Protein Engineering vol.7 no.2 pp.149–155, 1994

Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales

Michael J.E.Sternberg[1,3] and James S.Chickos[2,4]

[1]Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, PO Box 123, 44 Lincoln's Inn Fields, London WC2A 3PX and [2]Department of Chemistry, Whiteknights, PO Box 224, Reading RG6 2AD, UK

[3]To whom correspondence should be addressed

[4]On sabbatical leave from the Department of Chemistry, University of Missouri-St Louis, St Louis, MO 63121, USA

The loss of conformational entropy of protein side-chains is a major effect in the energetics of folding. The simplest approach is to enumerate the number of freely rotatable bonds. Recently, two scales of side-chain conformational entropy have been proposed based on the definition of entropy as the Boltzmann sampling over all accessible states ($S = -R\Sigma p_i \ln p_i$, where p_i is the probability of being in a rotameric state). In one scale, derived only for aliphatic and aromatic side-chains, the values of p_i were obtained from Monte Carlo simulations. In the other scale, the observed frequencies of different rotameric states in a database of protein crystal structures yielded an estimate for p_i. Here an empirical estimation of the fusion entropy of the side-chains is used to derive a third scale. The fusion entropy is obtained as a sum of empirically derived contributions from component hydrocarbon and functional groups. There is a good agreement between the fusion scale and the other two scales. This suggests that the magnitude of conformational entropy is being correctly established.

Key words: free energy/fusion entropy/hydrophobicity/mutagenesis/protein stability

Introduction

The reduction in the number of accessible main-chain and side-chain conformations when a protein folds into a compact globule yields an unfavourable entropic effect (Karplus and Kushick, 1981; Brady and Karplus, 1985; Dill, 1990; Pickett and Sternberg, 1993). This reduction in conformational entropy counters the hydrophobic effect (Kauzmann, 1959) favouring the folded state and in part explains the marginal stability of most globular proteins (Privalov and Gill, 1988). The conformational entropy of side-chains was considered by Leach *et al.* (1966) to explain the α-helix forming tendencies of Ile and Leu. Finkelstein and Ptitsyn (1977) included the entropic effect of different side-chains to evaluate the tendencies of residues to form different secondary structures. Shakhnovich and Finkelstein (1989) suggested that protein folding will be an 'all-or-none' transition, as protein expansion must achieve a certain threshold to permit the entropic benefit of freeing the rotations of buried side-chains. There are numerous scales of residue hydrophobicity (e.g. Fauchère and Pliska, 1983; Eisenberg and McLachlan, 1986; Sharp *et al.*, 1991a,b), with 38 different scales being reported and compared by Cornette *et al.* (1987). In contrast, attempts to quantify conformational entropy changes were based either on the number of rotatable bonds (Némethy *et al.*, 1966; Finkelstein and Janin, 1989; Novotny *et al.*, 1989) or require time-consuming calculations (e.g. Karplus and Kushick, 1981; Di Nola *et al.*, 1984; Brady and Karplus, 1985; Meirovitch *et al.*, 1992).

Recently, Creamer and Rose (1992) and Pickett and Sternberg (1993) introduced scales for the conformational entropy change of side-chains (ΔS_c) during protein folding. Both scales are based on the definition of entropy S as the Boltzmann sampling of states:

$$S = -R \sum_i p_i \ln p_i$$

where p_i is the probability of the side-chain being in state i. In the scale of Pickett and Sternberg (1993), a buried side-chain in the folded protein was taken as occupying one conformational state. An estimate of p_i for the accessible states for the side-chain in the unfolded protein was obtained from the observed distribution of side-chain rotamers in 50 non-homologous protein crystal structures. This approach directly yielded ΔS_c for 13 side-chains. Special considerations were given to rotations about bonds that yielded an identical arrangement of atoms (e.g. a χ^2 rotation of 180° for Phe), to restriction of free rotation of terminal amide and carboxyl groups and to the restriction of the rotation of hydroxyl groups if they form hydrogen bonds when buried. (Hydrogen bond formation could be evaluated from the coordinates of the protein or more generally be based on the extent of burial of the side-chain.) The resultant scale gave changes in free energy per residue due to restriction of side-chain conformational entropy that ranged from 0 to +2.1 kcal/mol at 300 K. Creamer and Rose (1992) estimated conformational entropy from a Monte Carlo simulation to evaluate p_i for the rotamer library. Values for Ala, Val, Ile, Leu, Met, Phe, Tyr and Trp were obtained in the extended and the α-helical conformations. Their values for these side-chains in the extended conformation agreed with the scale of Pickett and Sternberg (1993). Here we show that an alternative empirical approach based on the entropy of fusion of small organic compounds (Chickos *et al.*, 1990, 1991) yields comparable values for the conformational entropy of protein side-chains.

