PROTEINS: Structure, Function, and Genetics 36:332–339 (1999)

# Estimates of the Loss of Main-Chain Conformational Entropy of Different Residues on Protein Folding

**Debnath Pal and Pinak Chakrabarti***
*Department of Biochemistry, Bose Institute, Calcutta, India*

***ABSTRACT*** **The average contribution of conformational entropy for individual amino acid residues towards the free energy of protein folding is not well understood. We have developed empirical scales for the loss of the main-chain (torsion angles, ϕ and ψ) conformational entropy by taking its side-chain into account. The analysis shows that the main-chain component of the total conformational entropy loss for a residue is significant and reflects intrinsic characteristics associated with individual residues. The values have direct correlation with the hydrophobicity values and this has important bearing on the folding process. Proteins 1999;36:332–339.**


**Key words: conformation; conformational entropy; hydrophobicity; protein folding; protein engineering**

## INTRODUCTION

When a protein folds into a compact globule, the residues lose degrees of freedom as lesser number of conformations can be accessed by the main- and the side-chain—this reduction in conformational entropy opposes the folding process.[1,2] On the other hand, water molecules released from contact with nonpolar side-chains (that get buried in the folded state) gain in entropy. These are the source of the hydrophobic effect, which is a major driving force for folding.[3,4] However, as most globular proteins are only marginally stable (the free energy for a folding-unfolding reaction is around 5–20 kcal/mol)[5] it appears that the conformational entropy is the prime deterrent to folding.[6] While the calculation/measurement of the free energy associated with hydrophobicity has received wide attention, Cornette et al.[7] reporting a comparison of 38 different scales, the quantification of the conformational entropy changes has been rare and usually computation-intensive.[1,8–13] Some recent work in this area have determined scales for the conformational entropy change of side-chains during protein folding.[14–19] The backbone gets neglected in such studies, presumably with the assumption that the restriction on the conformation of the main-chain on folding is the same for all residues, and hence its inclusion will result in the addition of a constant term to all the entropy values such that the relative scale remains the same. D'Aquino et al.[20] have recently shown experimentally that this assumption is not valid and that the presence of the methyl group in alanine reduces the conformational entropy of the peptide backbone by 2.5 cal/K.mol with respect to that of glycine. They further substantiated their result by computing the energy profile of the backbone conformations as a function of the main-chain dihedral angles, ϕ and ψ. However, as discussed below, the ϕ,ψ angles can be influenced by the conformation of the side-chain of individual residues, and this needs to be taken into account while calculating the conformational entropy of the main-chain. Here we report such a calculation to show that the main-chain of each residue makes a distinct contribution towards the loss of conformational entropy on folding, and that the values are related to hydrophobicity.

### Influence of the Side-Chain

That each amino acid residue leaves its imprint on the protein structure on the basis of the chemical nature of its side-chain, is undisputed. However, at a more subtle level the side-chain can also influence the main-chain conformation of the residue,[21] and the assumption that the ϕ,ψ space is independent of the residue type while calculating the relative entropy values[15] is not justified. Because of the interdependence between the main-chain and the side-chain torsion angles,[21–23] a residue is represented not by a single ϕ,ψ distribution, but by distributions at three discreet values of $\chi_1$ (−60, 180 and 60°, which are designated as $g^+$, t and $g^-$ conformational states, respectively), or better yet, one should consider the distribution in the three-dimensional space (ϕ,ψ,$\chi_1$). Additionally, instead of considering the influence of the side-chain up to the γ position only ($\chi_1$-torsion), the effect of the whole side-chain can also be taken into account by calculating the ϕ,ψ distribution for all the rotameric states of the side-chain. All these methods to incorporate the effect of the side-chain while calculating the main-chain conformational entropy have been implemented by us.

### Folded and Unfolded States

If one attempts to calculate the change in conformational entropy in bringing a particular residue to the folded state, the unique conformation of the residue requires its entropy in the folded state to be set to zero.[15] On the other hand, instead of dealing with a residue in specific location in a protein structure, if we are interested in the

*Correspondence to: Pinak Chakrabarti, Department of Biochemistry, P-1/12 CIT Scheme VIIM, Bose Institute, Calcutta, 700 054, India. E-mail: pinak@boseinst.ernet.in
Received 17 December 1998; Accepted 8 April 1999

average value for all residues of the same type then we have to consider the distribution of the conformational angles of the amino acid residue both in the folded and the unfolded states. For example, Creamer and Rose,[14] while investigating the difference in conformational entropy between a series of apolar side-chains in the unfolded state and in the α-helical state, obtained these two distributions using Monte Carlo calculations. Although the rotational states that are being sampled by the main-chain in the folded state are directly available from the statistical analysis of known structures, the information on the unfolded state is not directly available. As a result, scales of main-chain entropy changes during folding have been proposed relative to the value for Gly (considering ϕ,ψ distributions in native states only).[8,24] Below we propose a procedure to fill in the lacuna on the unfolded state.

