\exp[-E(\phi,\psi,\chi)]/kT}{\sum\exp[-E(\phi,\psi,\chi)]/kT} \qquad (8)$$

and can be estimated if the energy profile for each amino acid is known as a function of the side chain dihedrals and backbone torsional angles.

Average Solvent-Accessible Surface Areas in the α-Helix Conformation. Previously, the energy profiles for the side chain of each amino acid in an α-helix conformation were calculated as a function of the χ_1 and χ_2 side chain dihedrals (Lee et al., 1994). In order to obtain the Boltzmann ASA average for each type of amino acid in a helical conformation, helices of the form $(Ala)_4$-Xaa-$(Ala)_4$ were built as described before (Lee et al., 1994). Accessibility profiles for the central residue were constructed by varying the χ_1 and χ_2 angles for each amino acid every 10°. Dihedrals further than χ_2 were not considered. For those side chains with a single dihedral (Cys and Val), the value of χ_1 was varied every degree between 0° and 360°. The energy values for each conformation were those calculated previously (Lee et al., 1994) using the potential functions and parameters proposed by Weiner et al. (1984) and by Jorgensen and Tirado-Rives (1988). Accessible surface areas were calculated according to the Lee and Richards algorithm (Lee & Richards, 1971) using a probe radius of 1.4 Å and a slice width of 0.25 Å as described before (Murphy et al., 1992). Figure 1 shows the energy, probability, and solvent accessibility profiles for the case of phenylalanine. Boltzmann weighted averages for each amino acid were obtained according to eq 8.

Figure 2 shows the difference between the Boltzmann weighted-average ASA values and the standard numerical average for 19 amino acids in an α-helix conformation. For comparison, the ASA values at the energy minimum are also shown. It is clear that the Boltzmann averages and the ASA values obtained at the energy minima are very close. These values are also close to those obtained by a simple numerical average indicating that for those conformations that are not forbidden by excluded volume the variation in solvent accessibility is not very large. Similar results were obtained if the procedure was applied to residue 44 in the 39/50 helix of lysozyme T4. In this case, the sequence of the helix is LeuAsnAlaAlaLysXaaGluLeuAspLysAlaIle. Of course, the overall ASA changes and stability of this helix are different, but position 44 is solvent-exposed and the relative differences at that position are comparable. This location has been mutated with all 20 amino acids, and high-resolution structures are available for some of them (Blaber et al., 1994).

Solvent-Accessible Surface Areas for Unfolded Conformations. The solvent accessibility for the central amino acid in tripeptides of the form Ala-X-Ala was calculated as a function of the torsional angles ϕ and ψ of the central residue. All main chain dihedral angles, ϕ and ψ, were set initially to the fully extended conformation. Side chain angles were set to the all-trans conformation. Accessibilities of the central residue were calculated for all combinations of the main chain dihedrals between 0° and 360° every 10°. Not every conformation generated in this way is feasible due to steric hindrance effects imposed by both the backbone and Ala side chains that limit the allowable conformational space. Those conformations that produced van der Waals collisions were rejected. The allowed regions are sensitive to the van der Waals radii assigned to each type of atom. In these calculations, the set of van der Waals radii given by Leach et al. (1966) were used.

The results of the calculations are shown in Figure 3. In this case, the standard deviations in ASA values are very small, averaging 15% and 16% for the apolar and polar surfaces, respectively. Since the variation is small, a decision was made to refine these values by applying the structural