Materials and methods

In the approach of Chickos *et al.* (1990, 1991), the entropy of fusion of a molecule is estimated from a sum of the contributions of component groups. These contributions depend on the structural environment of the group (see below) and were derived empirically from experimental fusion entropies of 649 organic compounds. This group additivity approach for the estimation of fusion entropies has been developed to evaluate the total phase change entropy occurring from 0 K to the melting point. Since only the sum of the major discontinuities in the heat capacity curve is evaluated by this approach, motions that are characterized by small activation energies and are likely to be active in the solid (and in the buried portions of the protein) are not evaluated. For

© Oxford University Press

Table I. Groups, group values and coefficients used in estimating fusion entropies of side-chains

Functional group	Group value (G_i)	Group coefficients (C_j)
Acyclic hydrocarbon groups		
Primary sp^3 carbon atom	4.38 (A1)	1 (A2)
Secondary sp^3 carbon atom	2.25 (B1)	1 (B2)
Tertiary sp^3 carbon atom	−3.87 (V1)	0.69 (V2)
Aromatic groups		
Tertiary sp^2 carbon	1.54 (D1)	1 (D2)
Quaternary sp^2 carbon adjacent to an sp^3 carbon	−2.47 (E1)	1 (E2)
Peripheral quaternary sp^2 carbon adjacent to an sp^2 carbon	−1.02 (F1)	1 (F2)
Ring equation for n atoms in cyclic group	$\Delta S = 8.41$ (W1) $+ 1.025[n - 3]$ (H1)	
Cyclic groups		
Cyclic tertiary sp^2 carbon	−1.04 (I1)	0.62 (I2)
Cyclic quaternary sp^2 carbon	−2.8 (J1)	0.86 (J2)

Functional group	Group value (G_K)		Group coefficients (C_K)						
			C_1	C_2		C_3[a]		C_4	
Type I									
Carboxylic acid	3.56	(K1)	1	1.83	(K2)	1.88	(K3)	1.72	(K4)
Alcohols	0.27	(L1)	1	12.6	(L2)	18.9	(L3)	26.4	(L4)
Phenol	3.96	(M1)	1	1.0	(M2)	1.0	(M3)	1.0	(M4)
Primary amide	6.26	(N1)	1	1.0	(N2)	NA	(N3)	NA	(N4)
Primary aliphatic amine	3.88	(O1)	1	1.82	(O2)	NA	(O3)	NA	(O4)
Thiol	4.3	(P1)	1	1.0	(P2)	NA	(P3)	NA	(P4)
Acyclic type II									
Secondary amides	−0.1	(Q1)	1	1.0	(Q2)	NA	(Q3)	NA	(Q4)
Secondary amines	−0.52	(R1)	1	1.0	(R2)	NA	(R3)	NA	(R4)
Sulphide	1.72	(S1)	1	NA	(S2)	0.36	(S3)	NA	(S4)
Cyclic type II									
Cyclic secondary amine	0.44	(T1)	1	NA	(T2)	NA	(T3)	NA	(T4)
Cyclic sp^2 nitrogen	0.4	(U1)	1	1.0	(U2)	NA	(U3)	NA	(U4)

The terms primary, secondary, tertiary and quaternary are defined by the number of hydrogen atoms attached to the carbon, 3, 2, 1 and 0, respectively. Data abstracted from Chickos *et al.* (1991). The terms in brackets (e.g. A1) are a device to explain the calculations reported in Table II. NA indicates that a value is not available.

[a]A value of 1 was used when the value of $C3$ is unavailable but required.

most compounds the total phase change entropy is equivalent to the fusion entropy.

At the melting temperature, the two phases are in equilibrium so $\Delta H_{fus} = T\Delta S_{fus}$. The enthalpy term is dominated by the loss of favourable packing that occurred in the solid state. The entropic effects are mainly due to the freeing up of both rigid-body motion and internal rotations on melting. Thus the entropy of fusion (ΔS_{fus}) provides an estimate of the entropy change due to these motions. The rigid-body term includes components due to translational and rotational entropy, both of which increase logarithmically with the molecular mass and moments of inertia (Benson, 1976). Thus, for a series of organic molecules of similar mass, the differences in ΔS_{fus} provide an estimation of entropic effects other than those due to rigid-body motion.

The organic solid state will be taken as a model for the interior of the folded protein in which atomic motion of side-chains is constrained to within a single rotameric conformation. Indeed, measurements of packing density with protein crystal structure are comparable to that of small organic molecules (Richards, 1974). Similarly, the non-rigid-body motions in the liquid state will model the unfolded protein in which side-chains are sampling allowed rotamers. In this model the rotation of hydrogen atoms must be considered. It is assumed that in the folded protein as well as in the denatured molecule there is free rotation of methyl groups and of other hydrogen atoms, apart from the restriction of hydroxyl groups that form hydrogen bonds when buried. The estimate of fusion entropy does not include contributions from effects such as the freeing up of methyl rotations that are characterized by small activation energies. Thus the relative values of the entropies of fusion of model compounds representing side-chains could provide a measure of the conformational entropy effects during protein folding. A similar model for the entropy of fusion of linear hydrocarbons has been used by Nicholls *et al.* (1991) to estimate the conformational entropy per rotational bond of the aliphatic side-chains.

The application of the group additivity scheme to evaluate fusion entropies of peptides depends on the structure and complexity of the amino acid residue. Equation (1) can be used for amino acids containing acyclic and aromatic components. Those amino acid residues containing non-aromatic cyclic components can be evaluated using equation (2).