We have recently shown[21] that amino acid residues (barring Gly, Ala and Pro) can be classified into five classes depending on how the ϕ,ψ distribution of individual residues get affected by a change of the side-chain torsion angle $\chi_1$. These are: Class I: Ser, Cys, Met, Glu, Gln, Lys and Arg; II: Leu; III: Asp and Asn; IV: His, Phe, Tyr and Trp; and V: Val, Ile and Thr. The classification is also chemically intuitive. Residues in a given class have similar topological arrangements at least up to the γ-position, beyond which the side-chain atoms do not significantly interact with the main-chain atoms, and are non-invasive as regard to the ϕ,ψ distribution.[21] For example, Class I has linear side-chains (up to γ-position), whereas Class V has aliphatic chains branched at $C_\beta$. As steric clash is the limiting factor that affects the ϕ,ψ angles in the unfolded state, one can safely assume that the ranges of the ϕ,ψ angles that all residues in a given class span in the unfolded state are identical. Because of the additional short range interactions (like hydrogen bonding) that are brought into play in the folded state (for example, due to the presence of different atoms, as in Val and Thr), the distribution of angles within this allowed ϕ,ψ space (in the unfolded state) becomes different for different residues in a class. Conversely then, the combined distribution of the ϕ,ψ (and $\chi_1$) values for all the members in a class can be taken as the representation of the distribution in the unfolded state for any member of the class. This translates to the assumption that in the unfolded state a residue cannot assume a ($\phi,\psi,\chi_1$) value not accessible to any member of the class in the native state, and that the distribution is random. Further justification for this assumption is provided by the energy calculations given in the next section.

## Theory

The change in the conformational entropy, ΔS, for a main-chain taken from the unfolded state (U) to the folded (F) state is given by:

$$\Delta S = S(F) - S(U).$$

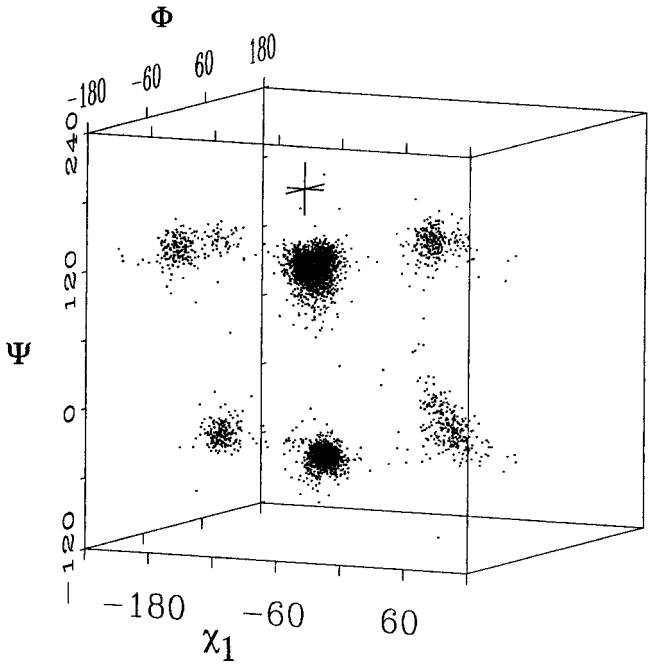


Fig. 1. Distribution of points corresponding to all Ile residues in the three-dimensional space. For one cluster the standard deviations (magnified four times) of the distribution of points are shown, which are the ellipsoidal axes used for the calculation of volume in Method 1.

In one of the earliest studies Némethy et al.[8] estimated the restrictions on folding of either the main-chain or the side-chain using:

$$\Delta S = -R \ln\left(\frac{\text{accessible conformational space unfolded}}{\text{accessible conformational space folded}}\right). \quad (1)$$

Although they used the area spanned by the torsion angles ϕ, ψ to determine the main-chain entropy changes during folding, we have used the equation by considering the volume occupied by different residues in the ($\phi,\psi,\chi_1$) space. Depending on the distribution of points, the 3-dimensional space is divided into a number of regions, such that each region encompasses a cluster of points (Fig. 1). For a given residue the mean and the standard deviation (σ) of the ϕ,ψ and $\chi_1$ values in a cluster are calculated.[21] The product of the three σ values gives the volume of an ellipsoid that provides an estimate of the dispersion of points in the cluster. The summation of these ellipsoidal volumes over all the regions give the conformational space available to the residue in the folded state. The same procedure repeated using the combined distribution of points for all the members in a class gives the value for the unfolded state of a member residue.

In a second method, we have directly used the distribution of points based on the definition of entropy[25] as the

Boltzmann sampling over all accessible states,

$$S = -R \sum_{i=1}^{N} p_i \ln p_i \quad (2)$$

where $p_i$ is the probability of the main-chain in state i (where each state is represented by a grid of fixed size in the ($\phi$,$\psi$,$\chi_1$) space (Fig. 1), the sum being performed over all the grid points (N) in the range of 360° along the three axes); R is the gas constant. As earlier, S(U) is obtained from the combined distribution of all the residues belonging to a class.

As Leu is the only member of the class, the above procedure of combining the $\phi$,$\psi$,$\chi_1$ distributions of the constituent class members to generate the unfolded state fails. Moreover, two residues (Gly and Ala) are left out in the above procedure (besides Pro, for which the main-chain $\phi$ and the side-chain torsion angles are restricted because of the pyrrolidine ring, and is not considered here). As a result, there is a need for an independent method to simulate the unfolded state for all the residues, and for this we have taken recourse to energy calculation in the $\phi$,$\psi$ space.

