PROTEINS: Structure, Function, and Genetics 40:443–450 (2000)

Side-Chain Conformational Entropy in Protein Unfolded States

Trevor P. Creamer
Center for Structural Biology, Department of Biochemistry, University of Kentucky, Lexington, Kentucky

ABSTRACT **The largest force disfavoring the folding of a protein is the loss of conformational entropy. A large contribution to this entropy loss is due to the side-chains, which are restricted, although not immobilized, in the folded protein. In order to accurately estimate the loss of side-chain conformational entropy that occurs upon folding it is necessary to have accurate estimates of the amount of entropy possessed by side-chains in the ensemble of unfolded states. A new scale of side-chain conformational entropies is presented here. This scale was derived from Monte Carlo computer simulations of small peptide models. It is demonstrated that the entropies are independent of host peptide length. This new scale has the advantage over previous scales of being more precise with low standard errors. Better estimates are obtained for long (e.g., Arg and Lys) and rare (e.g., Trp and Met) side-chains. Excellent agreement with previous side-chain entropy scales is achieved, indicating that further advancements in accuracy are likely to be small at best. Strikingly, longer side-chains are found to possess a smaller fraction of the theoretical maximum entropy available than short side-chains. This indicates that rotations about torsions after χ_2 are significantly affected by side-chain interactions with the polypeptide backbone. This finding invalidates previous assumptions about side-chain-backbone interactions. Proteins 2000;40:443–450.**

Key words: side-chain rotamer; protein folding; Monte Carlo computer simulation

INTRODUCTION

Proteins tend to have free energies of folding in the range -5 to -20 kcal mol^{-1}. These relatively small free energies are the sum of many much larger parts.[1,2] Contributions include hydrogen bonds, van der Waals interactions, electrostatics and entropy, which includes the hydrophobic effect. The entropic contributions to folding other than hydrophobicity are large and often unfavorable. The largest unfavorable entropic factor is the conformational entropy lost by the protein as it folds into its ensemble of native states. The protein backbone loses a significant amount of its conformational entropy as it becomes restricted to a narrow range of conformations.[3] Side-chains buried in the protein interior are generally thought to lose much, although certainly not all, of their conformational entropy as they become packed into the dense core.[4,5] Side-chains on the protein surface also lose significant amounts of entropy due to steric restrictions imposed by the newly crowded local environment.[5–8] Recent experimental studies indicate that the amount of entropy retained by side-chains in folded proteins is dependent upon fold and local environment.[9,10] In order to accurately estimate entropic contributions to protein stability and folding energetics it is necessary to be able to determine the conformational entropy of side-chains in unfolded states.

The entropy possessed by each side-chain in unfolded states has been estimated by a number of groups using computational methods,[7,8,11–13] from surveys of side-chain rotamers in proteins of known structure[6,14,15] and from entropies of fusion.[16] Such entropy scales have been used to rationalize the contributions of individual side-chains to α-helix formation.[7,8,12,17,18] They have also been used in studies of ligand binding.[19,20] Side-chain entropies could also be employed in the identification of residues that will be on the surface of folded proteins—such residues tend to be those that have the most entropy to lose upon folding.[6,21]

In this work, side-chain conformational entropies are estimated for all residues **Xaa** (except Ala, Gly, and Pro) using Monte Carlo computer simulations of Ace-Ala-**Xaa**-Ala-NMe tripeptides. Long simulations were run in order to generate good estimates of side-chain rotamer distributions. The rotamer distributions obtained are in good agreement with distributions derived from a dataset of protein structures. Statistically better entropy estimates were obtained in the cases of the longer (e.g., Lys and Arg) and/or rarer (e.g., Met and Trp) side-chains as compared to previous scales. It is demonstrated that assumptions that rotations about the latter torsions of long side-chains (e.g., χ_3 and χ_4 etc. of Arg and Lys) are independent of interactions with the polypeptide backbone are incorrect. The validity of tripeptides as models for the conformational properties of unfolded states was tested by simulating Leu in a series of peptides Ace-$(Ala)_n$-**Leu**-$(Ala)_n$-NMe (n = 1, 2, and 3): no variation in estimated Leu side-chain entropies was observed. The side-chain conformational entropy scale presented here will prove useful in estimating ener-