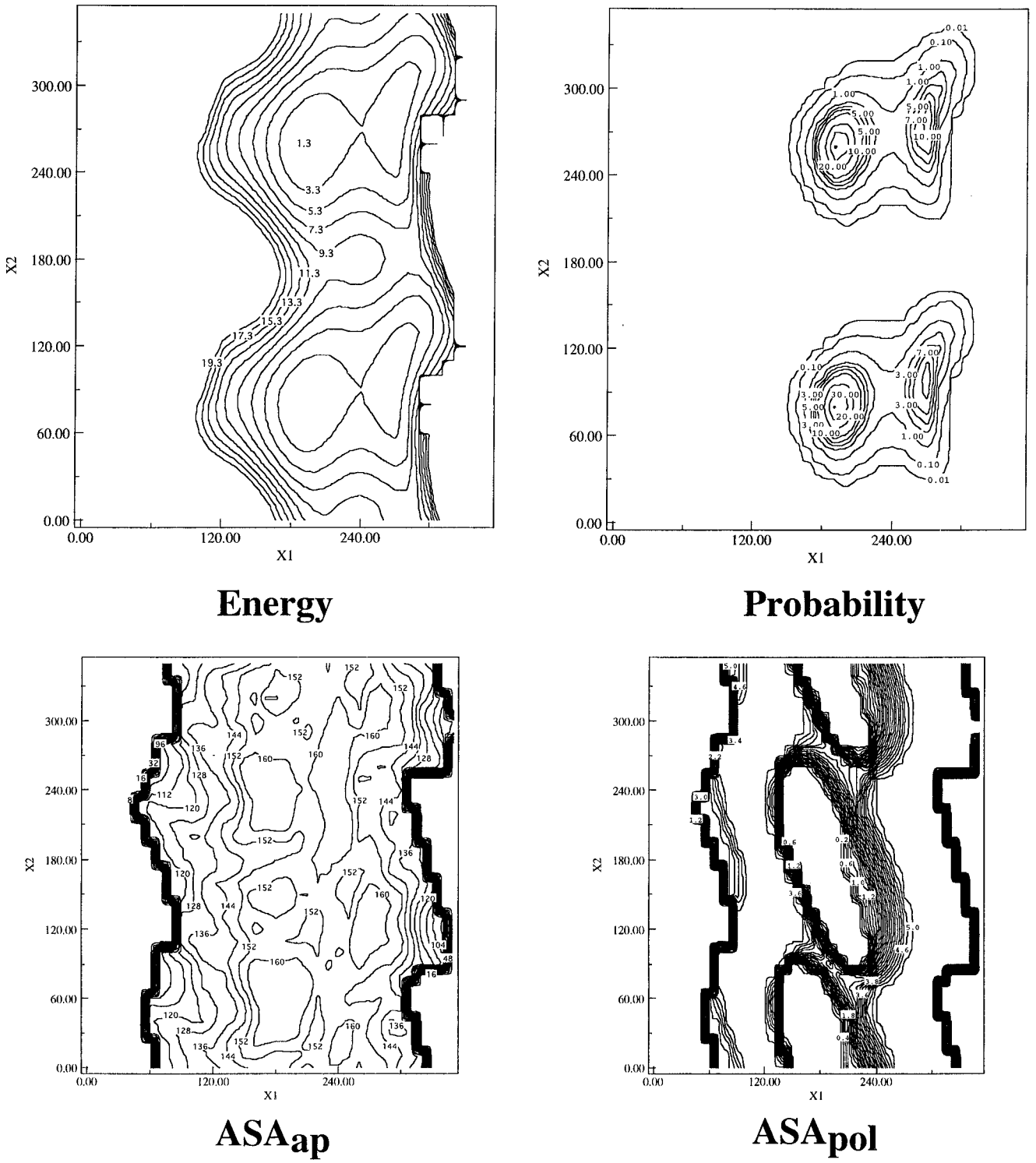

FIGURE 1: Schematic illustration of the procedure utilized for calculating average accessible surface areas. In this figure, the side chain energy profile for Phe as a function of the χ_1 and χ_2 dihedral angles is shown in the top left panel. The corresponding probability profile [$\exp(-E/kT)$] is shown in the top right panel. The bottom panels represent the apolar and polar accessible surface areas as a function of the χ_1 and χ_2 dihedral angles. The average accessible surface areas are obtained by a Boltzmann weighted average of the accessible surface area obtained as a function of the χ_1 and χ_2 dihedral angles using eq 8.

parameterization of the energetics to experimental systems in which $\Delta\Delta G$ values have been measured for all or a significant number of amino acids. The goal was 2-fold: (1) to check whether the experimental $\Delta\Delta G$ values can be correctly predicted by the structural parameterization using a unique set of ASA values for the amino acids in the unfolded state; and (2) to check whether the ASA values that account for the experimental data lie within the allowed set of values. For this analysis, the experimental $\Delta\Delta G$ values measured for residue 44 in T4 lysozyme (Blaber et al., 1994), residue 32 of barnase (Horovitz et al., 1992), a synthetic leucine zipper of 29 amino acids (O'Neil & DeGrado, 1990), and a monomeric synthetic peptide of 21 amino acids (Lyu et al., 1990) were used. In all four cases, the mutated amino acids are at a solvent-exposed location, thus minimizing the effects of the mutations on internal interactions in the native state and maximizing the effects of changes in solvent accessibilities in the unfolded state.