Acyclic and aromatic molecules

$$\Delta S_{fus} = \sum_i n_i G_j = \sum_j n_j C_j G_j + \sum_k n_k C_k G_k \quad (1)$$

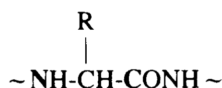
Cyclic molecules

$$\Delta S_{fus} = [8.41 + 1.025(n-3)] + \sum_i n_i G_i + \sum_j n_j C_j G_j + \sum_k n_k C_K G_k \quad (2)$$

$$\text{where } k = \sum_k n_k \quad (3)$$

The quantity n in these equations refers to the number of identical groups and in the case of cyclic molecules to the number of methylene groups necessary to simulate the size of the ring. The subscripts i, j and k, respectively, identify the hydrocarbon components, the carbons bearing the functional group and the functional groups, where a functional group is any non-hydrocarbon component. The terms C and G refer to group coefficients and group values, respectively. These terms are provided in Table I, which contains a summary of the groups used in estimating the amino acid residues in proteins (see also Figure 1). Group values for functional groups in Pro and Arg are not presently available and these side-chains are not considered here. In addition, the estimation of conformational entropy for Gly would include the different contribution due to main-chain effects and, although the value is reported, it is not considered in detail below.

In modelling the fusion entropies of multifunctional compounds, it was found empirically that the group values used in evaluating the contribution of some functional groups were not strictly additive, but often depended on the total number of functional groups in the molecule under consideration. This fact is reflected by the magnitudes of some group coefficients, C_j and C_K in Table I. Ideally the indices K determined by equation (3) and used in the symbols C_K should be determined based on the total number of functional groups in the molecule. Some of these C_K values change markedly as the number of groups increases. We have found, from work on polymers (work in progress), that C_3 and C_4 values are suitable for estimating fusion entropies of macromolecules and thus the C_4 values represent suitable maximum values. The calculation for ring compounds is shown empirically to apply to the modelling of fusion of low molecular weight organic compounds (typically $n < 8$). This calculation is not intended to apply to modelling the effect of disulphide linkages in protein chains. In modelling the entropy change of each side-chain (R) in going from a conformationally mobile to a rigid state, the total number of functional groups in each peptide has some effect on the magnitude of the entropy change estimated. Group coefficients for some functional groups in highly substituted amino acids are not presently available. We have chosen (see Table II) to calculate the contribution of each amino acid residue to the protein backbone as the contribution of a disubstituted secondary amide linkage (Q1 × Q2) and a trisubstituted methine carbon (V1 × V2), as illustrated in bold face for an R side-chain:



Each peptide in the protein contributes one amide linkage to the protein. Group values and group coefficients for side-chains were chosen on the basis of structure and substitution, respectively. Entropy changes for pure hydrocarbon side-chains (Gly, Ala, Val, Ile, Leu, Phe) are estimated as disubstituted molecules and other side-chains as trisubstituted. The value of C_K would be

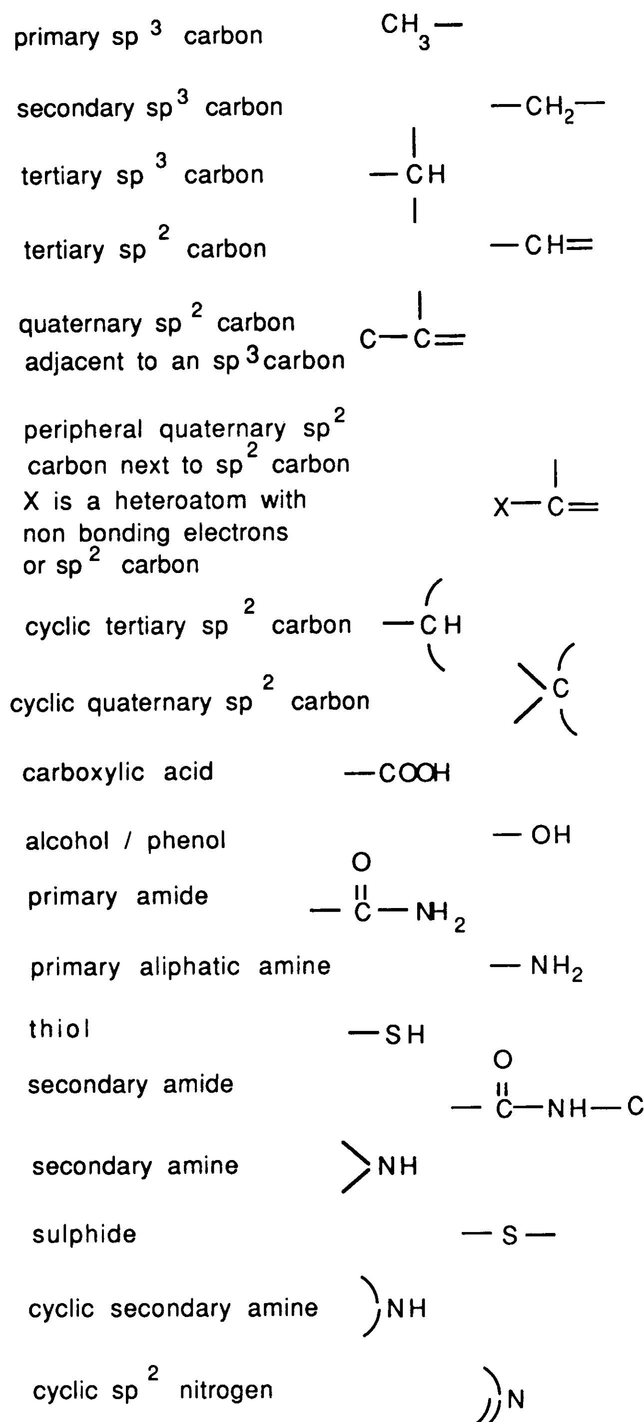


Fig. 1. Chemical structure of groups used to estimate fusion entropy (see Table 1). Curved bonds represent bonding contributions to a ring.

