PROTEINS: Structure, Function, and Genetics 20:68–84 (1994)

Estimation of Changes in Side Chain Configurational Entropy in Binding and Folding: General Methods and Application to Helix Formation

Kon Ho Lee,[1,2] Dong Xie,[3] Ernesto Freire,[3] and L. Mario Amzel[1]
Departments of [1]Biophysics and Biophysical Chemistry, School of Medicine, [2]Physics and Astronomy, and [3]Biology and Biocalorimetry Center, Johns Hopkins University, Baltimore, Maryland 21218

ABSTRACT **Theoretical estimations of changes in side chain configurational entropy are essential for understanding the different contributions to the overall thermodynamic behavior of important biological processes like folding and binding. The configurational entropy of any given side chain in any particular protein can be evaluated from the complete energy profile of the side chain. Calculations of the energy profiles can be performed using the side chain single bond dihedrals as the only independent variables as long as the structures at each value of the dihedrals are allowed to relax through small changes in the valence bond angles. The probabilities of different side chain conformers obtained from these energy profiles are very similar to the conformer populations obtained by analysis of side chain preferences in the proteins of the Protein Data Bank. Also, side chain conformational entropies obtained from the energy profiles agree extremely well with those obtained from the Protein Data Bank conformer populations. Changes in side chain configurational entropy in binding and folding can be computed as differences in conformational entropy because, in most cases, the frequency of the rotational oscillation around the energy minimum of any given conformer does not appear to change significantly in the reaction. Changes of side chain conformational entropy calculated in this way were compared with experimental values. The only available experimental data—the effect of side chain substitution on the stability of α-helices—were used for this comparison. The experimental values were corrected to subtract the solvent contributions. This comparison yields an excellent agreement between calculated and experimental values, validating not only the theoretical estimates but also the separability of the entropic contributions into configurational terms and solvation related terms.**

Key words: side chain conformation, protein folding, protein binding, helix formation, helix stability

INTRODUCTION

One of the major goals of structural energetics is the development of the ability to calculate association constants for binding reactions based on the three-dimensional structures of the complexes and of the free molecules. The constants can be calculated from the change in standard molar free energy in the reaction using the relation

$$K = \exp(-\Delta G^0/RT).$$

The change in free energy can in turn be estimated from the enthalpy and the entropy of binding

$$\Delta G = \Delta H - T\Delta S.$$

The same type of calculations can be used to estimate the free energy of protein folding/unfolding. The enthalpy and the entropy of binding both involve contributions from the reacting molecules as well as contributions from the solvent. In the case of the enthalpy, both contributions have been parameterized as a function of the change in exposed area before and after the reaction and, for the case of protein folding, calculations based on the proposed parameters give good agreement with experimental data (for a review, see Murphy and Freire[1]). In the case of the entropy, effective parameterization requires separation of the contributions into two terms

$$\Delta S = \Delta S_{solvent} + \Delta S_{configuration}.$$

The solvent contribution is primarily a consequence of the hydrophobic effect and to a lesser extent of the hydration of polar groups. Solvent contributions can be effectively parameterized as a function of the change of exposed polar and apolar areas. The change in configurational entropy, on the other hand, is refractory to this type of empirical parameterization and has to be estimated independently. As things stand, estimation of the change in configurational entropy is a major roadblock in the calcu-

Received March 17, 1994; revision accepted May 9, 1994.
Address reprint requests to L. Mario Amzel, Department of Biophysics and Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, MD 21205.

lation of free energies of binding, as well as in the calculation of the free energies of protein folding/unfolding.

Accurate estimation of configurational entropy changes requires that configurational entropies be calculated for all single bonds in the structures before and after the reaction (binding or folding/unfolding). Such calculations can only be performed if one can develop an accurate, efficient, deterministic algorithm that permits the calculation of configurational entropies based on the atomic coordinates for all molecules involved, each time a new reaction is studied. In this paper we describe such an algorithm and provide several ways to perform the calculations (depending on the required accuracy). Although the method is completely general and can be used to calculate the configurational entropy of any side chain in any structure, for the tests in this paper we calculated the configurational entropies for all amino acid side chains at the fifth position of a nine residue helix. A similar calculation using Monte Carlo sampling has been carried out by Creamer and Rose.[2] (A method that can be used with molecular dynamics, harmonic analysis, or Monte Carlo simulations has been proposed by Karplus and Kushick.[3] This method is computationally very intensive and does not give a direct separation between conformational and kinetic contributions to the configurational entropy.)

To test the accuracy of the calculations, as well as the validity of many of the assumptions, we attempted to extract values of the side chain configurational entropy from experimental data to compare with our calculated values. Measurements of the effects of individual side chains on the stability of α-helices provide the only available data relevant to these calculations. During the past few years, several research groups have estimated relative free energy changes resulting from specific amino acid substitutions in exposed α-helices in proteins[4–6]. These changes in free energy can be used to estimate the changes in configurational entropy by subtracting all other contributions estimated on the basis of previously reported empirical parameterizations.[9] We report here such a calculation and compare the results with the changes in configurational entropy calculated using the methods proposed in this paper.

CALCULATION OF CHANGES IN THE CONFIGURATIONAL ENTROPY

General Considerations

The change of side chain entropy in a binding (or folding) process can be calculated as the difference in the configurational entropy of all side chains between the initial and the final state. Entropies can be obtained (using statistical thermodynamics) from the Hamiltonian (energy), provided that it can be calculated at all points of configurational space. If this is possible, the partition function Z can be calculated using the expression

$$Z = 1/\sigma \mathbf{h}^f \int \exp(-E/RT)\, dp\, dq \qquad (1)$$

where the integration is performed over all accessible states and dp and dq are the generalized coordinates and conjugated momenta, respectively, σ is the product of the symmetry numbers of the f degrees of freedom, and **h** is the Plank constant. The entropy is then calculated with one of the following expressions:

$$S/R = -\Sigma\, p_i \ln p_i \qquad (2)$$

with

$$p_i = \exp(-E_i/RT)/Z$$

or

$$S/R = \langle E \rangle/RT + \ln Z \qquad (3)$$

with

$$\langle E \rangle = 1/(Z\, \sigma\, \mathbf{h}^f) \int E \exp(-E/RT)\, dp\, dq.$$

As mentioned before, accurate calculation of side chain entropies requires an evaluation of the energies of all conformations accessible to the side chain. This calculation must be performed for each particular situation because the energies depend on the structure of the entire protein that contains the residue being considered. The most important independent variables for the evaluation of side chain entropies are the rotations around the single bond dihedrals of the side chains. However, computation of the energy profiles using only these variables gives serious errors in the estimation of the energies of many side chain conformations. In particular, it exaggerates the number of unfavorable, high energy conformations. Some conformations that show unfavorable interactions at a given dihedral angle can relax the "bad" contacts through small changes in valence bond angles. Since small angle changes produce only small changes in the corresponding energy term, the total energy of the conformation at that dihedral angle value can actually be very low. (This statement is similar to the conclusions of Karplus and Kushick[3] about "important" internal coordinates.)

Calculation of Side Chain Energy Profiles

Structures

Most of the calculations were carried out using nine residue α-helices. The helices—of sequence $(Ala)_4$-Xaa-$(Ala)_4$—were constructed using the sequence builder in QUANTA (Molecular Simulations Inc., Palo Alt, CA) with main chain dihedral angles set to $\phi = -65°$ and $\psi = -40°$. The rest of the coordinates used in the calculations were obtained from the Protein Data Bank.

Energy calculations

Energies were calculated with the potential functions and parameters proposed by Weiner et al.[10] and by Jorgensen and Tirado-Rives.[11] Electrostatic contributions—calculated using a value of 78 for the dielectric constant—included partial charges. Dihedral angles were varied every 1° for χ_1 and every 10° for χ_2. For each conformation the structure was relaxed (taken to the equilibrium configuration) by minimizing the energy as a function of the valence bond angles θ_1 and θ_2 (θ_1: N-C_α-C_β; θ_2: C-C_α-C_β) while keeping single bond dihedrals constant. Minimization was carried out with a specially written code using conjugate gradients.

To test if the methods could be used with existing public domain software, the side chain energy profile of valine in an α-helix was calculated using the program XPLOR.[12] The side chain dihedral was changed every 5° and the energies of the resulting structures were minimized with XPLOR, selecting the coordinates of C_β, $C_{\gamma 1}$, and $C_{\gamma 2}$ as independent variables. During minimization, the side chain dihedral was constrained to each value of χ_1 (χ_i) by using the CONSTRAIN DIHEDRAL option of XPLOR with a constant of 400 kcal/mol Å^2. (This option adds to the total energy a quadratic constraint of the form $E_{\text{CONST-DIHEDRAL}} = 1/2\ K_x\ (\chi - \chi_i)^2$. After minimization, energies are calculated without this term.)

Calculation of Side Chain Entropies

Side chain entropies were calculated using expressions 2 and 3 with energies evaluated as described in the previous section. Three different methods were used:

Method 1: Use of expression 2 with conformer populations

In this procedure, expression 2 is used with the sum carried out over the appropriate number of conformers for each side chain dihedral. This provides the conformational entropy portion of the configurational entropy. Populations for dihedrals around single bonds between two sp^3 carbons are estimated for three conformers—*trans*, *gauche*$^+$, and *gauche*$^-$ (*trans*: 120° to 240°; *gauche*$^+$: 0° to −120°; *gauche*$^-$: 0° to 120°). For dihedrals around single bonds between an sp^3 and sp^2 atom, there are in principle six conformers that must be considered—$0° < \chi \leq 60°$, $60° < \chi \leq 120°$, $120° < \chi \leq 180°$, $180° < \chi \leq 240°$, $240° < \chi \leq 300°$, and $300° < \chi \leq 360°$. Since the selection of conformers affects the value of the entropy, care was taken to ensure that the conformers chosen corresponded to local minima in the energy profiles. Conformers centered at $\chi_2 + 90°$, −90°, 150°, and 330° covered all cases. If the planar group containing the sp^2 carbon is symmetrical (i.e., phenylalanine, aspartic, etc.) a symmetry number of 2 is used for the dihedral. Creamer and Rose[2] have used a method based on conformers with Monte Carlo sampling to calculate side chain entropies in a helix. In the present study the method is used by summing the individual probabilities of all conformations of the energy profile belonging to a given conformer.