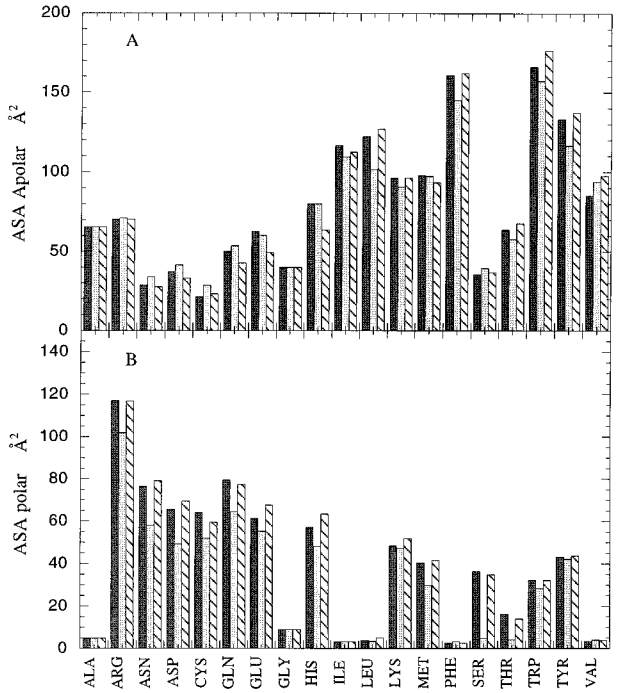

FIGURE 2: Apolar (panel A) and polar (panel B) accessible surface areas for 19 amino acids in the α-helix conformation. Shown are the Boltzmann weighted-average ASA (dark bars), the simple numerical average ASA over all χ_1 and χ_2 dihedral angles (dotted bars), and the ASA at the minimum of the potential well (dashed bars).

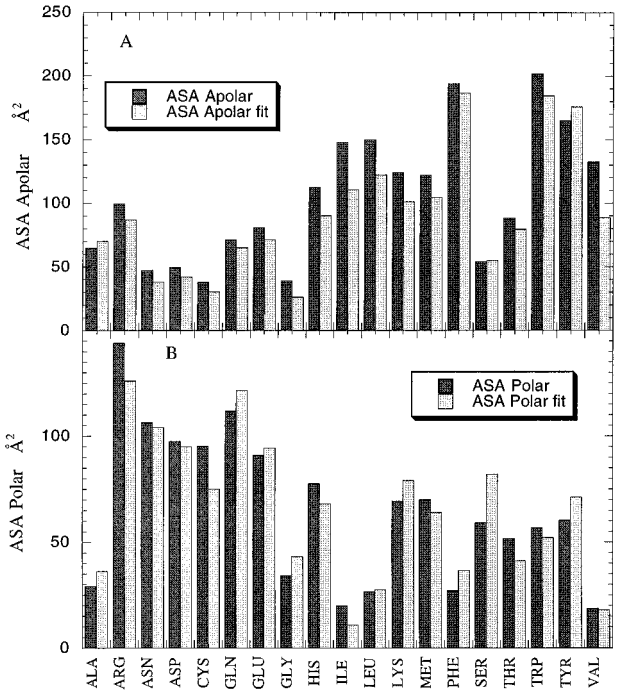

FIGURE 3: Apolar (panel A) and polar (panel B) accessible surface areas for 19 amino acids in the unfolded state. Shown are the numerical averages over the torsional angles ϕ and ψ and the best ASA values obtained from the global fit of the thermodynamic data for α-helix propensities in different experimental systems.