The conformational features of the unfolded state can also be obtained from the fact that when unfolded, the conformation of a residue is influenced only by short-range interactions which are adequately quantified through force-field calculations. The $\phi$,$\psi$ map obtained by Ramachandran and coworkers[26] is based on the simple assumption that the torsion angles that lead to steric clash are not permissible. This assumption is equally applicable both for the native and the denatured state, so that short range steric clash makes a portion of the whole map inaccessible even in the denatured state. Because of various secondary and tertiary interactions, residues in the native state are not spread all over the allowed region in the $\phi$,$\psi$ map, but occupy only a part of it. However, in the denatured state the main-chain torsion angles can be expected to span the whole space within the bounds of steep energy gradient. Consequently, we have calculated the $\phi$,$\psi$ energy maps for all the different combinations of the side-chain torsion angles (rotamers) of a given residue (X) in a tripeptide, Ala-X-Ala. For a given map, starting from the minimum energy value, contours are drawn at a fixed increment of energy and the area of the map enclosed by each contour level is calculated and compared with the area in the previous level. This process is repeated until a steep energy gradient is reached and no significant increase in area is observed. This area can then be taken as the maximum area that is accessible to the residue for the rotameric state under consideration. The area accessible in the folded state is obtained from the distribution of crystallographic data points for the residue in the same rotameric state (Fig. 2). Equation (1) then gives the $\Delta S$ value for the main-chain of the rotamer, and the average $\Delta S$ value can be obtained by considering all the rotamers.

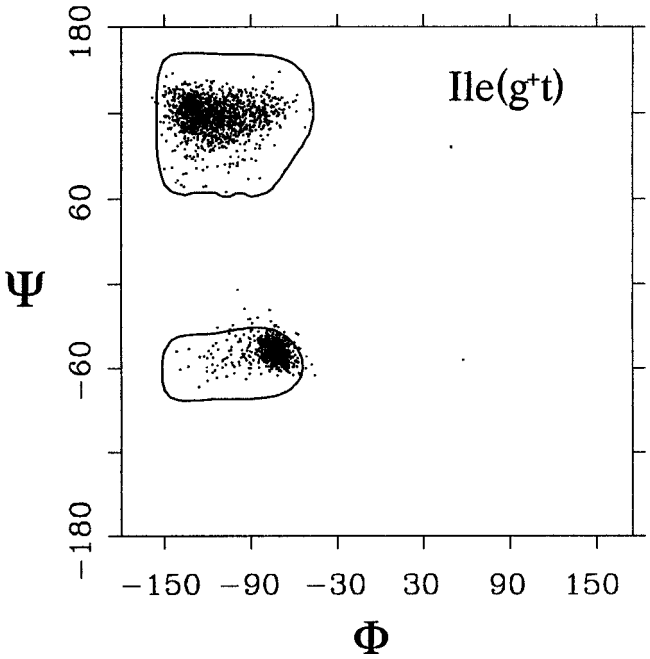


Fig. 2. Boundary obtained by energy calculation (Method 3, for the unfolded state) shown superimposed on the distribution of points corresponding to Ile in the rotameric state $g^+t$.

## METHODS

Two hundred ninety-four X-ray crystallographic protein coordinate files taken from Brookhaven Protein Data Bank (PDB),[27] 1998 version, having resolution ≤2.0 Å and R-factor ≤0.2 were selected conforming to <25% sequence homology criterion.[28] For proteins having more than one subunit, only those subunits were taken as prescribed in the list available at http://www.sander.embl-heidelberg.de. Torsion angles of the main-chain ($\phi$ and $\psi$) and side-chain ($\chi$) were calculated using the program DIHDRL provided by PDB. Small screening factors, discussed below, depending on the density/number of data points in a region/grid, were applied to get consistent entropy values in different methods.

### Method 1

It was found that for all the residues there were a maximum of twelve distinct regions in the ($\phi$,$\psi$,$\chi_1$) space where the points are clustered. The dispersion of the points in each region can be approximated by the standard deviation associated with the mean of each variable within the region (Fig. 1). The regions are defined[21] by the combination of the following ranges of $\phi$,$\psi$ and $\chi_1$:

| $\phi$(°) | $\psi$(°) | $\chi_1$(°) |
|---|---|---|
| −180,0 | −120,60 | −240,−120 (*t*) |
| | 60,240 | −120,0 (*g*+) |
| | | 0,120 (*g*−) |
| 0,180 | −90,90 | −240,−120 (*t*) |
| | 90,270 | −120,0 (*g*+) |
| | | 0,120 (*g*−) |

The product of standard deviations for each variable in a given range gave the volume dispersion of that cluster. The sum of all the clusters gave the total accessible volume for the residue in the 3-dimensional space. The dispersion volume for all the residues individually (except Ala, Gly and Pro) and combining them into classes (for the unfolded state) were obtained and the entropy values calculated using equation (1).

While combining the occupied regions in the ($\phi$,$\psi$,$\chi_1$) space of different residues to get the value for a class it was assumed that in the unfolded state these regions were randomly populated. As such, even though the distribution of points in the individual 10° × 10° × 10° grids within the occupied regions of a given residue may vary, only one point was retained per occupied grid per residue while getting the distribution (and subsequent calculation of the standard deviations) for the class in the unfolded state.

If a region was sparsely populated (containing <1% of the total data points), it was not considered; this cut-off was found to be necessary as each region gets equal weight in this calculation and a region with a very negligible number of widely dispersed points lead to a spuriously large volume contribution. As an example, for the aromatic residues in Class IV, only 7 out of 12 regions contributed to the volume calculation for the class, as well as the individual residues.