Method 2: Addition of kinetic terms as harmonic oscillators

The energy at the bottom of each potential well can be approximated by a quadratic equation of the form

$$E = E_{min} + 1/2\ k\ (\chi - \chi_{min})^2.$$

After adjustment of the energy values to this equation, the constant k can be used to calculate the frequency of the rotational oscillator ω and with it θ_ω, the Einstein temperature of the oscillator. (The momenta of inertia necessary to obtain ω were calculated as described for Method 3.) S_i, the contribution of the kinetic energy term to the entropy of the *i*th well can then be calculated with the expression for the Einstein oscillator

$$S_i/R = [(\theta_{\omega i}/T)/\exp(\theta_{\omega i}/T) - 1] - \ln[1 - \exp(-\theta_{\omega i}/T)].$$

The vibrational entropy is then calculated as

$$S_{vib} = \Sigma\ w_i\ S_i$$

where $w_i = \exp(-E_{min,i}/RT)/\Sigma \exp(-E_{min,i}/RT)$ is the Boltzmann weighting factor for each well. The contributions to the entropy calculated in this manner are added to the conformational entropy calculated using Method 1 to provide the complete configurational entropy.

Method 3: Integration using the complete Hamiltonian

The Hamiltonian for the configurations around a dihedral angle is a rigid body rotation Hamiltonian with two terms, one corresponding to the kinetic and the other to the potential energy:

$$H = \Sigma\ L_j^2/2I_j + V(\{\chi_j\}) \qquad (4)$$

where $I_j = \Sigma\ m_i \cdot r_i^2$ is the moment of inertia, $L_j = I_{j\omega j}$ is the angular momentum, and V is the potential energy as a function of the dihedral angles χ_j. (The momenta of inertia are calculated only for the side chain rotations around the axis of the dihedral. It is assumed that the rest of the protein is much larger than the side chains and therefore the center of mass of the system is very close to that of the rest of the protein.[13] The sum is performed over all the single bond dihedral angles considered for the given side chain.

The partition function can be calculated by integration of expression 1 using expression 4 to calculate the potential energy

$$Z = 1/\sigma \int \exp(-[\Sigma L_j^2/2I_j + V(\{X\})])\{dL \cdot dX\}/h^f$$

where, as above, σ is the product of the symmetry numbers of the rotations and f is the number of dihedrals of the side chain of interest.

The kinetic energy term can be integrated analytically, while the potential energy term is integrated numerically using the interaction energies obtained as described above. The entropy can then be obtained from the average energy and the partition function using expression 3. The average energy is calculated as

$$\langle E \rangle = \langle K \rangle + \langle V \rangle = f\,RT/2 + \langle V \rangle$$

and the entropy is then

$$S/R = f/2 + \langle V \rangle/RT + 1n\ Z$$

with

$$\langle V \rangle = \int V \exp(-V/RT)dX / \int \exp(-V/RT)\,dX.$$

Calculation of Entropy Changes

The calculations by either of the three methods only need to be performed for those side chains that are involved in the process and only for those states whose energy changes during the process. For these calculations it is some times convenient to estimate the configurational entropy as the sum of two terms

$$S_{configurational} = S_{conformational} + S_{vibrational}. \quad (5)$$

(This separation is particularly useful if, for example, there is good reason to believe that there is no change in the vibrational states during the process.) Since

$$\Delta S_{configurational} = S_{configurational,bound} - S_{configurational,free} \quad (6)$$

using expression 5 one obtains

$$\Delta S_{configurational} = (S_{conf.,bound} + S_{vib.,bound}) - (S_{conf.,free} + S_{vib,free}) \quad (7)$$

and

$$\Delta S_{configurational} = (S_{conf.,bound} - S_{conf.,free}) + (S_{vib.,bound} - S_{vib,free}) \quad (8)$$

or

$$\Delta S_{configurational} = \Delta S_{conformational} + \Delta S_{vibrational}.$$

When using entropies calculated with Method 1 to estimate the side chain contribution to the ΔS of binding one is assuming that $S_{vib.}$ is the same in the free and the bound states ($\Delta S_{vibrational} = 0$), and that only one of the rotamers is populated in the bound state for the side chains that become buried. The entropy change is then

$$\Delta S = S_{conf.,bound} - S_{conf.,free}$$

where $S_{conf.,bound}$ is equal to zero (R 1n 1) and $S_{conf.,free}$ is calculated by Method 1.

If, on the other hand, the shapes of the energy wells change during the process, it is also necessary to calculate $S_{vib,bound}$ and $S_{vib,free}$ as described above. The entropy is then calculated using expression 8. In the case of folding, the change in side chain configurational entropy is calculated as the difference between the folded and the unfolded state. In the thermodynamic expressions for folding the standard convention of using the native state as the reference state will be used. Therefore

$$\Delta S = S_{conf.,unfold.} - S_{conf.,fold.}$$

The change in side chain entropy is then the result of the restrictions imposed by the folded protein on the side chain conformations that were accessible in the unfolded state.

Calculations with Method 3 use expression 6 directly, with entropies calculated using the potential energy profiles computed for the complex and the free molecules, respectively.

Semiempirical Evaluation of Configurational Entropy Changes

The theoretically calculated side chain entropies discussed above can be compared with values derived from existing thermodynamic data on exposed α-helices in proteins[4–6] or model peptides.[7,8] At any temperature T, the relative changes in the free energy of stability [$\Delta\Delta G(T)$] are usually expressed in reference to glycine and can be written as:

$$\Delta\Delta G(T) = \Delta\Delta H(T_R) + \Delta\Delta C_p \cdot (T - T_R) - T \cdot (\Delta\Delta S(T_R) + \Delta\Delta C_p \cdot 1n\ (T/T_R)) \quad (9)$$

The entropy change at the reference temperature T_R can be obtained by rearranging equation 9:

$$\Delta\Delta S(T_R) = [\Delta\Delta H(T_R) + \Delta\Delta C_p \cdot (T - T_R) - \Delta\Delta G(T)]/T - \Delta\Delta C_p \cdot 1n\ (T/T_R) \quad (10)$$

where the double Δ is used to indicate that each thermodynamic parameter is relative to glycine.

RESULTS

Energy Profiles

Profiles were first calculated for all side chains with one or two dihedral angles. Dihedrals involving methyl hydrogens (or those of $-NH_3^+$) were considered not to contribute significantly to side chain entropy for the following two reasons: 1) the contribution of the kinetic energy states ($L_j^2/2I_j$; = 1/2 $I_j\,\omega_j^2$) is very low because the moment of inertia around the bond is extremely low; and 2) the conformational entropy is zero because the three possible conformers are indistinguishable. Dihedral angles involving polar hydrogens were included in the calculations for serine and threonine. These calculations showed very clearly that rotation of these polar hydrogens produces threefold symmetrical potential energy profiles. In subsequent calculations polar hydrogens

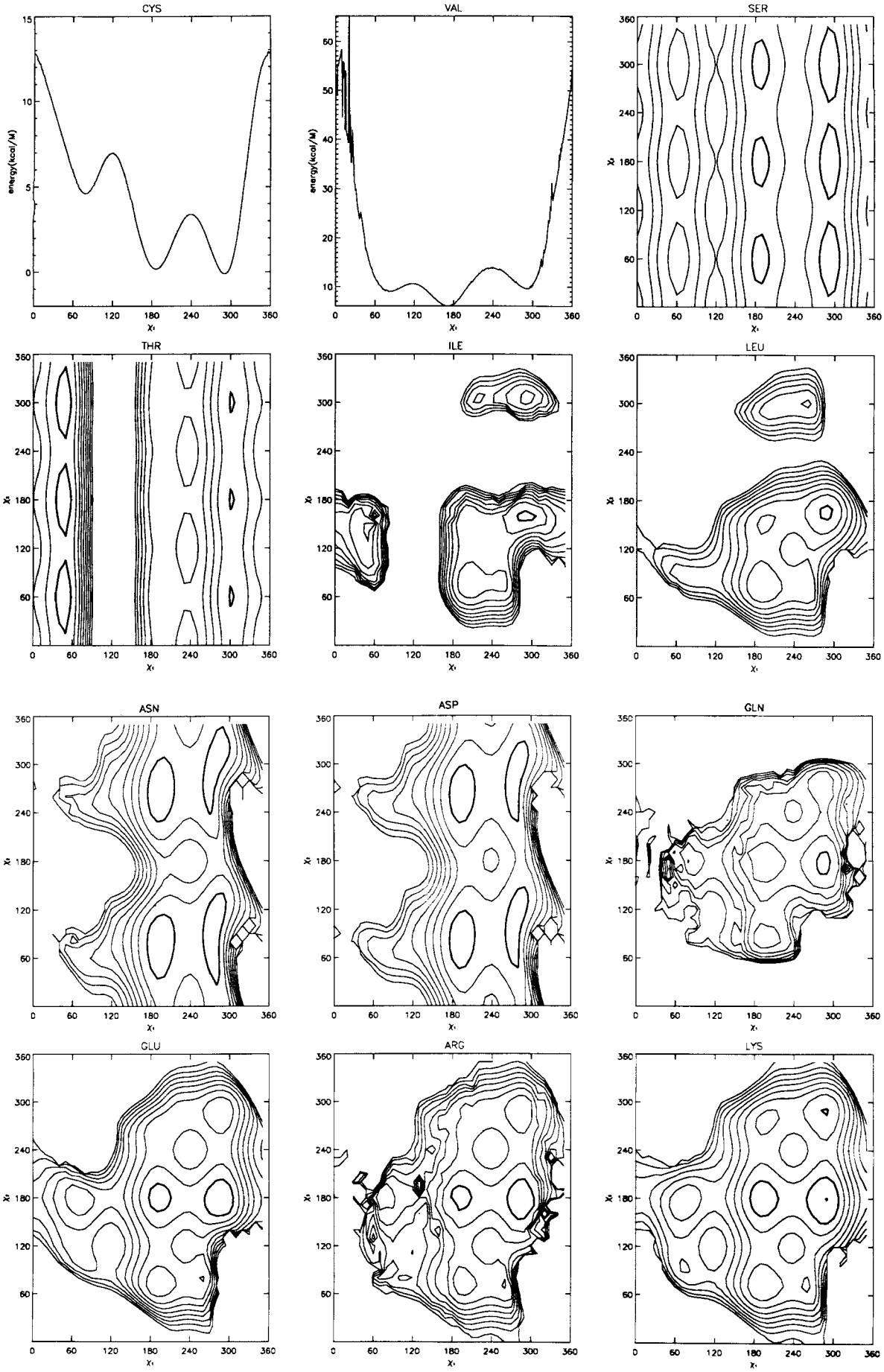
CYS
VAL
SER
THR
ILE
LEU
ASN
ASP
GLN
GLU
ARG
LYS
energy(kcal/M)
χ1
χ2

Fig. 1.