The free energy change for each amino acid mutation was calculated according to the structural parameterization described above, using the same set of elementary coefficients in all calculations. For each amino acid, the ASA_{ap} and ASA_{pol} values in the native state were calculated from the appropriate helix structure. Only the unfolded conformation, and consequently the ASA_{ap} and ASA_{pol} values of the amino acid in the unfolded state, was allowed to vary within the permitted structural range. The unfolded ASA values that minimized the difference between predicted and experimental $\Delta\Delta G$ values were identified. It must be noted that the four sets of data represent not only different systems but also different experimental temperatures. The data for T4 lysozyme are at 52 °C, the data for barnase are at 25 °C, the data for the leucine zipper are at 20 °C, and the data for the monomeric peptide are at 4 °C. Table 2 summarizes the differences in $\Delta\Delta G$ for all amino acids. In all cases, the agreement between experimental and calculated values was excellent. The absolute difference between calculated and experimental $\Delta\Delta G$ values is 0.09 kcal/mol for T4 lysozyme, 0.14 kcal/mol for barnase, 0.11 kcal/mol for the synthetic leucine zipper, and 0.08 kcal/mol for the synthetic peptide.

Table 2: Experimental and Predicted $\Delta\Delta G$ for Specific Amino Acid Mutations[a]

amino acid	$\Delta\Delta G_{calc} - \Delta\Delta G_{exp}$ (cal/mol) barnase	leucine zipper	T4−44	peptide
Ala	−94.76	22.01	−31.87	−77.45
Arg	−85.40	−37.43	112.74	−
Asn	42.86	203.73	−10.94	25.50
Asp	135.55	169.08	−8.96	−
Cys	163.20	−6.5	−29.3	−
Gln	80.94	161.56	−200.06	−56.88
Glu	78.66	151.37	−12.20	−
Gly	0.00	0.00	0.00	0.00
His	69.66	110.29	−235.30	−
Ile	241.0	84.1	−367.04	151.11
Leu	103.05	18.84	−139.1	116.36
Lys	−19.40	−551.35	76.08	−
Met	32.55	106.25	−102.66	−54.76
Phe	246.09	31.41	−15.89	−
Ser	−140.58	−42.68	82.30	−151.05
Thr	140.09	94.61	−6.73	−190.33
Trp	400.64	−10.59	33.59	−
Tyr	277.83	163.91	−206.55	−
Val	432.63	294.08	−31.33	−4.71

[a] $\Delta\Delta G$ values were evaluated at the experimental temperatures: T4 lysozyme, 52 °C (Blaber et al., 1994); barnase, 25 °C (Horovitz, et al., 1992); leucine zipper, 20 °C (O'Neil et al., 1990); and monomeric peptide, 4 °C (Lyu et al., 1990).

Figure 4 (panels A−D) shows the results of a correlation analysis between calculated and experimental $\Delta\Delta G$ values relative to glycine. For all data sets, the slopes vary between 0.8 and 1.27, averaging 1.03 ± 0.2, indicating the absence of systematic deviations that preferentially overestimate or underestimate the experimental values. The correlation coefficients vary between 0.83 and 0.91, which is within the same range reported before for other analysis reported in the literature (Muñoz & Serrano, 1995a).

The best set of unfolded ASA values is shown in Figure 3 (panels A and B). In general, the apolar accessible surface areas are smaller than those obtained by a simple numerical average of the allowed conformations. This result reflects the tendency of apolar groups to shield from water and expose to the solvent an area smaller than that corresponding to an extended conformation. This tendency is not as significant for the polar accessible surface areas. On average, the resulting values are about 85% of those corresponding to the numerical average, in agreement with the results of Creamer et al. (1995). These values are also close to those recently reported by Wimley et al. (1996). Table 3 sum-