### Method 2

Torsion angles $\phi$,$\psi$,$\chi_1$ of all residues were plotted in a 3-dimensional Cartesian axes having ranges 0 through 360°. The fraction of the number of points ($N_i$) in a 10° × 10° × 10° grid to the total number ($N_t$) of points of a residue gave the probability $p_i$ for the evaluation of entropy in the folded state using equation (2). The choice of a 10° grid size has been found to be suitable in an earlier study.[24] As discussed under Method 1, the condition that the whole occupied region is randomly accessible to a given residue in the denatured state was imposed by assuming all occupied grids to have just one point. This modified distribution was then combined to get the distribution for the class for the calculation of entropy in the denatured state.

### Method 3

The nonbonded-energy calculations over the whole $\phi$,$\psi$ range were performed for the sequence Ala-X-Ala (the peptide terminals being capped by protons), X being any residue, using the CVFF[29] forcefield as implemented in the program DISCOVER_3 (Biosym/MSI, San Diego, CA). Acidic and basic residues were considered with appropriate charges. Calculations were carried out by restraining the side-chain at discreet values of $\chi_1$ (and beyond); all the rotameric states were considered. As $\chi_1$ can have slightly different values at different regions of the $\phi$,$\psi$ map,[21] we found out the average of all the $\chi$ torsion angles of a residue when the main-chain conformation is in one of the 4 regions given above; however, if the mean value could not evaluated because of the paucity of data, the torsion angle was placed at the canonical value of ±60 or 180°. Next, the energy values within the individual regions were computed (at 5° intervals) by fixing the side-chain at these average values, and the total map was constructed.

Starting with the minimum energy ($E_m$) value in a map, contours were drawn at 5 kcal/mol intervals (this value is approximately 1/10th the value of the standard deviation of the average of all $E_m$ values), and areas enclosed (as the percentage of the total map-area) were found out. If a given increment increased the area by less than 10% of the map area then the area at that step was accepted as the accessible area for the rotamer in the denatured state (Fig. 2). For Cys, Ser and Thr, this area was found to be profoundly affected by the orientation of the -SH or -OH proton in the side-chain. As there is no way of ascertaining the position of the proton from crystallographic method, we have used six idealized orientations (given by the torsion angle, $C_\alpha$-$C_\beta$-O/$S_\gamma$-H, with values 0, ±60, ±120 and 180°) for each value of $\chi_1$ and the average value of the areas was accepted.

For the folded state of a residue, the $\phi$,$\psi$ values for each rotamer were plotted and the map was divided into squares 10° on a side. If a grid contained more than 0.25% of the data points it was assumed to be occupied (for Gly, however, as the points are distributed in all the four quadrants of the map, a threshold value of 0.1% was used) and the total number of such grids gave the accessible area. (On average, about 96% data points are enclosed in this area). If the population of a rotamer is less than 0.5% of the total number (78,662) of residues in our database, the size of the grid was increased to 20° and the threshold value for occupancy was changed to 1%.

The change in the conformational entropy for each rotamer of a residue was then calculated using equation (1), and the weighted $\Delta S$ value,

$$\Delta S = \frac{\sum_r N_r \Delta S_r}{\sum_r N_r}$$

was found out; $N_r$ is the number of data points for the rotamer that defined the accessible area; the summation is over all rotamers.

## RESULTS AND DISCUSSION

### Scales of the Main-Chain Conformational Entropy Values

Tables I and II provide the values for the average loss of main-chain conformational entropy, obtained by three different methods. As different considerations have gone into the calculation of each set of values the concurrence between them is remarkable (Table III). While the side-chain torsion, $\chi_1$, is directly used in Methods 1 and 2, the effect of the whole side-chain is taken into account, albeit indirectly, in Method 3. Also, the agreement between Set 3 and the other two sets suggests that the unfolded state

**TABLE I. Conformational Entropy Values Using Methods 1 and 2**

| Residue | | Set 1 | | Set 2 Entropy | | TΔS (kcal/mol at 300 K) | |
|---|---|---|---|---|---|---|---|
| Name | % of total population[a] | Volume[b] occupied | Entropy (ΔS/R) | Absolute (S/R) | Relative (ΔS/R) | Set 1 | Set 2 |
| a) Class I | | 4.13 | | 7.94 | | | |
| Cys | 1.42 | 1.99 | −0.73 | 6.23 | −1.72 | −0.43 | −1.02 |
| Ser | 6.13 | 2.66 | −0.44 | 6.84 | −1.10 | −0.26 | −0.66 |
| Met | 1.97 | 2.10 | −0.68 | 5.98 | −1.97 | −0.41 | −1.17 |
| Glu | 5.72 | 2.43 | −0.53 | 6.41 | −1.53 | −0.32 | −0.91 |
| Gln | 3.67 | 2.17 | −0.64 | 6.26 | −1.68 | −0.30 | −1.00 |
| Lys | 5.83 | 2.33 | −0.57 | 6.62 | −1.32 | −0.34 | −0.79 |
| Arg | 4.36 | 2.28 | −0.59 | 6.49 | −1.45 | −0.35 | −0.86 |
| b) Class II | | 2.50 | | 7.72 | | | |
| Leu[c] | 8.10 | 0.97 | −0.95 | 6.24 | −1.48 | −0.57 | −0.88 |
| c) Class III | | 4.89 | | 7.68 | | | |
| Asp | 6.18 | 2.74 | −0.58 | 6.62 | −1.06 | −0.35 | −0.63 |
| Asn | 4.79 | 3.04 | −0.47 | 6.80 | −0.88 | −0.28 | −0.52 |
| d) Class IV | | 2.71 | | 7.41 | | | |
| His | 2.26 | 2.22 | −0.20 | 6.47 | −0.95 | −0.12 | −0.57 |
| Phe | 4.10 | 1.62 | −0.51 | 6.36 | −1.05 | −0.30 | −0.63 |
| Tyr | 3.81 | 1.69 | −0.47 | 6.45 | −0.96 | −0.28 | −0.57 |
| Trp | 1.60 | 1.59 | −0.53 | 6.09 | −1.33 | −0.32 | −0.79 |
| e) Class V | | 3.09 | | 7.46 | | | |
| Val | 7.02 | 1.47 | −0.74 | 5.81 | −1.65 | −0.44 | −0.98 |
| Ile | 5.43 | 1.00 | −1.13 | 5.60 | −1.86 | −0.67 | −1.11 |
| Thr | 5.96 | 2.04 | −0.41 | 6.32 | −1.14 | −0.24 | −0.68 |