*Correspondence to: Trevor P. Creamer, Center for Structural Biology, Department of Biochemistry, University of Kentucky, 800 Rose Street, Lexington, KY 40536-0298. E-mail: trevor@euripides.gws.uky.edu

Received 21 January 2000; Accepted 10 April 2000

TABLE I. Side Chain Rotamers and Rotamer Classes Employed in This Work

Residue		Rotamers	Conformations	Number of rotamer classes[a] r_{Total}
All residues	χ_1	N-CA-CB-CG	*trans*, g^-, g^+	—
	χ_2	CA-CB-CG-CD	*trans*, g^-, g^+	
Arg	χ_3	CB-CG-CD-NE	*trans*, g^-, g^+	81
	χ_4	CG-CD-NE-CZ	*trans*, g^-, g^+	
Asn	χ_2	CA-CB-CG-OD1	*trans*, +90°, 0°, −90°	12
Asp	χ_2	CA-CB-CG-OD1	*trans*, +90°, 0°, −90°	6
Gln	χ_2	CA-CB-CG-CD	*trans*, g^-, g^+	36
	χ_3	CB-CG-CD-OE1	*trans*, +90°, 0°, −90°	
Glu	χ_2	CA-CB-CG-CD	*trans*, g^-, g^+	18
	χ_3	CB-CG-CD-OE1	*trans*, +90°, 0°, −90°	
His	χ_2	CA-CB-CG-ND1	*trans*, +90°, 0°, −90°	12
Ile	χ_2	CA-CB-CG1-CD1	*trans*, g^-, g^+	9
Leu	χ_2	CA-CB-CG-CD1	*trans*, g^-, g^+	9
	χ_2	CA-CB-CG-CD	*trans*, g^-, g^+	
Lys	χ_3	CB-CG-CD-CE	*trans*, g^-, g^+	81
	χ_4	CG-CD-CE-NZ	*trans*, g^-, g^+	
Met	χ_2	CA-CB-CG-SD	*trans*, g^-, g^+	27
	χ_3	CB-CG-SD-CE	*trans*, g^-, g^+	
Phe	χ_2	CA-CB-CG-CD1	*trans*, +90°, 0°, −90°	6
Trp	χ_2	CA-CB-CG-CD1	*trans*, +90°, 0°, −90°	12
Tyr	χ_2	CA-CB-CG-CD1	*trans*, +90°, 0°, −90°	6

[a]Total possible, including χ_1, after accounting for degenerate conformations.

getics of protein and peptide folding, in protein engineering, and in ligand binding and design.

METHODS

Entropy Calculations

Conformational entropy S can be calculated quite simply using Boltzmann's formulation:

$$S = \sum_i p_i \ln p_i \quad (1)$$

where the sum is taken over all conformational states of the system and p_i is the probability of being in state i. In the case of side-chain conformational entropy, conformational states are side-chain rotamers, and probabilities are calculated from the side-chain rotamer distributions.

Side-chain rotamers are defined in a similar manner to earlier work.[6–8,11] Rotamers employed in this work are defined in Table I for each side-chain. Note that only torsions involving four heavy (non-hydrogen) atoms are considered. Thus, Ser is considered to have just one side-chain torsion, N-CA-CB-OG, despite the hydroxyl hydrogen being included in the computer simulation. Included in Table I are the total number of possible rotamer classes for each side-chain after correction for symmetrically degenerate conformations. The IUPAC-IUB convention is used to define *trans* (±180°), *gauche*− ($g-$, −60°) and *gauche*+ ($g+$, +60°) conformations.[22]

Peptide Models

Seventeen residues are of interest in this work. Gly, Ala, and Pro are excluded: Gly because it lacks a side-chain, Ala since its side-chain consisting purely of a methyl group (non-polar hydrogens are not considered) and Pro because it is an imino acid with the side-chain covalently bonded to the backbone. Gly and Ala each possess zero side-chain conformational entropy, and the side-chain conformational entropy of Pro has yet to be determined. Each residue of interest is modeled in the guest position (**Xaa**) of the host peptide Ace-Ala-**Xaa**-Ala-NMe.