chosen as appropriate for each C_2 and C_3, respectively. Values of C_K for some functional groups in environments containing multiple substitutions are not available (e.g. N3, P3, Q3, T3, U3). Group coefficients for these cases were chosen selecting the most highly substituted group coefficient available. In most cases the group coefficients were not very sensitive to successive substitutions (functional groups Q, P, N, U; Table I). Only in one case (functional group T) and in two peptides (His, Trp) is the sensitivity of the functional group to successive substitutions unknown.

The groups contributing to each amino acid residue are shown

Table II. Calculation of side-chain conformational entropy

Peptide unit related to	Calculation of ΔS of peptide[a]	ΔS of peptide (e.u.)[b]	$T\Delta S_{c\text{-fus}}$ (kcal/mol)[c]	$T\Delta S_{c\text{-rot}}$ (kcal/mol)[d]	$T\Delta S_{c\text{-sim}}$ (kcal/mol)[e]	Nbond[f]
Ala	(Q1*Q2+V1*V2+A1)	1.61	0.00	0.00	0.0	0
Asn	(Q1*Q2+V1*V2+B1)+N1*N3	5.74	−1.24	−1.57	–	2
Asp	(Q1*Q2+V1*V2+B1)+K1*K3	6.17	−1.37	−1.25	–	2
Cys	(Q1*Q2+V1*V2+B1)+P1*P3	3.78	−0.65	−0.55	–	1
Gln	(Q1*Q2+V1*V2+B1)+B1+N1*N3	7.99	−1.91	−2.11	–	3
Glu	(Q1*Q2+V1*V2+B1)+B1+K1*K3	8.42	−2.04	−1.81	–	3
Gly	(Q1*Q2+B1)	2.15	(−0.16)	(0.0)	–	0
His	(Q1*Q2+V1*V2+B1)+W1+2*H1 +2*I1*I2+J1*J2+T1*T3+U1*U3	7.08	−1.64	−0.96	–	2
Ile	(Q1*Q2+V1*V2+B1)+2*A1+V1	4.37	−0.83	−0.89	−0.74	2
Leu	(Q1*Q2+V1*V2+B1)+2*A1+V1	4.37	−0.83	−0.78	−0.54	2
Lys	(Q1*Q2+V1*V2+B1)+3*B1+O1*O3	13.29	−3.50	−1.94	–	4
Met	(Q1*Q2+V2*V1+B1)+A1+B1+S1*S3	6.73	−1.54	−1.61	−1.54	3
Phe	(Q1*Q2+V1*V2+B1)+5*D1+E1	4.71	−0.93	−0.58	−1.27	2
Ser	(Q1*Q2+V1*V2+B1)+L1*L3	4.58	−0.89	−1.71	–	1
Thr	(Q1*Q2+V1*V2+A1)+V1*V2+L1*L3	4.04	−0.73	−1.63	–	1
Trp	(Q1*Q2+V1*V2+B1)+4*D1+W1 +2*H1+I1*I2+2*J1+J1*J2+T1*T3	7.89	−1.88	−0.97	−1.22	2
Tyr	(Q1*Q2+V1*V2+B1)+4*D1+E1+F1 +M1*M3	6.11	−1.35	−0.98	−1.73	2
Val	(Q1*Q2+V1*V2+V1+A1)+A1+V1	2.12	−0.15	−0.51	−0.44	1

[a]Terms used in Table I for calculation. Coefficients with a value of 1.00 are not shown.
[b]Calculated entropy of fusion in entropy units (e.u.) in units of cal/mol K.
[c]$T\Delta S_{c\text{-fus}}$ calculated from column 2 relative to Ala at T = 300 K (i.e. from fusion data).
[d]$T\Delta S_{c\text{-rot}}$ from Pickett and Sternberg (1993) based on the rotamer library.
[e]$T\Delta S_{c\text{-sim}}$ from the Monte Carlo simulations of Creamer and Rose (1992).
[f]Nbond is number of side-chain rotatable bonds leading to altered positions for heavy atoms, from Pickett and Sternberg (1993).