[a]The values for Gly, Ala and Pro are 8.23, 8.67 and 4.75, respectively.
[b]Given by $\Sigma\, \sigma_\phi * \sigma_\psi * \sigma_{\chi 1}$, where σs (in deg) are the standard deviations of different clusters in the (ϕ, ψ, χ1) space, the summation being over all the clusters, as given in Methods. The individual values are divided by that of Ile which has the smallest volume (not considering Leu, for reasons given below).
[c]As Leu is the only member of class II, its unfolded state has been simulated by putting it along with all other members of class I and excluding the $g^-$ conformation from the calculation. This is because as the branching in the side-chain is at the γ-position the χ1-dependence of the ϕ, ψ map of Leu resembles that of the unbranched residues of class I, except that its side-chain is hardly found in the $g^-$ conformation.[21]

**TABLE II. Conformational Entropy Values (Set 3) Using Method 3**

| Residue | No. of possible rotamers | Most populous rotamer | | | | Entropy (ΔS/R) | TΔS (kcal/mol at 300 K) | |
|---|---|---|---|---|---|---|---|---|
| | | Name | % of total population | % of the (ϕ, ψ) map occupied | | | | |
| | | | | Observed | Calculated | | Calculated | Scaled[a] |
| Gly | | | | 17.90 | 57.93 | −1.17 | −0.70 | −1.41 |
| Ala | | | | 5.63 | 19.19 | −1.23 | −0.73 | −1.45 |
| Cys | 3 | $g^+$ | 57 | 7.87 | 14.63 | −0.69 | −0.41 | −1.03 |
| Ser | 3 | $g^-$ | 47 | 6.10 | 17.61 | −0.95 | −0.57 | −1.24 |
| Met | 27 | $g^+g^+g^+$ | 16 | 6.17 | 14.93 | −0.76 | −0.45 | −1.08 |
| Glu | 27 | $g^+tg^+$ | 18 | 7.64 | 13.46 | −0.72 | −0.43 | −1.05 |
| Gln | 27 | $g^+tg^-$ | 14 | 4.94 | 14.24 | −0.86 | −0.51 | −1.16 |
| Lys | 81 | $g^+ttt$ | 18 | 5.56 | 10.8 | −0.57 | −0.34 | −0.93 |
| Arg | 81 | $g^+ttt$ | 11 | 8.02 | 11.25 | −0.59 | −0.35 | −0.95 |
| Leu | 9 | $g^+t$ | 56 | 6.10 | 13.45 | −0.80 | −0.48 | −1.12 |
| Asp | 9 | $g^+g^+$ | 43 | 5.79 | 10.19 | −0.67 | −0.40 | −1.01 |
| Asn | 9 | $g^+g^+$ | 33 | 6.64 | 12.81 | −0.63 | −0.38 | −0.99 |
| His | 9 | $g^+g^+$ | 28 | 7.48 | 9.39 | −0.17 | −0.10 | −0.61 |
| Phe | 6 | $g^+g^{+/-}$ | 42 | 8.02 | 13.37 | −0.58 | −0.35 | −0.95 |
| Tyr | 6 | $g^+g^{+/-}$ | 42 | 7.33 | 13.29 | −0.61 | −0.36 | −0.96 |
| Trp | 9 | $g^+g^-$ | 34 | 5.40 | 13.39 | −0.67 | −0.40 | −1.01 |
| Val | 3 | $g^+$ | 73 | 4.48 | 11.57 | −0.87 | −0.52 | −1.17 |
| Ile | 9 | $g^+t$ | 58 | 3.78 | 9.41 | −0.78 | −0.46 | −1.09 |
| Thr | 3 | $g^-$ | 47 | 6.48 | 13.62 | −0.84 | −0.50 | −1.15 |

[a]Scaled to the values of Set 2, using the equation $y = 1.33x - 0.48$, obtained by regression analysis between Set 2 (y) and Set 3 (x) (after removing Ser, Cys and Thr, as given in Table III).

**TABLE III. Correlation Coefficients Between the Three Sets of Conformational Entropy Values**

| | Set 2 | Set 3 |
|---|---|---|
| Set 1 | 0.73 | 0.49 (0.73) |
| Set 2 | | 0.44 (0.66) |

Values in brackets are obtained when Ser, Thr and Cys are omitted from calculations.

obtained from the energy calculation is similar to the one derived from the combined distribution of a class of residues, thereby lending a justification, based on energetics, for the earlier classification[21] of amino acid residues.