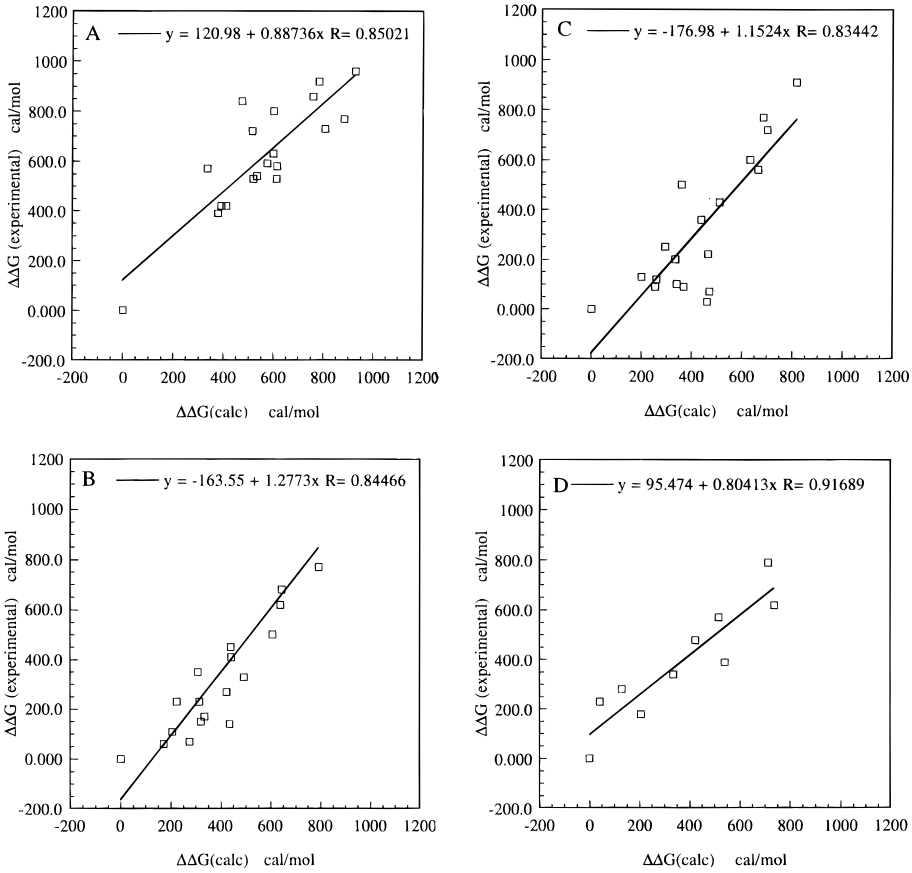

FIGURE 4: Correlation analysis between calculated and experimental $\Delta\Delta G$ relative to Gly for the α-helix propensities of all amino acids (panels A−D). The calculated values are at the corresponding experimental temperature. The data for T4 lysozyme (panel A) are at 52 °C (Blaber et al., 1994), the data for the leucine zipper (panel B) are at 20 °C (O'Neil et al., 1990), the data for barnase (panel C) are at 25 °C (Horovitz et al., 1992), and the data for the monomeric peptide (panel D) are at 4 °C (Lyu et al., 1990). The absolute difference between calculated and experimental ΔG values is 0.09 kcal/mol for T4 lysozyme, 0.14 kcal/mol for barnase, 0.11 kcal/mol for the synthetic leucine zipper, and 0.08 kcal/mol for the synthetic peptide.

Table 3: ASA Values for Amino Acids in Unfolded Conformations[a]

amino acid	ASA_{ap}	$ASA_{ap,num}$	ASA_{pol}	$ASA_{pol,num}$
Ala	70.0	64.7	36.1	29.2
Arg	87.1	99.5	126.1	143.9
Asn	38.1	47.2	104.0	106.6
Asp	42.1	49.7	95.0	97.6
Cys	30.3	38.1	75.1	95.3
Gln	60.5	71.2	112.5	111.8
Glu	71.1	80.9	94.3	91.1
Gly	26.2	38.9	43.1	34.0
His	90.0	112.4	68.0	77.6
Ile	110.7	148.2	10.9	19.8
Leu	122.3	149.9	27.5	26.6
Lys	101.3	124.1	79.0	69.4
Met	104.6	122.2	64.0	70.1
Phe	186.8	194.0	36.5	27.0
Ser	55.1	54.3	81.9	59.2
Thr	79.5	88.3	41.1	51.8
Trp	184.5	201.9	52.3	56.8
Tyr	175.8	165.1	71.1	60.3
Val	88.7	132.4	17.8	18.7

[a] ASA_{ap} and ASA_{pol} are the best apolar and polar ASA values obtained from the global fit of the α-helix propensity data. $ASA_{ap,num}$ and $ASA_{pol,num}$ are the values obtained by numerical averaging the apolar and polar area values over the main chain torsional angles.

marizes the set of average ASA values for the unfolded conformation that satisfactorily account for the α-helix propensities of all amino acids.