Table III. Comparisons of entropy scales

Scales	Residues	Regression equation	Correlation	Paired differences (mean ± SD)
fus/rot	not GPR	$T\Delta S_{c\text{-fus}} = 0.95\ T\Delta S_{c\text{-rot}} - 0.15$	0.69 (5%)	−0.096 ± 0.596
fus/rot	not GPRK	$T\Delta S_{c\text{-fus}} = 0.71\ T\Delta S_{c\text{-rot}} - 0.34$	0.68 (5%)	−0.004 ± 0.476
fus + OH/rot	not GPR	$T\Delta S_{c\text{-fus}} = 1.13\ T\Delta S_{c\text{-rot}} - 0.10$	0.81 (1%)	−0.246 ± 0.493
fus/sim	AVILMFYW	$T\Delta S_{c\text{-fus}} = 0.91\ T\Delta S_{c\text{-sim}} - 0.09$	0.84 (1%)	−0.004 ± 0.353
rot/sim	AVILMFYW	$T\Delta S_{c\text{-rot}} = 0.60\ T\Delta S_{c\text{-sim}} - 0.23$	0.77 (5%)	−0.145 ± 0.382
fus/rot	AVILMFYW	$T\Delta S_{c\text{-fus}} = 1.13\ T\Delta S_{c\text{-rot}} - 0.05$	0.81 (5%)	−0.148 ± 0.388
fus/Nbond	not GPR	$T\Delta S_{c\text{-fus}} = 0.76\ Nbond + 0.21$	0.89 (1%)	
rot/Nbond	not GPR	$T\Delta S_{c\text{-rot}} = 0.41\ Nbond - 0.36$	0.67 (5%)	
fus/Nbond	AVILMFYW	$T\Delta S_{c\text{-fus}} = 0.59\ Nbond + 0.10$	0.81 (5%)	
rot/Nbond	AVILMFYW	$T\Delta S_{c\text{-rot}} = 0.49\ Nbond + 0.06$	0.93 (1%)	
sim/Nbond	AVILMFYW	$T\Delta S_{c\text{-sim}} = 0.54\ Nbond + 0.01$	0.80 (5%)	

Fus, rot and sim denote the fusion, rotamer library (Pickett and Sternberg, 1993) and simulation (Creamer and Rose, 1992) scales. fus + OH denotes the addition of hydroxyl rotation for Ser, Thr and Tyr. Nbond is the number of rotatable bonds from Table I. The linear least squares regression is given, together with the Pearson correlation coefficient (r) with the significance level of a non-zero slope. The mean ± standard deviation of the paired differences of the two scales are given. All values are in kcal/mol.

in the second column of Table II. The results of the calculation are given in the third column of the table. Since we are interested in the loss of side-chain conformational entropy, Table II, column 4 gives $T\Delta S_c$ changes for each amino acid side-chain relative to Ala.

Results

Table II contrasts the values for $T\Delta S_c$ (T = 300 K) obtained from the fusion data (denoted $\Delta S_{c\text{-fus}}$) with the values from Pickett and Sternberg (1993), derived from the rotamer library ($\Delta S_{c\text{-rot}}$), and from Creamer and Rose (1993), based on Monte Carlo simulations ($\Delta S_{c\text{-sim}}$). Two approaches are used to quantify the agreement between the scales. First, the linear least-squares regression with the Pearson correlation coefficient (Altman, 1991) are reported. This quantifies the extent to which two scales are associated, but does not quantify the agreement. Thus a second measure is introduced that considers the paired differences between the two scales and reports the mean ± standard deviation. The comparisons between the scales are summarized in Table III.

Figure 2A shows that the fusion and rotamer-library scales are in broad agreement. At T = 300 K the two scales are related by the equation:

$$T\Delta S_{c\text{-fus}} = 0.95\ T\Delta S_{c\text{-rot}} - 0.15$$

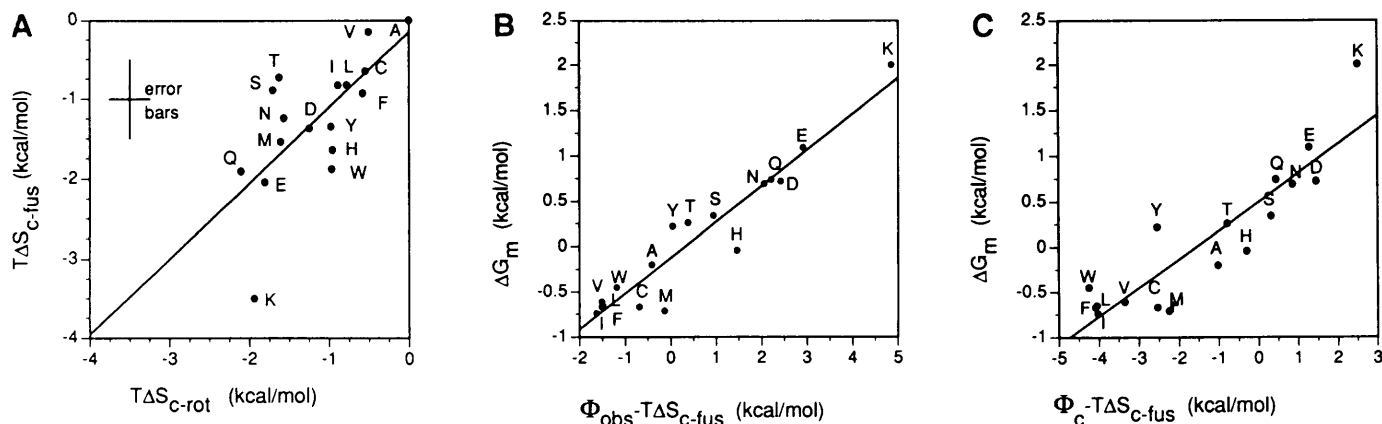


Fig. 2. (**A**) Values for side-chains of $T\Delta S_{c\text{-fus}}$ (i.e. fusion values) plotted against $T\Delta S_{c\text{-rot}}$ (i.e. rotamer-library values) at T = 300 K. (**B**) ΔG_m for residue burial from Miller *et al.* (1987) plotted against $\Phi_{obs} - T\Delta S_{c\text{-fus}}$ [hydrophobicity scale of Fauchère and Pliska (1983), minus fusion entropy times at 300 K]. The equation of the line is $\Delta G_m = 0.40(\Phi_{obs} - T\Delta S_{c\text{-fus}}) - 0.13$ with a correlation of 0.95. (**C**) ΔG_m plotted against $\Phi_c - T\Delta S_{c\text{-fus}}$ [corrected hydrophobicity scale of Sharp *et al.* (1991a,b) minus fusion entropy times at 300 K]. The equation of the line is $\Delta G_m = 0.32(\Phi_c - T\Delta S_{c\text{-fus}}) + 0.493$ with a correlation of 0.90.

The correlation coefficient for the regression is $r = 0.69$, which is significant at the 1% level. The mean ± standard deviation of the paired differences is −0.096 ± 0.696 kcal/mol. Thus, on average, the scales agree closely, but there is some variation in the agreement for the individual observations.

The largest difference in the two scales is for Lys ($T\Delta S_{c\text{-fus}}$ = −3.50 kcal/mol and $T\Delta S_{c\text{-rot}}$ = −1.94 kcal/mol). Without this side-chain, the mean ± standard deviation of the differences is −0.004 ± 0.476, i.e. the standard deviation is reduced markedly. The correlation coefficient is almost unaltered (r = 0.68). The fusion approach may have overestimated the number of allowed side-chain rotamers for Lys, by not considering steric restrictions due to main-chain/side-chain interactions. Alternatively, the rotamer-library approach may have inaccurately estimated p_i for the Lys rotamer library, possibly due to experimental errors in crystal structures. At present one cannot distinguish which value for Lys is more accurate.

There are also differences in the values of $T\Delta S_c$ for Ser and Thr. Pickett and Sternberg (1993) introduced an additional term of $-R\ln 6$ (= −1.07 kcal/mol at 300 K) to the values of ΔS_c of these side-chains to estimate the effects of the restriction of hydroxyl rotation due to hydrogen bonding. It is possible that the change in entropy on melting did not fully include the effect of freeing up hydroxyl rotation, as in both the liquid and the solid states these hydroxyls form hydrogen bonds. Then, an additional term of $-R\ln 6$ should be introduced to the values of $T\Delta S_{c\text{-fus}}$ for Ser and Thr, leading to a better agreement of the scales. However, an addition term of $-R\ln 2$ (= 0.41 kcal/mol) was also added to include the hydroxyl rotation of Tyr, and if this term were added to $T\Delta S_{c\text{-fus}}$ then the agreement for Tyr would be poorer. With these hydroxyl corrections, the correlation between the scales improves (r = 0.81), although the agreement becomes poorer, with the mean ± standard deviation of the differences being −0.25 ± 0.49 kcal/mol.

Other outliers in the correlation are for His and Trp. Both peptides contain unknown sensitivity to successive substitutions in the group additivity scheme, so the calculated fusion entropies may be less accurate than for other side-chains. Further analysis of organic compounds with the relevant functional groups would help in the evaluation of more accurate estimates of the fusion entropies for these two side-chains.

Creamer and Rose (1992) estimated ΔS_c for Ala, Val, Ile, Leu, Met, Phe, Tyr and Trp from a Monte Carlo simulation of the rotamer population (see Table II). For these side-chains the correlation (r) between $\Delta S_{c\text{-fus}}$ and $\Delta S_{c\text{-sim}}$ is 0.84, with the mean ± standard deviation of the paired differences being 0.00 ± 0.35 kcal/mol. For these side-chains the corresponding values for $\Delta S_{c\text{-fus}}$ and $\Delta S_{c\text{-rot}}$ are r = 0.81 and the mean ± standard deviation of paired differences is −0.15 ± 0.39 kcal/mol. The slightly closer agreement of the fusion values with the Creamer and Rose (1992) scale results mainly from the value for Trp. For these side-chains the correlation coefficient between $\Delta S_{c\text{-fus}}$ and $\Delta S_{c\text{-rot}}$ is 0.81.

Table III also compares these three scales with the simpler approach based on the number of side-chain bonds about which rotation leads to different heavy atom positions (see Pickett and Sternberg, 1993, Table 5). For the 17 residues in Table II, the correlation coefficients of the number of rotatable bonds with fusion and rotamer scales are 0.89 and 0.67. The lower correlation for the rotamer scale results in part from the treatment of the additional terms for free rotations of hydroxyl, amide and carboxyl groups. For the eight residues considered by Creamer and Rose (1993), the correlation coefficients of the number of rotatable bonds with fusion, rotamer and simulation scales are 0.81, 0.93 and 0.80. Thus, although all three scales of entropy are related to the number of rotatable bonds, they also include other terms that should provide a more accurate description of the effect.