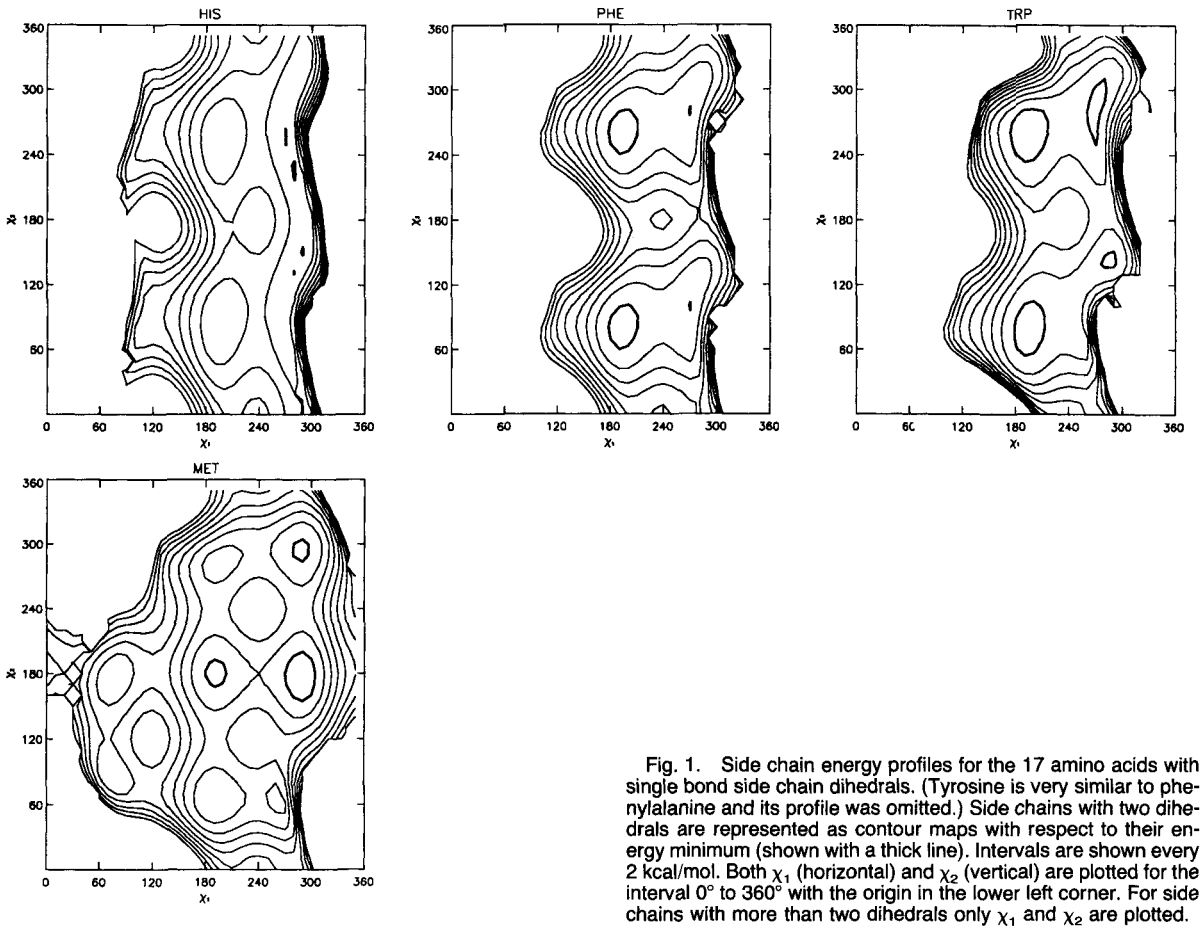

Fig. 1. Side chain energy profiles for the 17 amino acids with single bond side chain dihedrals. (Tyrosine is very similar to phenylalanine and its profile was omitted.) Side chains with two dihedrals are represented as contour maps with respect to their energy minimum (shown with a thick line). Intervals are shown every 2 kcal/mol. Both χ_1 (horizontal) and χ_2 (vertical) are plotted for the interval 0° to 360° with the origin in the lower left corner. For side chains with more than two dihedrals only χ_1 and χ_2 are plotted.

were considered to contribute R ln 3 (for cysteine) or R ln 2 (for tyrosine) to the overall entropy. Profiles for side chains with three dihedral angles were calculated for χ_1 and χ_2 as before, for values of χ_3 kept at each of the expected conformers. These calculations showed that the profiles on χ_1 and χ_2 were very similar for all χ_3 values. For side chains with more than three dihedrals, profiles were calculated for χ_1 and χ_2 at fixed values of the other dihedrals.

The profiles for the side chains studied in the α-helical environment show very distinct conformational preferences (Fig. 1). The rotation of the hydroxyl hydrogen for serine and threonine (χ_2) is threefold symmetrical, indicating that the hydrogen does not interact with either the main chain or the β carbons of adjacent residues. Residues branched at the C_β (valine, threonine, and isoleucine) have strong conformational preferences around χ_1. Residues that extend beyond Cγ (glutamic acid, glutamine, lysine, and arginine) have extremely similar energy profiles, indicating that these side chains experience similar conformational restrictions in an α-helix. As expected, serine and cysteine have very similar energy profiles. (In the case of serine we present the variation around χ_2, that is, the rotation of the hydroxyl hydrogen around the C-O bond. This rotation has almost perfect threefold symmetry and has no effect on the profile around χ_1.)

The probabilities of the conformational states calculated with these energy profiles (Fig. 2) correlate very well with the frequencies of side chain dihedrals observed in α-helices in proteins of known three-dimensional structure.[4,14] In the case of valine the probability profile is also very similar to that calculated by Creamer and Rose[2] using a Monte Carlo algorithm. (Although these authors analyze many of the amino acid side chains, they only report the probability profile for valine.)

Side Chain Entropies

The entropies calculated for the central residue of the nine residue α-helix using Method 1 can be directly compared with those reported by Creamer and Rose[2] (Table I). As expected, for the side chains reported by these authors, there is very good agreement between both calculations. It appears that in the case of phenylalanine and tyrosine Creamer and Rose[2] do not correct for the symmetry of the ring for the rotation around χ_2 ($\sigma = 2$). We report the correct values in Table I and, in addition, we give (in pa-

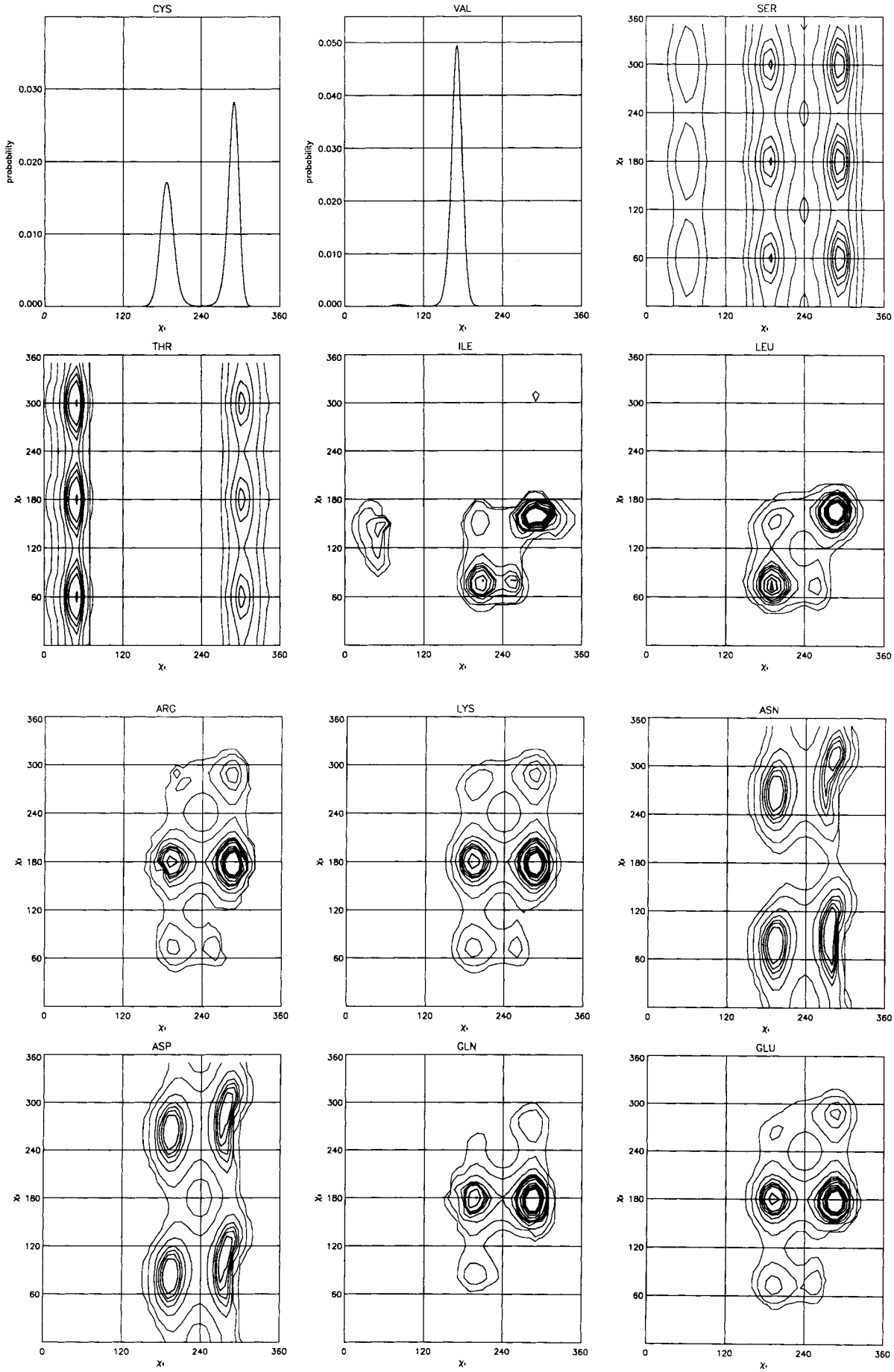
CYS
VAL
SER
THR
ILE
LEU
ARG
LYS
ASN
ASP
GLN
GLU
probability
χ1
χ2
0
120
240
360
60
180
300
0.000
0.010
0.020
0.030
0.040
0.050

Fig. 2.