Ser, Thr and Cys have maximum discrepancies in the Set 3 values compared to the other sets. This is because the results of the energy calculation for these residues are highly sensitive to the location of the terminal proton of the side-chain. The calculated area can change by as much as 100% depending on the conformation (that defines the position of the proton) selected (see Methods). In the ensuing discussion we use the Set 2 values (along with those of Gly and Ala from Set 3 scaled to Set 2) as these values are obtained directly from the observed distribution.

Although not used in the discussion, our aim for employing Method 1 was to see if a simple volume calculation using three $\sigma$ values can give result comparable to other methods. One may argue that this ellipsoid enclosing the distribution of $\phi,\psi,\chi_1$ points is aligned with the three axis, whereas a more sophisticated method would be to calculate intertial tensor of the points and then use the deduced principal axes of an enclosing nonaligned ellipsoid for the volume calculation. As can be seen, $\Delta S$ values calculated by Method 1 are reasonable, and besides the volumes given in Table I, are also correlated with other parameters like the $\beta$-sheet propensities of residues (manuscript under preparation). This method has the added advantage that one may also include the influence of other side-chain torsion angles, say $\chi_2$ by considering a hypervolume consisting of $\phi,\psi,\chi_1$ and $\chi_2$.

### Substitution of Gly by Ala

Gly and Ala are bypassed in all work dealing with the side-chain conformational entropy. But even without any side-chain torsion angle they do have large and comparable negative $\Delta S$ values, which has important connotation for protein engineering experiments aimed at enhancing protein thermostability. Matthews et al.[30] have proposed that substitutions which decrease the chain entropy of the unfolded state might shift the equilibrium to the folded state, provided there are no enthalpic penalties for these changes in the folded state. It was assumed that Gly residues would lose the maximum conformational entropy upon folding and its substitution, even by a slightly larger residue like Ala, would be stabilizing, as indeed was the case in many instances.[20,30–37] However, inasmuch as $\Delta S$ for Gly and Ala are nearly the same, the degree to which the conformational flexibility of the main-chain is reduced due to a change from a Gly to an Ala residue is similar both in the denatured and the folded states. Consequently, our data do not support the hypothesis that the higher protein stability arising out of such mutations is due to entropic effect; rather the enthalpic contribution involving the methyl group may also be important.

This result needs to be put in proper perspective. Our method of finding out the maximum area that is accessible to a residue in the unfolded state is similar to Ramachandran's method of delineating the contact map,[26] *i.e.*, the region devoid of any steric clash. Gly residues in known structures, however, cover only a limited portion of this allowed region as can be seen in the diagrams of Ramakrishnan and Srinivasan.[38] Even for Ala, the points are not distributed for the whole allowed region and indeed Stites and Pranata[24] have commented that of all residues (except Gly and Pro), Ala is the most ordered. Our result quantifies the reduction in the area over which the $\phi,\psi$ points in the known structures are distributed as compared to the total allowed region, and the values obtained for Gly and Ala are nearly identical. The relative $\Delta S$ values enumerated here provide guidelines for the judicious substitutions that would decrease the conformational entropy of unfolding and may thereby stabilize the native structure.

### Main-Chain Conformational Entropy, Hydrophobicity and Protein Folding

Inasmuch as the contribution of the side-chain torsion $\chi_1$ is taken into account in our calculation, the entropy values reported here should approximate the value for the whole residue if it does not have any torsion angle beyond $\chi_1$. Moreover, it is the main-chain that determines the protein fold, and it is highly plausible that at the initial stages of folding, the main-chain atoms and the close side-chain atoms ($\chi_1$ torsion) get restrained, leaving the side-chain atoms further off to retain their full flexibility characteristic of the unfolded state. As such, even for the longer residues (with $\chi_2$ and beyond) the estimated entropy values should correspond closely to the loss in the conformational entropy for the whole residue as these are folding to make the protein core. Additionally, side-chains of residues on protein surface are highly mobile (so much so that the distribution of their side-chain conformations may be what one expects in the unfolded state[15]) and for these residues, our $\Delta S$ values would be the major component of the overall value.

In order to analyse if the entropy values reflect the residue hydrophobicities, we have compared the $-T\Delta S$ values with a few hydrophobicity scales (Table IV). Values are in good agreement with the scales due to Kyte-Doolittle and Wolfenden (Fig. 3). Some workers have considered hydrophobicity and side-chain entropy to be distinct entities.[19] However, as there is a substantial correlation between the main-chain conformation entropy and hydrophobicity it can be surmised that both get a major contribution from a common factor. In case of

**TABLE IV. Correlation Coefficients Between Conformational Entropy Values (−ΔS) and Some Representative Hydrophobicity Data**

| Hydrophobicity scale of | Set 1 | Set 2 | Set 2 + Gly + Ala[a] |
|---|---|---|---|
| Fauchère and Pliska[40] | 0.44 (0.53) | 0.32 (0.60) | 0.20 (0.30) |
| Kyte and Doolittle[41] | 0.65 (0.78) | 0.47 (0.77) | 0.46 (0.63) |
| Miller et al.[42] | 0.38 (0.59) | 0.30 (0.77) | 0.32 (0.59) |
| Ponnuswamy et al.[43] | 0.55 (0.63) | 0.52 (0.79) | 0.36 (0.46) |
| Wolfenden et al.[44] | 0.55 (0.77) | 0.37 (0.75) | 0.51 (0.80) |

Values obtained on exclusion of Arg, Lys, Glu and Gln are given in parentheses.
[a]Values for Gly and Ala are from Set 3 scaled to Set 2 (last column of Table II).