Pickett and Sternberg (1993) suggested that one should include terms for both hydrophobicity and conformational entropy to explain the tendency of residue types to be buried. Two scales for the hydrophobicity were considered—the experimental values (Φ_{obs}) of Fauchère and Pliska (1983) and these values corrected (Φ_c) for the different volumes of solute and solvent (Sharp *et al.*, 1991a,b). The observed frequencies of residue burial were quantified by the free energy transfer scale (ΔG_m) of Miller *et al.* (1987), derived from analysis of protein crystal structures. Figure 2B and C shows the agreement between ΔG_m and two hydrophobicity scales modified by $T\Delta S_{c\text{-fus}}$ for all residues except Arg, Gly and Pro. [Note that Pickett and Sternberg (1993), excluded Lys and Pro, following Miller *et al.* (1987).]

Formally, the correlation coefficients of ΔG_m with Φ_{obs} alone, with $\Phi_{obs} - T\Delta S_{c\text{-rot}}$ and with $\Phi_c - T\Delta S_{c\text{-fus}}$ are 0.878, 0.908 and 0.954. The significance of the improvements in correlation coefficients can be estimated by the approach suggested by Neill and Dunn (1975). $\Phi_{obs} - T\Delta S_{c\text{-fus}}$ is significantly more highly

correlated with ΔG_m than Φ_{obs} (significance level $P = 0.025$). $\Phi_{obs} - T\Delta S_{c\text{-rot}}$ is not significantly more highly correlated with ΔG_m than Φ_{obs} ($P = 0.32$). $\Phi_{obs} - T\Delta S_{c\text{-fus}}$ is not significantly more highly correlated with ΔG_m than $\Phi_{obs} - T\Delta S_{c\text{-rot}}$, but it is almost so ($P = 0.08$). The correlation coefficients of ΔG_m with Φ_c alone, with $\Phi_c - T\Delta S_{c\text{-rot}}$ and with $\Phi_c - T\Delta S_{c\text{-fus}}$ are 0.737, 0.811 and 0.900. $\Phi_c - T\Delta S_{c\text{-fus}}$ is significantly more highly correlated with ΔG_m than Φ_c ($P = 0.00006$). $\Phi_c - T\Delta S_{c\text{-rot}}$ is also significantly more highly correlated with ΔG_m than Φ_c ($P = 0.038$). $\Phi_c - T\Delta S_{c\text{-fus}}$ is significantly more highly correlated with ΔG_m than $\Phi_c - T\Delta S_{c\text{-rot}}$ ($P = 0.002$).

In interpreting these results, one must consider that the ΔG_m values are only a guide to the tendency of a residue to be buried in a protein, for example, relying on a cut-off for the definition of burial. The general conclusion is that including a scale of side-chain entropy with hydrophobicity tends to lead to a better model for the tendency of residues to be buried than hydrophobicity alone. There is also some suggestion that the fusion approach leads to a better correlation, but this mainly results from the difference in the value for Lys in the two scales.

Discussion and conclusion

In any application of the entropy scale, it is useful to have an estimate of the likely errors. The scale based on the rotamer library provides a direct evaluation from protein structure for 13 side-chains. Sampling errors were shown to introduce at most an error of ±5% (Pickett and Sternberg, 1993). In addition, heuristic terms were introduced to model free rotation in seven side-chains. The largest of these terms was $-R\ln 6$ for Asn, Gln, Ser and Thr. The actual value could be between $-R\ln 4$ and $-R\ln 8$, leading to an error of up to ±0.25 kcal/mol. This error value is shown in Figure 2A. The fusion approach is based on experimental values of entropy changes. The mean deviation between predicted and experimental values of ΔS_{fus} was ±2 cal/mol K (= 0.6 kcal/mol at 300 K) for multisubstituted compounds in the database. This provides a lower limit on the probable errors in the values of $\Delta S_{c\text{-fus}}$. However, in the fusion scale only values relative to Ala rather than absolute magnitudes are required, and so we suggest an error value of ±0.5 kcal/mol for $\Delta S_{c\text{-fus}}$, which is shown in Figure 2A.

To evaluate the correspondence of the two scales of conformational entropy, it is instructive to acknowledge the discrepancies and controversies over the correct values for residue hydrophobicity. Cornette *et al.* (1987) reports the correlations between 45 measures of hydrophobicity derived from 38 scales. For example, two widely used measures are the scales of Hopp and Woods (1981) and of Kyte and Doolittle (1982), and although the range of values is similar, the correlation between them is 0.70, compared to the correlation of 0.69 between the two entropy scales. More importantly, it has recently been suggested (Sharp *et al.*, 1991a,b) that the results of experimental transfer experiments should be corrected for the sizes of the solvent and solute, leading to nearly a doubling of the hydrophobic effect. In this paper the correlation between the empirical ΔG_m values of Miller *et al.* (1987) and the calculated combined effects of hydrophobicity and conformational entropy was explored. Although correlations of at least 0.9 can be obtained, the slopes of the graphs in Figure 2B and C are 0.40 and 0.32, which are far removed from the 1.0 required for agreement. In contrast, the agreement between the two entropy scales in Figure 2A has a slope of 0.95.