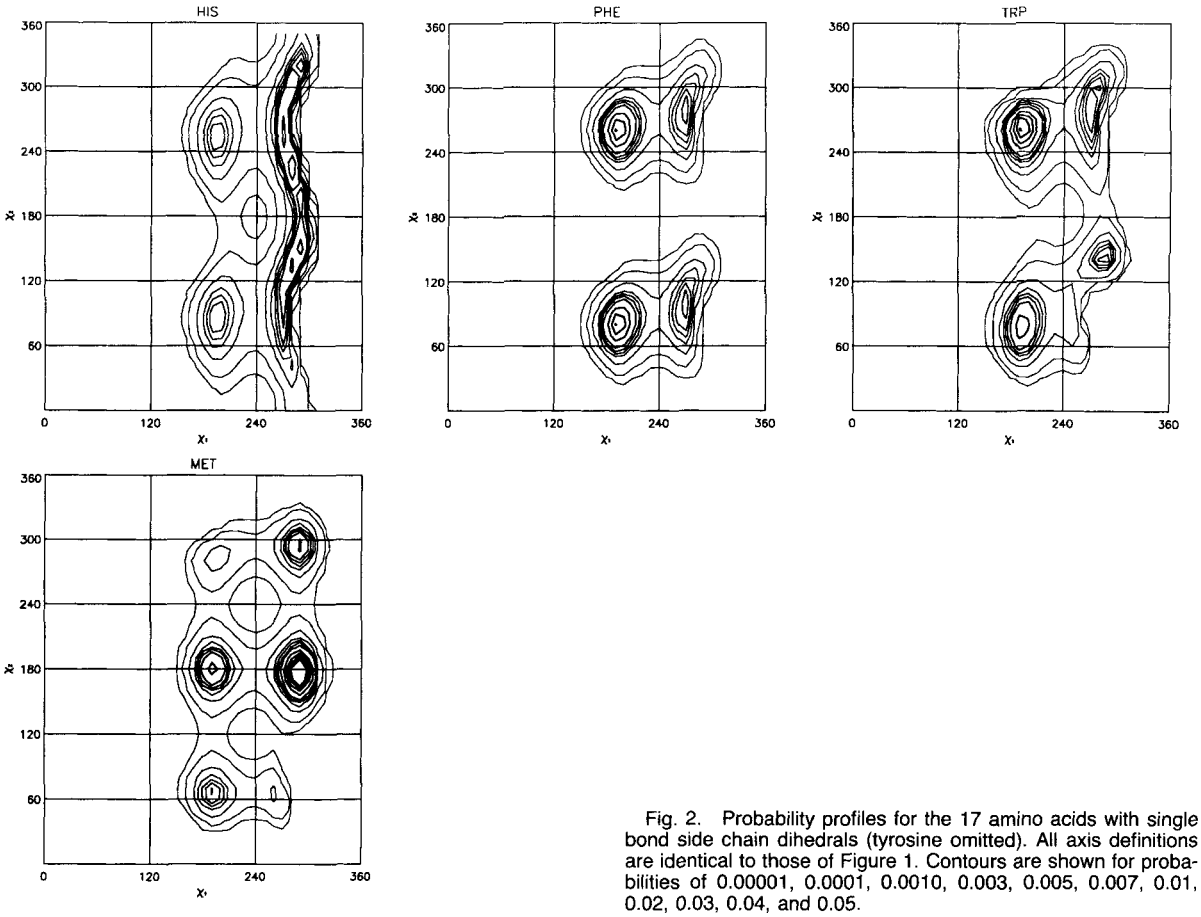

Fig. 2. Probability profiles for the 17 amino acids with single bond side chain dihedrals (tyrosine omitted). All axis definitions are identical to those of Figure 1. Contours are shown for probabilities of 0.00001, 0.0001, 0.0010, 0.003, 0.005, 0.007, 0.01, 0.02, 0.03, 0.04, and 0.05.

rentheses) the values not corrected for the symmetry to be directly comparable to those of Creamer and Rose.[2] (These values are just the correct values plus R 1n 2.)

The conformational entropies can also be compared directly with the values obtained from the frequencies of occurrence of the different side chain rotamers in the helices of proteins of known structure; Blaber et al.[4] report such a calculation with values expressed as TS^{α}—the absolute temperature times the side chain conformational entropy. There is outstanding agreement between our values and those of Blaber et al.[4] for all amino acids compared (Table I), the maximum difference being less than 0.3 kcal/mol. Since the method we proposed can be used for any individual residue in any protein of known structure, this is an important validation of the accuracy of the procedure. Abagyan and Trotov[15] reported an analysis similar to that of Blaber et al.[4] but did not differentiate the populations of the conformers according to the main chain conformations. Although their side chain conformer populations are averaged over all main chain conformations and not only over α-helices, there is an excellent agreement between their TS values and those calculated based on energy profiles of α-helices (Table I). This agreement is not completely surprising because the conformers in the data base do contain a high proportion of helical cases.

The entropies calculated with Method 3 provide the total configurational entropy for the side chain movement (Table II; they are presented only for amino acids with one or two side chain single bond dihedrals because the complete energy profile is required for the calculation.). They are larger than those calculated by Method 1 for two reasons: 1) the conformational entropy as calculated in Method 3 is obtained by integrating over the whole angular range; and 2) the calculation includes the contribution of the kinetic energy states. Although these entropies are intrinsically much more accurate than those calculated with Method 1, they can only be used for computing entropy changes between two states in situations in which the complete calculation can be performed in both states. (For example, this calculation would be difficult to perform for side chains in the unfolded state of proteins.) A simpler calculation of the complete configurational entropy can be carried out using Method 2. In this method it is assumed that the wells in the simplified potential energy surface can be described by an harmonic potential (quadratic function). Under this assumption,

TABLE I. Side Chain Entropy (Method 1)*

Amino acid	Angle	Probability			S/R	σ	$N^{\dagger}_{add}$	$(S/R)^{\ddagger}_{tot}$	(S/R)**	TS[¶] (kcal/M)	TS[§] (kcal/M)	TS[¶¶] (kcal/M)
Cys	χ_1	0.0002	0.4312	0.5686	0.69	1				(0.44)	0.41	
							3	1.78		1.15		1.14
Ser	χ_1	0.021	0.367	0.612	0.75	1				(0.49)	0.60	
							3	1.85		1.20		1.19
		0.007	0.007	0.007								
	χ_1, χ_2	0.123	0.122	0.124	1.85			1.85				
		0.202	0.203	0.205								
Thr	χ_1	0.748	0.000	0.252	0.56	1				(0.36)	0.30	
							3	1.662		1.073		1.12
	χ_1, χ_2	0.251	0.249	0.246								
		0.000	0.000	0.000	1.67	1		1.67				
		0.084	0.084	0.085								
Val	χ_1	0.008	0.990	0.002	0.06	1		0.06	0.100	0.04	0.23	0.50
Ile		0.002	0.023	0.000								
	χ_1, χ_2	0.198	0.019	0.000	0.88	1		0.88	0.755	0.57	0.47	0.75
		0.044	0.713	0.000								
Leu		0.000	0.000	0.000								
	χ_1, χ_2	0.300	0.040	0.000	0.82	1		0.82	0.831	0.53	0.65	0.75
		0.010	0.650	0.000								
Met		0.000	0.000	0.000								
	χ_1, χ_2	0.073	0.193	0.015	1.19	1	3	2.29	2.327	1.48		1.53
		0.006	0.577	0.135						(0.77)	0.82	
Gln		0.000	0.000	0.000								
	χ_1, χ_2	0.008	0.198	0.003	0.73	1	6	2.53		1.63		2.02
		0.000	0.632	0.016						(0.47)	0.86	
Glu		0.000	0.000	0.000								
	χ_1, χ_2	0.010	0.220	0.002	0.68	2	6	1.78		1.15		1.65
		0.000	0.750	0.018						(0.44)	0.89	
Arg		0.000	0.000	0.000								
	χ_1, χ_2	0.010	0.188	0.001	0.69	1	18	3.58		2.30		2.13
		0.002	0.769	0.030						(0.44)	0.77	
Lys		0.000	0.000	0.000								
	χ_1, χ_2	0.014	0.223	0.003	0.75	1	9	2.95		1.90		2.21
		0.001	0.730	0.028						(0.49)	0.81	
His		0.000		0.000								
		0.125		0.119								
	χ_1, χ_2	0.114	0.200	0.340	1.68			1.68		1.09		0.99
		0.102				(2)		(0.99)		(0.64)	0.84	
Asp		0.000		0.000								
		0.256		0.256								
	χ_1, χ_2	0.159	0.085	0.157	1.70	2		1.01		0.65	0.52	0.61
		0.086						(2.04)				
Asn		0.000		0.000								
		0.268		0.254								
	χ_1, χ_2	0.217	0.055	0.081	1.66	1		1.66		0.99		
		0.124				(2)		(0.96)		(0.62)	0.75	0.81
Phe		0.000		0.000								
		0.367		0.370								
	χ_1, χ_2	0.106	0.025	0.106	1.40	2		0.71	0.998[§§]	0.46	0.60	0.58
		0.026							(1.691)			
Tyr					1.40	2	2	1.40		0.90	0.62	0.99
Trp		0.000		0.000								
		0.290		0.452								
	χ_1, χ_2	0.021	0.063	0.128	1.38	1		1.38	1.846	0.89	0.79	0.97
		0.046										

*1, Probabilities based on rotamer class classification. For χ_1, the order of column is g^-, t, g^+. For χ_1, χ_2, the column shows χ_1 and the row shows χ_2. 2, $(S/R)_{tot} = S/R + 1n\, N_{ad} - 1n\, \sigma$.

†N_{add} is the number of rotamer class of side chain dihedral angles χ_i where i is index larger than that of the angle considered in energy calculations.