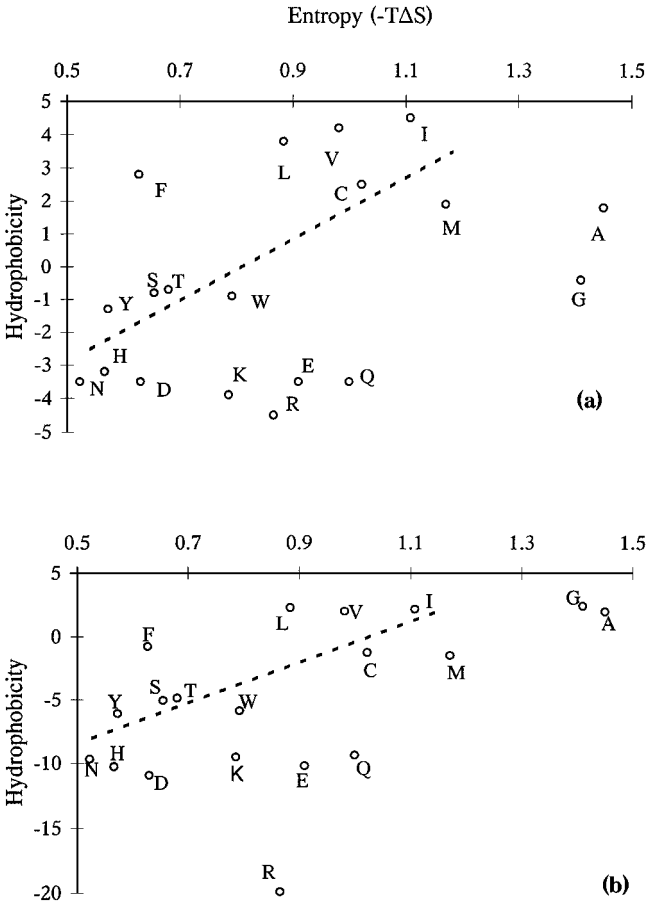


Fig. 3. Plot of the loss of main-chain conformational entropy (−TΔS values, Set 2 + Gly + Ala, explained in Table IV footnote) against the hydrophobicity values (both in kcal/mol) due to (**a**) Kyte and Doolittle, and (**b**) Wolfenden. The least-squares lines [(a) y = 10.30x −7.64; (b) y = 15.93x −16.35] are obtained by excluding Gly, Ala, Arg, Lys, Glu and Gln from the regression analysis; the corresponding correlation coefficients are provided in Table IV, third column.

hydrophobicity it is water, and it is likely that water also contributes to ΔS values, albeit in an indirect manner. Hydrophobicity is determined by the chemical nature of the side-chain and its surface area, which in turn also control the ($\phi$,$\psi$,$\chi_1$) values that a particular residue can adopt, and thus the main-chain entropy.

Residues that show large deviations from the fitted lines in Figure 3 are Arg, Lys, Glu and Gln. Interestingly, however, these are the very residues for which the hydrophobicity values derived by various methods also show wide variations.[39] These residues have large components of both polar and nonpolar atoms which may contribute differently depending on the systems used to generate the hydrophobicity scales, whereas in our calculations, it is the nearby nonpolar part that is likely to affect the main-chain conformation and thereby the resultant entropy values.

It is remarkable that the hydrophobic residues (Fig. 3), and Gly and Ala (Table II) show the maximum reduction of the main-chain conformational entropy on folding, unlike the earlier studies concerned only with the side-chain entropies which showed, as expected, that the entropy values depend on the length of the side-chain. As these constitute a major proportion of amino-acid residues in a protein (Table I), ΔS values will have a pronounced influence in the folding process. In the hydrophobic collapse model,[5] it is the hydrophobic residues which initiate the folding process, but our study indicates that the main-chain of these same residues get the maximum resistance, owing to conformational entropy to take up the native form. This may be beneficial for a productive folding process as it means that specific non-covalent interactions (like hydrogen bonding) must start to be formed at this stage to overcome the entropic cost. Without the countering influence from ΔS the hydrophobic residues might aggregate, leading to non-native like forms. Seen from this angle the term "hydrophobic collapse" may be a too severe term, as it does not adequately convey the opposition of the main-chain entropy of the hydrophobic residues.

## CONCLUDING REMARKS

In summary, any attempt to elucidate thermodynamic parameters neglecting the main-chain on account of its invariance within the naturally occurring amino acid residues may not be correct; although chemically equivalent, the backbones from different residues have varying degree of conformational restrictions imposed by the side-chain. The main hindrance for the calculation of the loss in the main-chain conformational entropy has been the lack of any concrete data on the possible conformations in the denatured state. We have overcome this by energy calculations, and by combining the distribution of residues that show similar $\chi_1$-dependence of the $\phi$,$\psi$ plot.[21] Hydrophobic residues, and Gly and Ala have larger entropy values with important implications for the folding process. The derived entropy values can be used for protein stabilization by rational amino acid replacements based on the concept of entropic stabilization.[30]

## ACKNOWLEDGMENTS

We thank the Council of Scientific and Industrial Research for a fellowship to D.P., the Department of Science and Technology for a grant to P.C., and the Department of

Biotechnology for the National Facility for Biomolecular Modeling.