Leu was also modeled in the peptide series Ace-$(Ala)_n$-**Leu**-$(Ala)_n$-NMe (n = 1, 2, and 3) in order to determine the effect of chain length upon calculated side-chain entropies. Longer peptides were not simulated since they collapse into relatively deep local energetic minima and consequently do not adequately sample conformational space.

Monte Carlo Computer Simulations

Peptide conformational behavior was simulated using the Monte Carlo computer simulation method with standard Metropolis sampling.[23] The protocol followed is essentially that employed in earlier work.[7,8] Each Monte Carlo step consists of rotations about randomly chosen dihedrals by a random amount in the range −180° to +180°, and perturbations of atoms by random amounts up to 0.005 Å in randomly chosen directions. Peptides were simulated for N_{cycle} cycles, where a cycle consists of $N_{dihedral}$ steps ($N_{dihedral}$ is set to equal the number of rotatable dihedrals in each peptide) plus an initial 10^4 cycles to equilibrate each peptide system. Conformations were collected for analysis after each 1,000 cycles, with the exception of the Arg simulation, where conformations were collected every 2,000 cycles. The total number of cycles generated in each simulation, N_{cycle}, is listed in Table II.

The AMBER/OPLS potential[24,25] was employed. This potential uses the united atom approximation: non-polar hydrogens are not modeled explicitly, rather the radii of

TABLE II. Number of Cycles, N_{cycle}, and Autocorrelations at lag 1, r_1, of the Conformational Energies in Each Simulation

Simulation	Peptide	N_{cycle} ($\times 10^6$)	r_1
Arg	Ace-Ala-**Arg**-Ala-NMe	6	0.40
Asn	Ace-Ala-**Asn**-Ala-NMe	1	0.41
Asp	Ace-Ala-**Asp**-Ala-NMe	1	0.38
Cys	Ace-Ala-**Cys**-Ala-NMe	1	0.38
Gln	Ace-Ala-**Gln**-Ala-NMe	2	0.45
Glu	Ace-Ala-**Glu**-Ala-NMe	2	0.41
His	Ace-Ala-**His**-Ala-NMe	1	0.38
Ile	Ace-Ala-**Ile**-Ala-NMe	1	0.42
Leu	Ace-Ala-**Leu**-Ala-NMe	1	0.42
Lys	Ace-Ala-**Lys**-Ala-NMe	3	0.46
Met	Ace-Ala-**Met**-Ala-NMe	2	0.40
Phe	Ace-Ala-**Phe**-Ala-NMe	1	0.39
Ser	Ace-Ala-**Ser**-Ala-NMe	1	0.33
Thr	Ace-Ala-**Thr**-Ala-NMe	1	0.40
Trp	Ace-Ala-**Trp**-Ala-NMe	1	0.45
Tyr	Ace-Ala-**Tyr**-Ala-NMe	1	0.42
Val	Ace-Ala-**Val**-Ala-NMe	1	0.39
Leu5	Ace-Ala_2-**Leu**-Ala_2-NMe	1.5	0.35
Leu7	Ace-Ala_3-**Leu**-Ala_3-NMe	2	0.42

the atoms to which they are attached are inflated slightly. Solvent was modeled as a dielectric continuum set to have the dielectric of bulk water ($\epsilon = 78$). The temperature, T, was set to 300 K.

Data Correlation and Error Estimation

Each collected conformation from a Monte Carlo simulation should have little or no correlation with the preceding collected conformation. Correlations between conformations were monitored using a standard measure, the autocorrelation at lag 1, r_1, of the conformational energies U.[26] This is estimated from

$$r_1 = \frac{\sum_{i=2}^{n} (U_i - \langle U \rangle)(U_{i-1} - \langle U \rangle)}{\sum_{i=1}^{n} (U_i - \langle U \rangle)^2}, \qquad (2)$$

where U_i is the energy of conformation i, $\langle U \rangle$ is the ensemble average energy and n is the total number of conformations used in the analysis. Generally, a dataset with $r_1 < 0.7$ is considered to possess sufficiently little correlation between successive data points for accurate estimation of conformational properties.[26] Calculated autocorrelations from each simulation are included in Table II.