‡Symmetry corrected $(S/R)_{tot}$ values. The values is parenthesis are not symmetry-corrected for comparison with Creamer and Rose.[2]

**Values in Creamer and Rose.[2]

¶Our TS values at temperature 325K for comparison with Blaber et al.[5] The values in parenthesis are obtained from considering the χ_1 or χ_1 and χ_2 excluding the polar hydrogen for comparison with Blaber et al.[5]

§Values in Blaber et al.[5]

¶¶Values in Abagyan and Totrov[15] at T = 300K.

§§Symmetry-corrected value, which was not considered in their work.

TABLE II. Side Chain Entropy (Method 3)

Res	Moment of inertia (10^{-40} gcm^2)	Symmetry number	1n Z	⟨V⟩/RT	S/R
Val	73	1	0.84	0.60	1.94
Cys	98	1	1.55	0.77	2.82
Leu	162	1			
	73	1	2.76	1.50	5.26
Ile	111	1			
	37	1	2.05	1.57	4.63
Phe	731	1			
	90	2	4.04	1.32	6.37
Trp	1540	1			
	435	1	5.67	1.54	8.21
His	515	1			
	55	1	4.67	1.30	6.97
Asp	182	1			
	40	2	3.72	1.08	5.79
Ser	32	1			
	0.83	1	1.94	1.30	4.24
Thr	69	1			
	0.84	1	2.06	1.24	4.30

the entropy can be calculated as the sum of two terms, one corresponding to the populations of the different conformations and the other corresponding to the entropy of one harmonic oscillator per well in the energy surface. The entropy of the harmonic oscillators is calculated using the equations of an Einstein oscillator. The approximations used in this method were tested for the nine residue helix with valine at position 5 ([Ala]$_4$-Val-[Ala]$_4$). The use of quadratic expressions for the representation of the potential well around each of the three minima appears to provide an excellent functional form for the potential. The rms errors ($1/2\{\Sigma[E_{true}-E_{quadr}]^2\}^{1/2}$) range from 0.2 to 2.1 kcal/mol for dihedrals varying up to 20° from the equilibrium angle. The three wells have very similar "spring constants" (Table III), reflecting the similar shapes and curvatures of the potential energy curves in the wells. Accordingly, the three oscillators have very similar Einstein temperatures (Table III). The major differences in their contributions to the total entropy reflect mainly the difference in the depth of their potential wells. The total entropy is calculated as the sum of the contributions of the three oscillators (Table III) plus the conformational entropy (Table I). The value of S/R obtained in this manner (1.911) is in excellent agreement with the value obtained by complete integration (1.940).

The side chain configurational entropy was also calculated for a side chain that becomes buried in a binding process. In a previous paper we described calculations on the thermodynamics of binding in a complex of angiotensin II (AII) with the high affinity

TABLE III. Side Chain Entropy (Method 2): Residue Val*

Angle (degree)	E_{min} (kcal/M)	Spring constant k	$\omega = (k/I)^{1/2}$	Θ_ω/T	S/R
81	3.0	28.22	1.64×10^{13}	0.42	0.01
171	0.0	31.04	1.72×10^{13}	0.44	1.83
292	3.5	36.54	1.87×10^{13}	0.48	0.01

*$S_{vib}/R = 0.01 + 1.83 + 0.01 = 1.85$.

monoclonal antibody Mab 131.[1a] ValH134 of Mab 131, a buried residue, has an energy profile with a single well at $\chi_1 = 170°$, very close to the position of the side chain in the X-ray structure. Least-squares fit of the profile of this well with a quadratic function gives a spring constant of 30.3 kcal/mol · rad^2. This constant is right in the middle of the range of the constants found for the three wells of the exposed valine in the center of a nine residue α-helix. This result indicates that when a side chain becomes buried the number of conformational wells accessible to the side chain is drastically reduced, but the oscillations within the well do not appear to be damped. These results provide a justification for the use of simplified methods of calculating changes in configurational entropy: in cases in which the shape of the potential function is expected to be similar in the initial and final states, one can use the changes in conformational entropy as the total change in configurational entropy.

Entropy Changes

Changes in side chain configurational entropy were calculated for the transition between the unfolded state and a folded state in which the side chain is exposed and forms part of an α-helix. The value of the configurational entropy (computed with either Method 2 or Method 3) is difficult to calculate for side chains in the unfolded state because complete energy profiles would have to be calculated for all the conformations of the main chain contributing to the unfolded state. If the shapes of the potential wells of all conformers contributing to both the unfolded and the helical state are similar, one could use a calculation based only on side chain conformational entropies. We performed such a calculation in order to evaluate the validity of the approximation by comparing values computed in this manner with experimental values. We used the values in Table I for the configurational entropy of the side chains in the α-helix and the values of Creamer and Rose[2] for the conformational entropies in the unfolded state.

Calculation of Side Chain Configurational Entropy From Experimental Data

In order to estimate the change in entropy at the reference temperature $\Delta\Delta S(T_R)$, values for $\Delta\Delta G(T)$,

TABLE IV. Relative Free Energies for α-Helix Preferences of Different Amino Acids

Amino acid	ΔΔG(25) Barnase* cal/mol	ΔΔG(20) OD† cal/mol	ΔΔG(52) T4-44‡ cal/mol	ΔΔG(4) LLMK** cal/mol
ALA	910.0	770.0	960.0	790.0
ARG	770.0	680.0	770.0	
ASN	250.0	70.0	390.0	180.0
ASP	200.0	150.0	420.0	
CYS	90.0	230.0	420.0	
GLN	430.0	330.0	800.0	480.0
GLU	360.0	270.0	530.0	
GLY	0.0	0.0	0.0	0.0
HIS	130.0	60.0	570.0	
ILE	100.0	230.0	840.0	390.0
LEU	560.0	620.0	920.0	620.0
LYS	720.0	1230.0	730.0	
MET	600.0	500.0	860.0	570.0
PHE	220.0	410.0	590.0	
SER	500.0	350.0	530.0	280.0
THR	120.0	110.0	540.0	230.0
TRP	70.0	450.0	580.0	
TYR	90.0	170.0	720.0	
VAL	30.0	140.0	630.0	340.0

*Horovitz et al.[6] Data obtained at 25°C.
†O'Neil and Degrado.[8] Data obtained at 20°C.
‡Blaber et al.[5] Data obtained at 52°C.
**Lyu et al.[7] Data obtained at 4°C.

the enthalpy change [$\Delta\Delta H(T_R)$], and the heat capacity change ($\Delta\Delta C_p$) for each amino acid are needed. Of these three parameters, ΔΔG(T) values have been the easiest to measure experimentally and several sets of data are available in the literature.[4–8] These experimental values obtained for different systems are summarized in Table IV and Figure 3. At the present time, experimental enthalpies and heat capacities for single amino acids in α-helices are not available; therefore, we estimated these parameters from changes in solvent accessible polar and apolar surface areas calculated from the atomic coordinates as described recently.[9]

Calculated accessible polar and apolar surface areas for each amino acid in the α-helix conformation and in an extended conformation (Table V) can be used to predict ΔC_p and ΔH(25), the enthalpy of helix unfolding at 25°C. The values for the α-helix were obtained after energy minimizing the same $(Ala)_4$-Xaa-$(Ala)_4$ helices used in the previous calculations. The minimization was performed with the program XPLOR using the Powell conjugate gradient option. Accessible surface areas (ASA) were calculated as described before[9] using the program ACCESS (Scott R. Presnell, University of California, San Francisco, CA), an implementation of the Lee and Richards algorithm,[17] with a probe radius of 1.4 Å and a slice width of 0.25 Å. Accessible surface areas in the unfolded state were modeled as the sum of the ASA values of the individual residues in the extended Ala-Xaa-Ala tripeptide. These values show that different amino acids bury different proportions of polar and apolar surface even in a solvent exposed α-helix. The calculations gave similar results with the α-helix containing residue 44 in T4 lysozyme (serine in the wild type and replaced by all other amino acids).[4,5] As a consequence, the unfolding of the helix results in the hydration of different amounts of polar and apolar surface for each amino acid. Since the hydration of polar and apolar surfaces contributes differently to ΔC_p, it follows that each amino acid will have a characteristic ΔC_p. In general, the hydration of polar surfaces contributes negatively to ΔC_p (-0.26 cal/K·mol-Å^2) while the hydration of apolar surfaces contributes positively (0.45 cal/K·mol-Å^2). The amino acids that bury a larger polar fraction in the α-helix conformation are characterized by negative ΔC_p of unfolding, whereas those that bury a significant apolar surface are characterized by positive ΔC_p values. For example, glycine has a ΔC_p of -11 cal/K·mol, which arises primarily from the exposure of the peptide backbone to the solvent upon unfolding. On the other hand, the aromatic amino acids tryptophan, phenylalanine, and tyrosine bury large apolar areas in the helical state and exhibit a positive ΔC_p upon unfolding. These calculations agree with the conclusion of Matthews and co-workers[4,5] that solvent related interactions and not only side chain and backbone configurational entropy changes contribute to the different helix forming propensities of amino acids. In general, however, the expected ΔC_p values are relatively small and do not convey a significant temperature dependence to the enthalpy change over these short temperature ranges.

The differences in solvent interactions for the various amino acid side chains in the α-helix are also reflected in the enthalpy values. Overall, the predicted enthalpy change for helix unfolding at 25° [ΔH(25)] averages 1.1 kcal/mol, which is close to various estimates in the literature[18–20] (see also refs. 1 and 21 for reviews). For alanine, for which the enthalpy change has been recently measured calorimetrically,[21] the predicted value is 1.3 kcal/mol at the experimental transition temperature of 45°C. This value is in excellent agreement with the experimental value of 1.3 kcal/mol, lending further experimental support to the applicability of the structural parameterization to these systems. The enthalpy values listed in Table V are within the range expected except for tryptophan, which is predicted to exhibit a negative enthalpy change at 25°C. This peculiar behavior of tryptophan in the calculation arises from the relatively large apolar surface that is predicted to be buried from the solvent in the energy minimized α-helical structure.