## REFERENCES


1. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 1989;3:1–3.
2. Brady GP, Sharp KA. Entropy in protein folding and in protein-protein interactions. Curr Opin Struct Biol 1997;7:215–221.
3. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
4. Chothia C. Hydrophobic bonding and accessible surface area in proteins. Nature (London) 1974;248:338–339.
5. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
6. Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. Adv Protein Chem 1988;39:191–234.
7. Cornette JL, Cease KB, Margalit H, Spouge JL, Berzsofsky JA, DeLisi C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J Mol Biol 1987;195: 659–685.
8. Némethy G, Leach SJ, Scheraga HA. The influence of amino acid side-chains on the free energy of helix-coil transitions. J Phys Chem 1966;70:998–1004.
9. Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. Macromolecules 1981;14:325–332.
10. Di Nola A, Berendsen HJC, Edholm O. Free energy determination of polypeptide conformations generated by molecular dynamics. Macromolecules 1984;17:2044–2050.
11. Brady J, Karplus M. Configuration entropy of the alanine dipeptide in vacuum and in solution: a molecular dynamics study. J Am Chem Soc 1985;107:6013–6015.
12. Meirovitch H, Kitson DH, Hagler AT. Computer simulation of the entropy of polypeptides using the local states method: application to cyclo-(ala-pro-D-phe)$_2$ in vacuum and in the crystal. J Am Chem Soc 1992;114:5386–5399.
13. Novotny J, Bruccoleri RE, Saul FA. On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3 and Hy-HEL-5. Biochemistry 1989;28:4735–4749.
14. Creamer TP, Rose GD. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix forming propensities. Proc Natl Acad Sci USA 1992;89:5937–5941.
15. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
16. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side-chain configurational entropy in binding and folding: General methods and application to helix formation. Proteins 1994;20:68–84.
17. Blaber M, Zhang X, Lindstorm JD, Pepiot SD, Baase WA, Mathews BW. Determination of α-helix propensity within the context of a folded protein. Sites 44 and 131 in bacteriophage T4 lysozyme. J Mol Biol 1994;235:600–624.
18. Koehl P, Delarue M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J Mol Biol 1994;239:249–275.
19. Sternberg MJE, Chickos JS. Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales. Protein Eng 1994;7:149–155.
20. D'Aquino JA, Gomes J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. Proteins 1996;25:143–156.
21. Chakrabarti P, Pal D. Main-chain conformational features at different conformations of the side-chains in proteins. Protein Eng 1998;11:631–647.
22. Dunbrack Jr RL, Karplus M. Backbone dependent rotamer library for proteins: application to side-chain prediction. J Mol Biol 1993;230:543–571.
23. Dunbrack Jr RL, Karplus M. Conformational analysis of the backbone dependent rotamer preferences of protein side-chains. Nature Struct Biol 1994;1:334–340.
24. Stites WE, Pranata J. Empirical evaluation of the influence of side chains on the conformational entropy of the polypeptide backbone. Proteins 1995;22:132–140.
25. Hill TL. Statistical mechanics. New York: McGraw Hill, 1956.
26. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem 1968;23:283–437.
27. Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The protein data bank: a computer based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
28. Hobohm U, Sander C. Enlarged representative set of protein structures. Protein Sci 1994;3:522–524.
29. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DJ, Wolff J, Genest M, Hagler AT. Structure and energetics of ligand binding to proteins: E. coli dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins 1988;4:31–47.
30. Matthews BW, Nicholson H, Becktel WJ. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. Proc Natl Acad Sci USA 1987;84:6663–6667.
31. Ganter C, Plückthun A. Glycine to alanine substitutions in helices of glyceraldehyde-3-phosphate dehydrogenase: effects on stability. Biochemistry 1990;29:9395–9402.
32. Green SM, Meeker AK, Shortle D. Contributions of the polar, uncharged amino acids to the stability of staphylococcal nuclease: evidence for mutational effects on the free energy of the denatured state. Biochemistry 1992;31:5717–5728.
33. Hecht MH, Sturtevant JM, Sauer RT. Stabilization of λ repressor against thermal denaturation by site-directed Gly → Ala changes in α-helix. Proteins 1986;1:43–46.
34. Imanaka T, Shibazaki M, Takagi M. A new way of enhancing the thermostability of proteases. Nature (London) 1986;324:695–697.
35. Margarit I, Campagnoli S, Frigerio F, Grandi G, De FV, Fontana A. Cumulative stabilizing effects of glycine to alanine substitutions in Bacillus subtilis neutral protease. Protein Eng 1992;5:543–550.
36. Shortle D, Stites WE, Meeker AK. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. Biochemistry 1990;29:8033–8041.
37. Takagi M, Imanaka T. Addition of a methyl group changes both catalytic velocity and thermostability of the neutral protease from Bacillus stearothermophilus. FEBS Lett 1989;254:43–46.
38. Ramakrishnan C, Srinivasan N. Glycyl residues in proteins and peptides: an analysis. Curr Sci 1990;59:851–861.
39. Karplus PA. Hydrophobicity regained. Protein Sci 1997;6:1302–1307.
40. Fauchère J, Pliska V. Hydrophobic parameters π of amino acid side-chains from the partitioning of N-acetyl-amino-acid amides. Eur J Med Chem 1983;18:369–375.
41. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
42. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol 1987;196:641–656.
43. Ponnuswamy PK, Prabhakaran M, Manavalan P. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. Biochim Biophys Acta 1980;623:301–316.
44. Wolfenden R, Andersson L, Cullis PM, Southgate CCB. Affinities of amino acid side-chains for solvent water. Biochemistry 1981;20: 849–855.