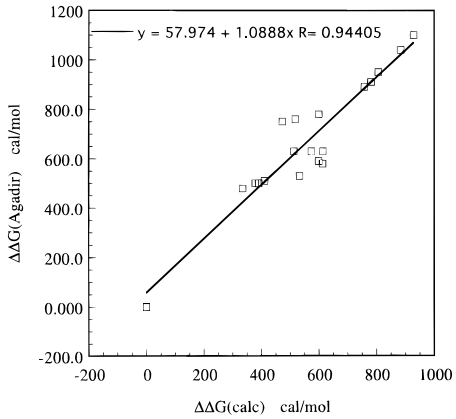

FIGURE 5: Correlation between $\Delta\Delta G$ values calculated with the structural parameterization of the energetics ($\Delta\Delta G_{calc}$) and those calculated with AGADIR (Muñoz & Serrano, 1995a).

Comparison with Other Approaches. The best correlation was observed between the structure-based thermodynamic analysis presented here and the AGADIR algorithm developed by Muñoz and Serrano (1995a,b). Depending on the exact temperature, the correlation coefficient between the two treatments ranged between 0.9 and 0.95, and the slope was very close to 1, as illustrated in Figure 5. The average difference between the structural parameterization and AGADIR was 0.1 ± 0.08 kcal/mol without any single difference exceeding 0.27 kcal/mol. The excellent agreement between these two methods is remarkable, especially if one considers

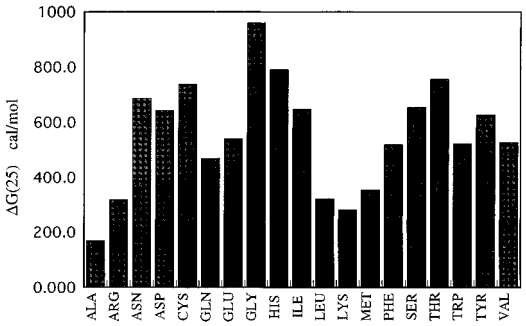

FIGURE 6: Helix propensities of 19 amino acids. Shown in the figure are the calculated ΔG values for helix formation at 25 °C.

the different nature of the approaches used in their development.

The structural parameterization of the energetics developed in this laboratory is of a general thermodynamic nature and aims at predicting the Gibbs energy of a specific peptide or protein conformation from its structure. It is not based on structural considerations specific to α-helix formation and was not developed with the purpose of creating a helix propensity scale. The AGADIR algorithm, on the other hand, is based upon helix−coil transition theory and explicitly considers specific interactions occurring in α-helices devoid of tertiary structure interactions. In their work, Muñoz and Serrano (1995a,b) have considered the contributions of the main chain hydrogen bond, intrinsic helical tendencies, capping interactions, charge-dipole interactions, and side-chain to side-chain interactions. The quantitative agreement observed between the two algorithms is an indication that the structural parameterization effectively captures the interactions existing in helix formation, albeit in an implicit fashion, and provides a thermodynamic validation to AGADIR.

The α-helix propensity scale developed in this paper correlates very well with the empirical scales shown in Figure 4. The average slope of the correlation is 1.03, suggesting that systematic deviations are absent. The correlation coefficient with the values obtained by Chakrabartty et al. (1994) for alanine-based peptides is 0.84, which is also within the observed range. The slope, however, is close to 1.8, which can be partly attributed to the larger difference between Gly and Ala reported for the alanine-based peptides.

Free Energy of Helix Formation. The results presented above represent the relative average differences in free energy with respect to Gly for the formation of a solvent-exposed helix. Calculation of helix stabilities, however, requires ΔG values and not only $\Delta\Delta G$ values. In this respect, Lyu et al. (1990) have estimated ΔG for Gly at 4 °C as 0.47 kcal/mol. The value reported by Baldwin (1995) for Gly at 0 °C is 1.6 kcal/mol. At the same temperatures, the ΔG value predicted by the structural parameterization is 0.76−0.80 kcal/mol, which is within the experimental range. The calculated ΔG values for 19 amino acids are summarized in Figure 6. The ΔG values for all amino acids are bracketed within 0.8 kcal/mol with Ala being the most stable and Gly the least stable helix former. The difference of 0.8 kcal/mol between Ala and Gly is mostly due to the difference in backbone entropies in the unfolded state, that amounts to 0.7 kcal/mol (see Table 1).