It has been shown before that at 112° the solvent related entropy change associated with the hydra-

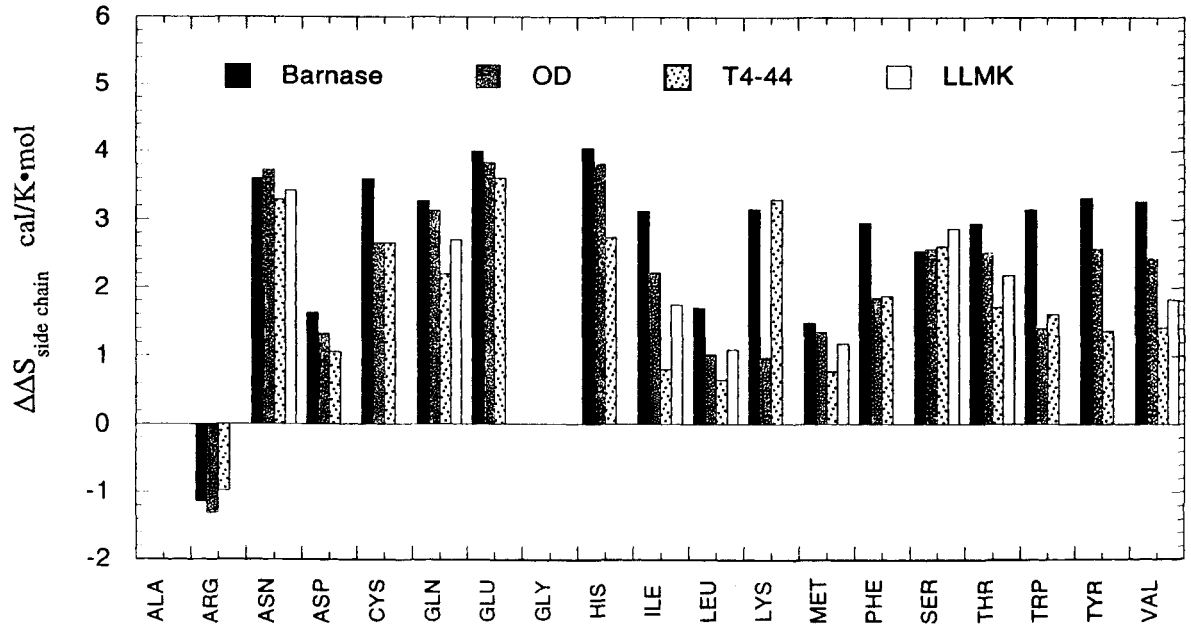

Fig. 3. Comparison of the changes in side chain configurational entropy on helix formation computed from four different experimental data sets. The values are obtained from experimental ΔΔG values of the effect amino acid substitutions on the stability of α-helices using equation 10. The values are from Horovitz, Matthews and Fersht,[6] O'Neil and DeGrado,[8] Blaber et al.,[5] and Lyu et al.,[7] respectively.

tion of apolar groups is zero.[22] Thus, after correction for protonation or other electrostatic effects, ΔS(112) is essentially equal to the configurational entropy change.[19] If those conclusions are correct, application of equation 2 together with the thermodynamic parameters listed in Tables IV and V should provide good estimates of the configurational entropy change per amino acid. The results of these calculations are listed in Table VI. Except for alanine, whose methyl group is not motionally restricted when the amino acid is in α-helical conformation, the configurational entropy changes include changes due to the backbone and changes due to the side chains. For alanine, the average value of −2.36 ± 0.30 cal/K · mol (calculated from the four sets of data) is due entirely to the restrictions imposed by the methyl group on the degrees of freedom of the peptide backbone, relative to glycine. The configurational entropy of the peptide backbone has been estimated on theoretical grounds, and is expected to be on the order of 5.5 cal/K·mol for glycine.[23] Nemethy et al.[24] estimated that for alanine the backbone restrictions amount to −2.4 cal/K · mol, a figure extremely close to the average value given in Table VI. These results are consistent with the conjecture that ΔΔS(112) is essentially equal to the configurational entropy change.

The largest steric restriction on the degrees of freedom of the peptide backbone is imposed by the β carbon. The presence of additional groups has little effect on the backbone entropy. Nemethy et al.[24] estimated that the additional restrictions imposed by the γ carbon amount only to 0.3 cal/K · mol and that additional terms contribute negligibly. The same authors concluded that only valine and isoleucine can be expected to make a larger contribution due to the presence of two γ carbons. So, if the ΔΔS(112) value for alanine is subtracted from the ΔΔS(112) values for all remaining amino acids except glycine, the resulting values should reflect predominantly the side chain entropy change plus the remaining higher order terms contributing to the backbone entropy restrictions. The results of this calculation are shown in Figure 3. As shown in the figure, the agreement between the values calculated using the four sets of experimental ΔΔG values is generally good. (Except for the charged residues the numerical agreement is expected not to vary significantly with pH or solvent.) These values can be compared with independent estimates of side chain entropy changes such as those described above or those calculated by Creamer and Rose[2] using Monte Carlo calculations. If the analytical values presented above together with the published values for the unfolded state[2] are used to calculate the theoretical side chain entropy changes, an excellent numerical agreement with the experimental values is obtained (Table VII; the values in the column labeled $\Delta\Delta S_{calc}$ are compared with the experimental values contained in the last three columns.) The best quantitative agreement is found

TABLE V. Calculated Thermodynamic Parameters for α-Helix Unfolding

Amino acid	ASA_{ap} Å^2	$ASA_{ap,ext}$ Å^2	ASA_{pol} Å^2	$ASA_{pol,ext}$ Å^2	$\Delta\Delta A_{ap}$ Å^2	$\Delta\Delta A_{pol}$ Å^2	$\Delta C_{p,calc}$* cal/K · mol	$\Delta H(25)_{calc}$† cal/mol
ALA	65.5	60.41	4.8	28.71	−5.09	23.91	−8.51	1,474.9
ARG	80.6	91.16	124.0	133.75	10.56	9.75	2.22	174.97
ASN	27.8	30.66	76.4	111.99	2.86	35.59	−7.97	1,843.1
ASP	35.2	50.49	64.7	81.74	15.29	17.04	2.45	412.6
CYS	24.8	40.84	65.3	96.06	16.04	30.76	−0.78	1,135.1
GLN	58.9	49.7	75.2	113.47	−9.2	38.27	−14.09	2,396.2
GLU	62.1	60.88	49.4	91.76	−1.22	42.36	−11.56	2,349.8
GLY	39.8	25.74	9.1	28.71	−14.06	19.61	−11.43	1,543.3
HIS	101.2	99.05	37.8	73.95	−2.15	36.15	−10.37	2,042.7
ILE	116.3	127.41	1.6	28.71	11.11	27.11	−2.05	1,102.5
LEU	122.2	135.45	0.6	28.71	13.25	28.11	−1.35	1,084.8
LYS	109.2	116.15	32.4	77.79	6.95	45.39	−8.67	2,239.2
MET	94.3	122.06	45.8	72.95	27.76	27.15	5.43	542.78
PHE	140.3	176.73	0.0	28.71	36.43	28.71	8.93	335.18
SER	42.4	41.15	30.6	64.43	−1.25	33.83	−9.36	1,885.9
THR	64.0	74.79	37.1	63.31	10.79	26.21	−1.96	1,064.3
TRP	125.9	200.83	27.0	52.41	74.93	25.41	27.11	−1,144.0
TYR	109.0	153.93	42.4	70.25	44.93	27.85	12.98	1.44
VAL	95.0	113.04	0.0	26.19	18.04	26.19	1.31	818.5

*ΔC_p were calculated from the changes in accessible surface areas using the formula $\Delta C_p = \Delta c_{p,ap}\,\Delta\Delta A_{ap} + \Delta c_{p,pol} \cdot \Delta\Delta A_{pol}$ where the elementary parameters are $\Delta c_{p,ap}$ = 0.45 cal/K · mol − apolar Å^2 and $\Delta c_{p,pol}$ = −0.26 cal/K · mol · polar Å^2.[9]
†$\Delta H(25)$ was calculated with the standard thermodynamic equation $\Delta H(25) = \Delta H(100) + \Delta C_p$ (25-100), where the enthalpy change as the reference temperature of 100 has been shown to scale linearly with $\Delta\Delta A_{pol}$ as $\Delta H(100) = 35 \cdot \Delta\Delta A_{pol}$.[9]

TABLE VI. Calculated Contribution for Configurational Entropy

Amino acid	ΔΔS(112) Barnase* cal/K · mol	ΔΔS(112) OD† cal/K · mol	ΔΔS(112) T4-44‡ cal/K · mol	ΔΔS(112) LLMK** cal/K · mol
ALA	−2.53	−2.06	−2.70	−2.13
ARG	−3.68	−3.38	−3.68	
ASN	1.05	1.66	0.58	1.29
ASP	−0.91	−0.74	−1.65	
CYS	1.05	0.58	−0.05	
GLN	0.74	1.07	−0.50	0.57
GLU	1.46	1.76	0.89	
GLY	0.0	0.0	0.0	0.0
HIS	1.51	1.74	0.03	
ILE	0.59	0.15	−1.89	−0.38
LEU	−0.84	−1.04	−2.04	−1.04
LYS	0.62	−1.09	0.59	
MET	−1.05	−0.72	−1.92	−0.95
PHE	0.42	−0.22	−0.82	
SER	0.002	0.50	−0.1	0.74
THR	0.41	0.45	−0.99	0.04
TRP	0.62	−0.66	−1.09	
TYR	0.77	0.51	−1.34	
VAL	0.73	0.36	−1.28	−0.31

*Horovitz et al.[6] Data obtained at 25°C.
†O'Neil and Degrado.[8] Data obtained at 20°C.
‡Blaber et al.[5] Data obtained at 52°C.
**Lyu et al.[7] Data obtained at 4°C.

between these $\Delta\Delta S_{calc}$ values and the semiempirical values derived from the T4 lysozyme data[4,5] (Fig. 4 and Table VII). For the aliphatic amino acids (Gly, Ala, Val, Leu, and Ile) the average deviation is 0.13 cal/K · mol (or 0.2 cal/K · mol if Ala and Gly are not included). For the aromatic residues it is of the order of 0.3 cal/K · mol. The agreement is also very good for the charged amino acids including the prediction of the negative sign of the value for arginine. The largest deviations occur for the aliphatic hydroxyl side chains (Ser and Thr) and for Cys that contains a sulfydryl group. The origin of this discrepancy appears to be in the structure-based calculation of ΔC_p for these amino acids. Recent heat of solution experiments on solid dipeptides (Laynez, Freire, and Murphy, personal communication) suggest that the hydroxyl contribution to the ΔC_p might be positive and not negative as previously assumed. Refinement of these methods will be possible as experimental ΔH an ΔC_p values for individual amino acids become available and the algorithm to calculate the changes in solvent accessible area is improved by using probability weighted areas for the side chains.