Errors in the calculated side-chain entropies were estimated from the standard deviation in entropy within the last half of each simulation. The cumulative conformational entropy was calculated every 100 stored conformations in the second half of the simulation, with the standard deviation being determined from these calculated entropies. These entropies were calculated using all prior stored conformations, including those from the first half of the simulation. Standard deviations calculated using data from the first half of the simulation would be expected to be erroneously large due to insufficient sampling.

RESULTS

Equilibration and Autocorrelations

A major concern in all computer simulations is whether the simulation has been run long enough that sufficient conformations have been generated to approximate the equilibrium distribution. In this work the sampling of conformational space was tracked by observing the behavior of side-chain entropy as a function of conformations generated in the simulation. Figure 1 shows plots of entropy (presented as TS) versus number of conformations for four residues; Val, Leu, Met, and Arg. Similar plots (not shown) were obtained for all side-chains simulated. In all cases TS can be seen to approach an equilibrium value rapidly within the first half of the simulation, indicating that sufficient data was generated in all cases to approximate the equilibrium distribution of side-chain conformations.

The autocorrelation at lag 1, r_1, of the conformational energy was calculated in order to determine whether there was a sufficient lack of correlation between successive stored conformations in each simulation. The calculated autocorrelations for are shown in Table II. In all cases r_1 is significantly less than 0.7, indicating that the level of correlation was sufficiently low.[26]

Peptide Length Dependence

The dependence of estimated side-chain conformational entropy upon length of the simulated peptide was tested using the series of peptides Ace-$(Ala)_n$-**Leu**-$(Ala)_n$-NMe ($n = 1$, 2, and 3). The calculated conformational entropies for the Leu side-chain are given in Table III for these simulations (**Leu**, **Leu5**, and **Leu7**). Little variation in entropy is seen with increasing peptide length, indicating that, at least for relatively short peptides, the conformational behavior of the Leu side-chain is not dependent upon chain length. The side-chain rotamer distributions also do not differ significantly (data not shown).

Side-Chain Rotamer Distributions

χ_1 distributions obtained for each simulated residue are shown in Figure 2. The corresponding distributions from a dataset of protein structures[27] are shown for comparison. Distributions derived from the simulations are generally in good agreement with those observed in proteins. The simulation data does appear to slightly underestimate the population of the $g+$ conformation for many residues. This underpopulation of $g+$ is most noticeable for Ser, the simulation of which results in the $g-$ conformation being the most favorable. This is in contrast to the protein data, which suggests that $g+$ should be the more favorable conformation.

Side-Chain Conformational Entropy

Plots of the conformational entropies calculated in this work against three of the previously published scales are