The enthalpy change is sensitive to the context of a particular residue in the sense that changes in hydration due to different solvent exposures as well as intrinsic interactions play a major role. The structural parameterization implicitly incorporates both contributions (Hilser et al., 1996). The enthalpy change needs to be calculated on a case by case basis because these contributions have opposite signs and because they are also accompanied by compensating enthalpy−entropy effects. The desolvation of both polar and nonpolar groups is known to exhibit a positive enthalpy change (Cabani et al., 1981; Murphy & Gill, 1989) and is also accompanied by an entropy term (eq 6). On average, for the formation of a helix in the absence of tertiary interactions, the enthalpy change predicted by the structural parameterization is about −0.7 kcal/mol residue at 25 °C. This value is similar to the one derived by Muñoz and Serrano (1995b). At 25 °C, the predicted enthalpy change is somewhat larger for Gly (∼−1.7 kcal/mol) and Ala (∼−1.1 kcal/mol) since these two amino acids lack side chains that become buried or partially buried from the solvent upon helix formation. The larger value observed for Gly is due to the differences in solvent accessibility created by the presence of the β carbon. This difference in enthalpy between Ala and Gly is almost completely compensated by the difference in solvation entropy [$T\cdot\Delta\Delta S_{solv} = 0.65$ kcal/(K·mol)]. For amino acids other than Gly and Ala, helix formation is also accompanied by a larger partial burial of the side chain from water and therefore by the presence of an additional small positive contribution to ΔH and a compensating ΔS. For Ala, the only amino acid for which direct calorimetric data are available, the measured enthalpy value ranges between −1.0 and −1.3 kcal/mol (Baldwin, 1995; Scholtz et al., 1991), which is very close to the predicted value.

Helix formation is opposed by a loss in conformational entropy of backbone and side chains (Table 1). This conformational entropy loss is not the same for all amino acids, and its magnitude contributes to a large extent to define the α-helix propensity scale (Creamer & Rose, 1992). The difference in helix propensity between Ala and Gly, for example, can be almost entirely attributed to differences in conformational entropy. For other amino acids, the conformational entropy loss is also partially compensated by a positive solvent-related entropy associated with the burial from the solvent of portions of the side chain as mentioned above. Overall, the entropy change for helix formation is negative and therefore unfavorable.

CONCLUSIONS

The excellent agreement observed between predicted and experimental ΔG values for all amino acids validates the use of the structural parameterization in free energy calculations of protein folding or binding. In addition, these results allow rationalization of the origin of the different α-helix propensities of amino acids. According to the results presented here, several factors contribute to the observed α-helix propensities. In general, α-helix formation is enthalpically favored and entropically unfavored. For Ala and Gly, the enthalpy change is on the order of −1.3 kcal/mol and originates from the formation of the hydrogen bond and the simultaneous burial from the solvent of a sizeable portion of the backbone. For other amino acids, the enthalpy change is smaller due to the additional positive enthalpic contribution associated with the burial from the solvent of portions of the side chain. The difference in helix propensity between

Ala and Gly is due primarily to the entropic difference arising from the larger conformational entropy of Gly in the unfolded state. The presence of the β-carbon in Ala and all other amino acids induces a large decrease in the conformational entropy of the unfolded state (2.4 eu) (DAquino et al., 1996; Leach et al., 1966). Except for β branching (Val, Ile), the presence of additional groups does not induce further major restrictions in the conformational entropy of the unfolded state (DAquino et al., 1996; Leach et al., 1966). For amino acids other than Gly and Ala, the side chain loses conformational entropy in the helical conformation relative to the unfolded state (Creamer & Rose, 1992; DAquino et al., 1996; Lee et al., 1994). These side chain conformational restrictions lower their helical propensities relative to Ala. Side chain entropy, however, is not sufficient to account completely for the observed propensity values. In addition, for amino acids other than Gly and Ala, the burial of nonpolar portions of the side chain from the solvent provides a compensating favorable entropy contribution that partially compensates the unfavorable desolvation enthalpy. Thus, the observed α-helix propensities are the result of a combination of different enthalpy, conformational entropy, and solvation entropy contributions.