DISCUSSION

Theoretical Calculations

Accurate calculation of the configurational entropies of a given side chain in a given structure requires the calculation of the energy surface for all the internal coordinates of the side chain in that structure. For a side chain of n_s atoms there are $6 \times n_s$ internal coordinates (degrees of freedom) accessible to the side chain atoms. Of those, only the rotations around the single bond side chain dihedrals can contribute to entropy changes for processes at constant temperature, i.e., their energy states

TABLE VII. Side Chain Entropies

Amino acid	S_u (CR) (cal/K · mol)	S_{ex} (cal/K · mol)	$S_u - S_{ex}$ ($\Delta\Delta S_{cal}$)	$S_u - S_{ex}$(CR) ($\Delta\Delta S_{cal,CR}$)	ΔΔS (T4-44)	ΔΔS (Barnase)	ΔΔS (deGrado)	ΔΔS (Kallenbach)
ALA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ARG	6.27	7.11	−0.84		−0.71	−1.15	−1.38	
ASN	5.53	3.29	2.24		3.29	3.59	3.72	3.42
ASP	4.16	2.00	2.16		1.16	1.62	1.30	
CYS	4.16	3.55	0.61		2.57	3.59	2.66	
GLN	7.14	5.02	2.13		1.92	3.27	3.22	3.03
GLU	5.80	3.53	2.28		3.24	4.00	3.91	
GLY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HIS	4.23	3.44	0.79		2.48	4.05	3.88	
ILE	2.42	1.74	0.67	0.98	0.85	3.12	2.22	1.74
LEU	1.88	1.63	0.25	0.14	0.73	1.70	1.01	1.02
LYS	6.88	5.86	1.02		3.02	3.16	1.00	
MET	5.13	4.55	0.59	0.50	0.96	1.48	1.31	0.96
PHE	4.29	2.78	1.51	0.86	2.03	2.96	1.80	
SER	4.23	3.68	0.55		2.37	2.54	2.62	3.11
THR	3.79	3.31	0.48		1.68	2.95	2.53	2.20
TRP	3.89	2.74	1.15	0.41	2.10	3.15	1.27	
TYR	5.90	4.16	1.74	1.09	1.63	3.31	2.52	
VAL	1.41	0.12	1.30	1.28	1.47	3.26	2.42	1.74

change significantly during the process. Thus, it is possible to have a complete representation of the conformational space by considering only the side chain dihedrals. At each value of the dihedrals it is necessary to find the equilibrium conformation with respect to all the other internal modes. This is easily accomplished by minimizing the energy while keeping the dihedrals at their fixed values. Energy minimization at constant dihedrals can be accomplished in cartesian coordinates—by imposing a strong constraint against change of the dihedral—or in angular coordinates—by minimizing with respect to the valence bond angles. (These are the "softer" side chain internal coordinates at fixed side chain dihedrals). The refinement in cartesian coordinates can be carried out with public domain or commercial programs. Refinement in angular coordinates, carried out with the software we developed, converges very quickly while keeping the dihedral exactly at the selected value because it uses angular coordinates that are essentially orthogonal to the dihedral angles. (All calculations were performed with the code in angular coordinates written by us. The calculation for valine was repeated using the program XPLOR. The values of the probabilities of the three conformers using XPLOR were 0.025, 0.965, and 0.01, very similar to the values obtained with the refinement in angular coordinates—0.008, 0.990, and 0.002.)

The values of the probabilities of the different conformers obtained from the energy profiles agree extremely well with the populations of conformers found in the helices of proteins of known structure.[4,15] This agreement suggests that the populations of conformers found for the side chains of folded proteins are just a reflection of the Boltzman distribution of the energies of the conformers.

Once a complete energy profile is available, the partition function can be computed by direct integration and used to calculate the configurational entropy. There is excellent agreement between values of the side chain configurational entropies calculated in this paper and the values obtained from the analysis of proteins of known structure (Table I). Analysis of the shape of the potential wells around the minima of the energy profile indicates that many of the wells have very similar shapes. They can be accurately represented by a quadratic expression and most of them have similar spring constants. In addition, spring constants appear to be similar for buried and for exposed residues. In processes in which the spring constants are similar in the initial and final states, the changes in configurational entropy can be approximated by changes in the conformational entropy calculated using rotamer distributions computed from the energy profiles. This approach was used without a specific justification in previous publications. Our semiempirical calculations of the configurational entropy differences for individual side chains between the unfolded state and an α-helix using previously published experimental data agree very well with the conformational entropy differences calculated with Method 1 (Table IV). This suggests, among other things, that indeed the shape of the conformational wells in the unfolded state and in the exposed-folded state are very similar. As shown above, the conformational wells of buried residues can also be similar to those of exposed residues.

The results of these studies indicate that the en-

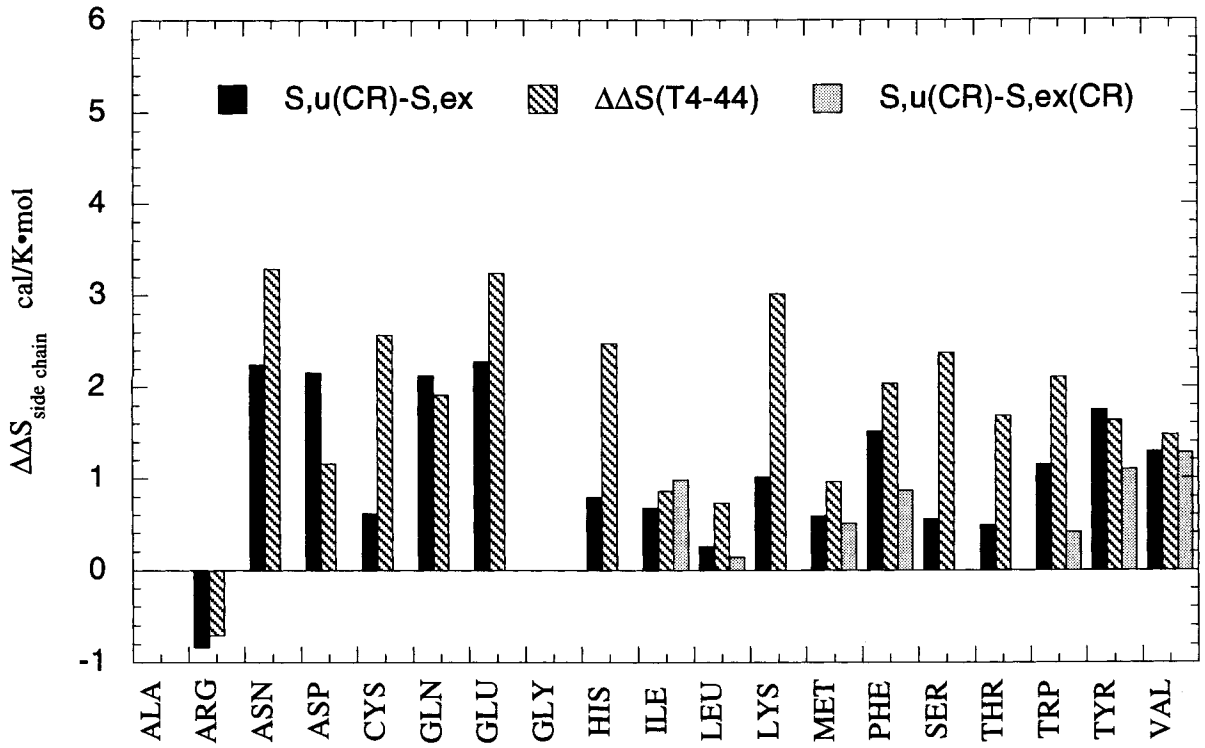

Fig. 4. Comparison of experimental and theoretical values of side chain entropy changes. Striped bars show values obtained from the experimental data of Blaber et al.[5] using equation 10. Black bars are the theoretical values calculated as described in this paper. S_{ex} is the entropy of an exposed side chain in an α-helix. S_u is the entropy of the side chain in the unfolded state. Gray bars show the values obtained using the results of Creamer and Rose[2] for both the helical and the unfolded state. CR refers to values from Creamer and Rose[2] and T4-44 to the results of Blaber et al.[5] for replacements of residue 44 in T4 lysozyme.

tropy changes associated with protein folding or peptide binding can be functionally divided into two terms: a solvent related term associated primarily with the hydrophobic effect, and to a lesser extent to the hydration of polar groups, and to electrostatic or protonation effects, and a purely configurational term related to changes in the degrees of freedom of the peptide backbone and the amino acid side chains. In this paper, we have presented theoretical estimates of side chain entropies for amino acid side chains located in exposed sites in α-helices. The side chain entropy changes calculated in this manner and those obtained from experimental ΔΔG values in conjunction with a semiempirical structural parameterization of the enthalpy and heat capacity changes developed earlier[9] are in excellent agreement (Fig. 4). This convergence suggests that the structural parameterization effectively accounts for all interactions and that the separation of the entropy changes in solvent effects and configurational entropy is adequate; otherwise the magnitude of the configurational entropies computed from the experimental values would have been much different from the independently calculated theoretical values.