At present, the agreement of these two scales derived independently and from quite different considerations provides confidence that the values are roughly correct. This strongly suggests that the change in conformational entropy of buried side-chains during protein folding results from the reduction in the rotameric conformational space. Recently, Privalov and Makhatadze (1993) evaluated the entropy of residue hydration and using experimental values for the entropy of protein folding estimated the change in side-chain conformational entropy during protein folding. Their values are about five times higher for individual proteins than the values estimated by Pickett and Sternberg (1993), based on freezing buried side-chains. Indeed, Privalov and Makhatadze (1993) suggest that the conformational entropic effect for side-chains is between the values for fusion and sublimation, whilst in this paper a scale is based solely on fusion as the model. As information correlating residues and protein stability accumulates, particularly via site-directed mutations (e.g. Matsumura *et al.*, 1988; Sandberg and Terwilliger, 1991; Serrano *et al.*, 1992), it may be possible to identify the best scale of entropy to use. More generally, entropic effects due to restriction of rotation occur in many biological processes, not just protein folding. This approach, based on fusion data, could be extended to other modelling areas, such as investigating the binding of ligands, including drugs, to receptors.

Acknowledgements

We thank Dr Stephen Pickett (Rhône-Poulenc Rorer Ltd, Dagenham, UK) for helpful discussions and Mr D.G.Altman (ICRF) for statistical advice. J.S.C. thanks the University of Missouri-St Louis for a Research Fellowship, Professors Monty Frey and Robin Walsh and Reading University for their hospitality and the USEPA, Office of Exploratory Research (Grant R81-9067-010) for partial financial support.

References

Altman,D.G. (1991) *Practical Statistics for Medical Research.* Chapman and Hall, London, pp. 299–316.

Benson,S.W. (1976) *Thermochemical Kinetics—Methods for the Estimation of Thermochemical Data and Rate Parameters*, 2nd edn. John Wiley, New York, Chapter 2.

Brady,J. and Karplus,M. (1985) *J. Am. Chem. Soc.*, **107**, 6103–6105.

Chickos,J.S., Braton,C.M., Hesse,D.G. and Liebman,J.F. (1990) *J. Org. Chem.*, **55**, 3833–3840.

Chickos,J.S., Hesse,D.G. and Liebman,J.F. (1991) *J. Org. Chem.*, **56**, 927–938.

Cornette,J.L., Cease,K.B., Margalit,H., Spouge,J.L., Berzofsky,J.A. and DeLisi,C. (1987) *J. Mol. Biol.*, **195**, 659–685.

Creamer,T.P. and Rose,G.D. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5937–5941.

Di Nola,A., Berendsen,H.J.C. and Edholm,O. (1984) *Macromolecules*, **17**, 2044–2050.

Dill,K.A. (1990) *Biochemistry*, **29**, 7133–7155.

Eisenberg,D. and McLachlan,A.D. (1986) *Nature*, **319**, 199–203.

Fauchère,J.-L. and Pliska,V. (1983) *Eur. J. Med. Chem.–Chim. Ther.*, **18**, 369–375.

Finkelstein,A.V. and Janin,J. (1989) *Protein Engng*, **3**, 1–3.

Finkelstein,A.V. and Ptitsyn,O.B. (1977) *Biopolymers*, **16**, 469–495.

Hopp,T.P. and Woods,K.R. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 3824–3828.

Karplus,M. and Kushick,J.N. (1981) *Macromolecules*, **14**, 325–332.

Kauzmann,W. (1959) *Adv. Protein Chem.*, **14**, 1–63.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.

Leach,S.J., Némethy,G. and Scheaga,H.A. (1966) *Biopolymers*, **4**, 369–407.

Matsumura,M., Becktel,W.J. and Matthews,B.W. (1988) *Nature*, **334**, 406–410.

Meirovitch,H., Kitson,D.H. and Hagler,A.T. (1992) *J. Am. Chem. Soc.*, **114**, 5386–5399.

Miller,S., Janin,J., Lesk,A.M. and Chothia,C. (1987) *J. Mol. Biol.*, **196**, 641–656.

Neill,J.J. and Dunn,O.J. (1975) *Biometrics*, **31**, 531–543.

Némethy,G., Leach,S.J. and Scheraga,H.A. (1966) *J. Phys. Chem.*, **70**, 998–1004.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) *Proteins: Struct. Funct. Genet.*, **11**, 281–296.

Novotny,J., Bruccoleri,R.E. and Saul,F.A. (1989) *Biochemistry*, **28**, 4735–4749.

Pickett,S.D. and Sternberg,M.J.E. (1993) *J. Mol. Biol.*, **231**, 825–839.

Privalov,P.L. and Gill,S.J. (1988) *Adv. Protein Chem.*, **39**, 191–234.

Privalov,P.L. and Makhatadze,G.I. (1993) *J. Mol. Biol.*, **224**, 715–723.

Richards,F.M. (1974) *J. Mol. Biol.*, **82**, 1–14.
Sandberg,W.S. and Terwilliger,T.C. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1706–1710.
Serrano,L., Kellis,J.T., Cann,P., Matouschek,A. and Fersht,A.R. (1992) *J. Mol. Biol.*, **224**, 783–804.
Shakhnovich,E.I. and Finkelstein,A.V. (1989) *Bioploymers*, **28**, 1667–1680.
Sharp,K.A., Nicholls,A., Fine,R.F. and Honig,B. (1991a) *Science*, **252**, 106–109.
Sharp,K.A., Nicholls,A., Friedman,R. and Honig,B. (1991b) *Biochemistry*, **30**, 9686–9697.

Received on July 19, 1993; revised on September 27, 1993; accepted on October 1, 1993