ACKNOWLEDGMENT

We thank Mario Amzel and George Rose for many helpful discussions and the reviewers of the manuscript for helpful suggestions, especially the correlation with AGADIR.

REFERENCES

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069−8072.
Baldwin, R. L. (1995) *Biophys. Chem. 55*, 127−135.
Blaber, M., Zhang, X.-j., Lindstrom, J. L., Pepiot, S. D., Baase, W. A., & Matthews, B. W. (1994) *J. Mol. Biol. 235*, 600−624.
Cabani, S., Gianni, P., Mollica, V., & Lepori, L. (1981) *J. Solution Chem. 10*, 563−595.
Chakrabartty, V., Kortemme, T., & Baldwin, R. L. (1994) *Protein Sci. 3*, 843−852.
Connolly, M. L. (1983) *Science 221*, 709−713.
Creamer, T. P., & Rose, G. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5937−5941.
Creamer, T. P., Srinivasan, R., & Rose, G. D. (1995) *Biochemistry 34*, 16245−16250.
DAquino, J. A., Gómez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., & Freire, E. (1996) *Proteins: Struct., Funct., Genet. 25*, 143−156.
Freire, E. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 141−165.
Gomez, J., & Freire, E. (1995) *J. Mol. Biol. 252*, 337−350.
Gomez, J., Hilser, J. V., Xie, D., & Freire, E. (1995) *Proteins: Struct., Funct., Genet. 22*, 404−412.
Habermann, S. M., & Murphy, K. P. (1996) *Protein Sci. 5*, 1229−1239.
Hilser, V. J., Gomez, J., & Freire, E. (1996) *Proteins: Struct., Funct., Genet.* (in press).
Horovitz, A., Matthews, J. M., & Fersht, A. R. (1992) *J. Mol. Biol. 227*, 560−568.
Jorgensen, W. L., & Tirado-Rives, J. (1988) *J. Am. Chem. Soc. 110*, 1657−1666.
Leach, S. J., Nemethy, G., & Scheraga, H. A. (1966) *Biopolymers 4*, 369−407.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
Lee, K. H., Xie, D., Freire, E., & Amzel, L. M. (1994) *Proteins: Struct., Funct., Genet. 20*, 68−84.
Lyu, P. C., Liff, M. I., Marky, L. A., & Kallenbach, N. R. (1990) *Science 250*, 669−673.
Muñoz, V., & Serrano, L. (1995a) *J. Mol. Biol. 245*, 275−296.
Muñoz, V., & Serrano, L. (1995b) *J. Mol. Biol. 245*, 297−308.
Murphy, K. P., & Gill, S. J. (1989) *Thermochim. Acta 139*, 279−290.
Murphy, K. P., & Freire, E. (1992) *Adv. Protein Chem. 43*, 313−361.
Murphy, K. P., Bhakuni, V., Xie, D., & Freire, E. (1992) *J. Mol. Biol. 227*, 293−306.
Murphy, K. P., Xie, D., Garcia, K. C., Amzel, L. M., & Freire, E. (1993) *Proteins: Struct., Funct., Genet. 15*, 113−120.
O'Neil, K., & DeGrado, W. (1990) *Science 250*, 646−651.
Ooi, T., Oobatake, M., Némethy, G., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 3086−3090.
Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M., & Bolen, D. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2854−2858.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351−372.
Spolar, R. S., & Record, M. T. (1994) *Science 263*, 777−784.
Spolar, R. S., Livingstone, J. R., & Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., & Weiner, P. A. (1984) *J. Am. Chem. Soc. 106*, 765−784.
Wimley, W. C., Creamer, T. P., & White, S. H. (1996) *Biochemistry 35*, 5109−5124.
Xie, D., & Freire, E. (1994a) *Proteins: Struct., Funct., Genet. 19*, 291−301.
Xie, D., & Freire, E. (1994b) *J. Mol. Biol. 242*, 62−80.
Xie, D., Fox, R., & Freire, E. (1994) *Protein Sci. 3*, 2175−2184.

BI961319S