Semiempirical Calculations

As discussed previously,[25,26] the configurational entropy change for peptide folding or binding can be subdivided into three different terms: 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. To a first approximation, the side chain entropies calculated in this paper (Table I) correspond to the term $\Delta S_{bu\rightarrow ex}$, since the amino acid side chains buried in the interior of a protein can be considered to be immobilized and have an entropy near zero (see for example ref. 14). The term $\Delta S_{ex\rightarrow u}$ is smaller in magnitude. Creamer and Rose[2] theoretically estimated $\Delta S_{ex\rightarrow u}$ for seven amino acids, finding that it ranges from about 0.14 cal/K · mol for leucine to 1.28 cal/K · mol for valine and averages 0.75 ± 0.46 cal/K · mol. For all 19 side chains we obtained an average of 0.87 ± 1.2 cal/K · mol. Recently Blaber et al.[4] have also estimated $\Delta S_{ex\rightarrow u}$ theoretically from an analysis of side chain conformations in 100 crystal-

lographic structures. These authors obtained a mean value of 0.56 ± 0.48 cal/K · mol for 19 amino acids, which is close to the value obtained by Creamer and Rose[2] and also similar to the value of 0.5 cal/K · mol obtained much earlier by Nemethy et al.[24] Finally, the entropy gained by the backbone is a function of the steric constraints imposed by the amino acid side chains. ΔS_{bb} is maximal for glycine, and on the order of 5.5 cal/K · mol.[23] The presence of a β carbon reduces ΔS_{bb} by −2.4 cal/K · mol according to Nemethy et al.,[24] a figure extremely close to the average value obtained in this paper (Table VI). The presence of additional side chain constituents has an almost negligible effect on the degrees of freedom of the peptide backbone, except perhaps for valine and isoleucine, in which the branching at the β carbon position might induce additional steric restrictions. Nemethy et al.[24] estimated that the added steric restrictions in valine and isoleucine may contribute an additional −2.4 cal/K · mol to ΔS_{bb}. This theoretical calculation, however, appears to overestimate the effect of branching, judging by the small differences observed in the experimental ΔΔG values obtained for leucine and isoleucine. If the theoretical side chain entropy changes,[2,4] (this work) are subtracted from the overall entropies for leucine and isoleucine, the maximal difference is on the order of only 0.75 cal/K · mol. Finally, the magnitude of the overall backbone entropy change upon unfolding will be diminished by the presence of disulfide bridges or other covalent links. An analysis of the restrictions in backbone entropy by the presence of disulfide bridges has been presented before.[27]

In the absence of covalent links, the total configurational entropy change associated with the unfolding of a protein will be a function of the amino acid composition, as well as the fraction of side chains that are buried in the interior of a protein. According to the partitioning of the configurational entropy discussed above, only those amino acids that are buried contribute to $\Delta S_{bu\rightarrow ex}$, while all the amino acids contribute to $\Delta S_{ex\rightarrow u}$ and ΔS_{bb}. If we take into consideration the average amino acid composition of a protein,[28] the term $\Delta S_{bu\rightarrow ex}$ is expected to be around 1.7 cal/K · mol on the average. This term needs to be multiplied by the average fraction of side chains that are buried in the interior of a protein, which is approximately 50%, yielding an overall $\Delta S_{bu\rightarrow ex}$ contribution of 0.85 cal/K · mol of residue. For a typical globular protein, the total configurational entropy contribution upon unfolding due to side chains is expected to be on the order of 1.4 cal/K · mol of residue. The backbone contribution is expected to be about 3 cal/K · mol of residue if only the steric restrictions due to β carbons are included, and about 2.7 cal/K · mol of residue if higher order terms are included. It follows that the overall configurational entropy change upon unfolding is between 4.1 and 4.4 cal/K · mol of residue. This figure is very close to the value of 4.3 ± 0.2 cal/K · mol of residue obtained at 112° for globular proteins. At this temperature Baldwin[23] has shown that the hydration entropy of apolar residues is very close to zero, and the resultant value has been attributed to the configurational entropy change after correction for protonation and electrostatic effects.[1,9]

In recent years Oobatake and Ooi[29] and Privalov and Makhatadze[30] have arrived at estimates of the configurational entropy change for protein unfolding on the order of 9 cal/K · mol of residue and 19 cal/K · mol of residue, respectively, significantly larger than the other estimates of this term discussed above. (Their values are a consequence of using large negative values for the hydration entropy of polar groups at 112°C obtained from the differences between the observed ΔS° of dissolution of linear alkanes and linear alcohols.) Our results seem to indicate that the original suggestion of Privalov and Khechinashvili[31] is the correct one, and that the entropy change at 112° provides a good measurement of the configurational entropy change for protein unfolding.

CONCLUSIONS

We have presented a general method to calculate the complete energy profile of any side chain in any structure. This energy surface can be used directly to calculate the configurational entropy of any given side chain. The contribution of the side chains to the changes in configurational entropy that take place in folding/unfolding and in binding can be calculated as the difference in the entropy of the side chain in the initial and the final states. In cases in which the shapes of the wells in the potential energy surfaces are similar in both states, the change in entropy can be simply evaluated as the change in conformational entropy.

For these calculated values to be useful it is necessary that the configurational entropy effects can be separated from the solvent effects. To investigate this point, side chain configurational entropies were estimated from experimental values of free energies of helix stabilization. The excellent agreement between the calculated and the experimental values validate several of the assumptions used in the calculations. First and most important, it validates the separation of the entropic effects in configurational and solvent contributions, that is, these two effects can be evaluated independently and then added. Second, it provides additional support for the parameterization of the solvent effects on ΔH, ΔC_p, and ΔS in terms of changes in buried polar and apolar areas. Third, it validates the conclusion that at 112°C the entropy change reflects primarily changes in configurational entropy. Thus, the results obtained indicate that the ΔG of folding and the ΔG of binding can in principle be accurately estimated by the methods

we propose from the structures of the molecules present in the initial and in the final states of the processes. At this point in time, the most important term for which no accurate parameterization is available is the backbone entropy. Calculations based on small modifications of the methods presented here can be used for estimating this term.

ACKNOWLEDGMENTS

We thank K. Murphy and G. Bains for helpful discussions and George Rose and Trevor Creamer both for helpful discussions and for providing their unpublished data on the conformational entropies in the unfolded state. This work was supported by NIH grants RR04328 and GM37911 (to E.F.), and GM44692 (to L.M.A.).

REFERENCES

1. Murphy, K.P., Freire, E. Thermodyanmics of structural stability and cooperative folding behavior in proteins. Adv. Protein Chem. 43:313–361, 1992.
1a. Murphy, K.P., Xie, D., Garcia, K.C., Amzel, L.M., Freire, E. Structural energetics of peptide recognition: Angiotensin II/antibody binding. Proteins 15:113–120, 1993.
2. Creamer, T.P., Rose, G.D. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc. Natl. Acad. Sci. USA 89: 5937–5941, 1992.
3. Karplus, M., Kushick, J.N. Method for estimating the configurational entropy of macromolecules. Macromolecules 14:325–332, 1981.
4. Blaber, M., Zhang, X.-J., Lindstrom, J.L., Pepiot, S.D., Baase, W.A., Matthews, B.W. Determination of alpha-helix propensity within the context of a folded protein: Sites 44 and 131 in bacteriophage T4 lysozyme. J. Mol. Biol. 235:600–624, 1994.
5. Blaber, M., Zhang, X.-J., Matthews, B.W. Structural basis of amino acid alpha helix propensity. Science 260:1637–1640, 1993.
6. Horovitz, A., Matthews, J.M., Fersht, A.R. α-Helix stability in proteins. II. Factors that influence stability at an internal position. J. Mol. Biol. 227:560–568, 1992.
7. Lyu, P.C., Liff, M.I., Marky, L.A., Kallenbach, N.R. Side chain contributions to the stability of alpha-helical structure in peptides. Science 250:669–673, 1990.
8. O'Neil, K., DeGrado, W. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science 250:646–651, 1990.
9. 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.
10. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
11. Jorgensen, W.L., Tirado-Rives, J. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and Crambin. J. Am. Chem. Soc. 110:1657–1666, 1988.
12. Brünger, A.T. "X-PLOR, Version 3.1. A System for Crystallography and NMR." New Haven: Yale University Press, 1992.
13. Janz, G.J. "Estimation of Thermodynamic Properties of Organic Compounds." New York: Academic Press, 1958: 26–28.
14. Pickett, S.D., Sternberg, M.J.E. Empirical scale of side-chain conformational entropy in protein folding. J. Mol. Biol. 231:825–839, 1993.
15. Abagyan, R., Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J. Mol. Biol. 235:983–1002, 1994.
16. Garcia, K.C., Ronco, P.M., Verroust, P.J., Brumger, A.T., Amzel, L.M. Three dimensional structure of an angiotensin II-Fab complex at 3 Å: Horomone recognition by an anti-idotypic antibody. Science 257:502–507, 1992.
17. Lee, B., Richards, F.M. The Interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55: 379–400, 1971.
18. Ben-Naim, A. J. The role of hydrogen bonds in protein folding and protein association. Phys. Chem. 95:1437–1444, 1991.
19. Ooi, T., Oobatake, M. Prediction of the thermodynamic of protein unfolding: The helix-coil transition of poly(L-alanine). Proc. Natl. Acad. Sci. USA 88:2859–2863, 1991.
20. Rialdi, G., Hermans, J.J. J. Am) Chem. Soc. 88:5719–5720, 1966.
21. Scholtz, J.M., Baldwin, R.L. The mechanism of alpha-helix formation by peptides. Annu. Rev. Biophys. Biomol. Struct. 21:95–118, 1992.
Scholtz, J.M., Marquese, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M., Bolen, D.W. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. Proc. Natl. Acad. Sci. USA 88:2854–2858, 1991.
22. Baldwin, R.L. Temperature dependence of the hydrophobic interaction in protein folding. Proc. Natl. Acad. Sci. USA 83:8069–8072, 1986.
23. Schellman, J.A. Thermodynamics of urea solutions and the heat of formation of the peptide hydrogen bond. C.R. Trav. Lab. Carlsberg Ser. Chim. 29:223–229, 1955.
24. Nemethy, G., Leach, S.J., Scheraga, H.A. The influence of amino acid side chains on the free energy of helix-coil transitions. J. Phys. Chem. 70:998–1004, 1966.
25. Freire, E., Haynie, D.T., Xie, D. Molecular basis of cooperativity in protein folding. IV. CORE: A general cooperative folding model. Proteins 17:111–123, 1993.
26. Murphy, K.P., Xie, D., Thompson, K., Amzel, L.M., Freire, E. Entropy in biological binding processes: Estimation of translational entropy loss. Proteins in press, 1994.
27. Pace, C.N., Grinsley, G.R., Thomson, J.A., Barnett, B.J. Conformational stability and activity of ribonuclease T1 with zero, one and two intact disulfide bonds. J. Biol. Chem. 263:11820–11825, 1988.
28. Klapper, M.H. The independent distribution of amino acid near neighbor pairs into polypeptides. Biochem. Biophys. Res. Commun. 78:1018–1024, 1977.
29. Oobatake, M., Ooi, T. Hydration and heat stability effects on protein unfolding. Prog. Biophys. Molec. Biol. 59:237–284, 1992.
30. Privalov, P.L., Makhatadze, G.I. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. J. Mol. Biol. 232:660–679, 1993.
31. Privalov, P.L., Khechinashvili, N.N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. J. Mol. Biol. 86:665–684, 1974.