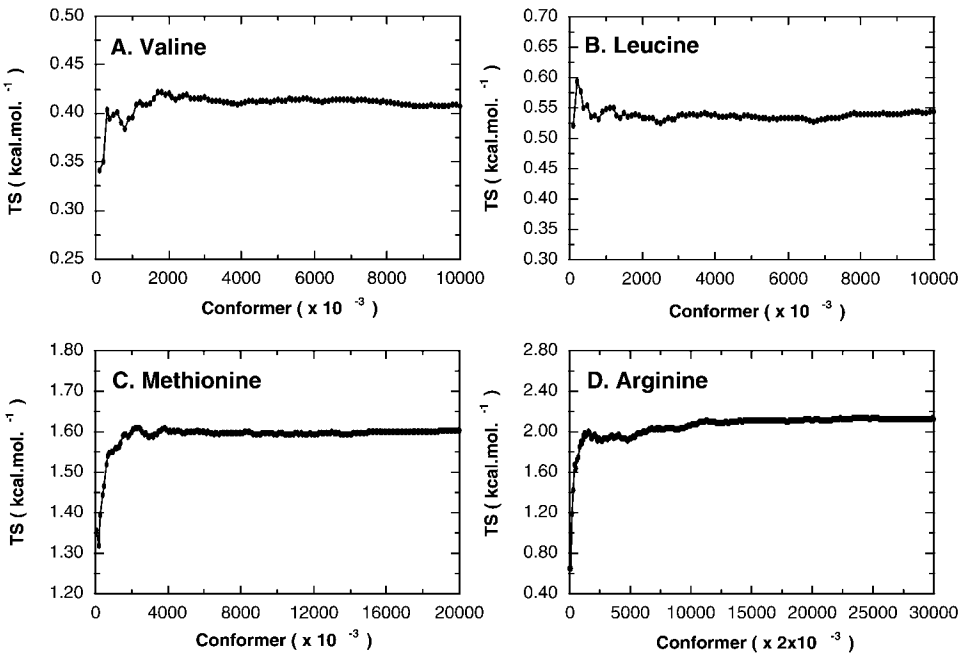

Fig. 1. Simulation equilibration plots for **A:** valine, **B:** leucine, **C:** methionine, and **D:** arginine. The cumulative average side-chain conformational entropy, *TS* (kcal mol^{-1}) is plotted against number of conformations generated in the simulation.

TABLE III. Leucine Side Chain Conformational Entropy as a Function of the Host Peptide Chain Length

Simulation	*TS* (kcal mol^{-1})
Leu	0.54 ± 0.01
Leu5	0.52 ± 0.01
Leu7	0.53 ± 0.01

shown in Figure 3. Comparisons are with the scales of Pickett and Sternberg,[6] Abagyan and Totrov,[13] and Lee et al.[12] The entropies estimated here were recalculated using the rotamer partitioning employed in each of these previous scales and are reported in Table IV. Excellent agreement is obtained with each of these scales. The correlation coefficients range from 0.93 upwards, intercepts are all close to zero and the slopes all approach one. Similar results are obtained when the entropies estimated here are compared to other scales such as those of Blaber et al.,[15] Koehl and Delarue,[14] and Sternberg and Chickos[16] (data not shown).

DISCUSSION

Models

One could question the relevance of using Ala-based peptides as hosts in the calculations presented here. It is certainly true that the results may well be context dependent. Polar residues in particular would be expected to show some level of context dependence—the presence of other polar side-chains nearby in sequence could lead to hydrogen bonding and/or electrostatic interactions that would have profound effects upon the conformational behavior of the side-chains and backbone. However, under current computing limitations it would be nearly impossible to simulate all possible tripeptide (or longer) sequences in a reasonable amount of time. Ala-based hosts are considered the best compromise.

Another point of contention is the use of a dielectric continuum as the model for solvent. Clearly an explicit solvation model would be better. However, such a model would lead to large increases in computation time, which would negate the major advantage of these calculations over previous side-chain entropy scales—the number of conformations generated. As noted previously, for short peptides such as the tripeptides examined here, use of a dielectric continuum is not expected to result in significantly different conformational behavior to that which would be observed using an explicit solvation model.[7,8] The good agreement between the rotamer distributions generated in this work and distributions derived from protein structures supports this argument (see Fig. 2).

The results in Table III demonstrate that conformational behavior of a Leu side-chain inserted into the center of an Ala-based peptide is independent of peptide length. It would of course be of interest to simulate even longer peptides in order to examine whether this independence holds in the presence of longer-range interactions. However, longer peptides tend to "collapse" into compact conformations during the early parts of a simulation and do not adequately sample conformational space.

Rotamer Distributions

Very good agreement was obtained when comparing χ_1 rotamer distributions from the simulations with distributions derived from a protein dataset by MacGregor et al.[27] (Fig. 2). The calculated distributions do apparently slightly overestimate the populations of the $g-$ rotamers and consequently underestimate $g+$ populations. There are at least two probable causes for this. One is that the dataset used by MacGregor et al.[27] was too small to be truly

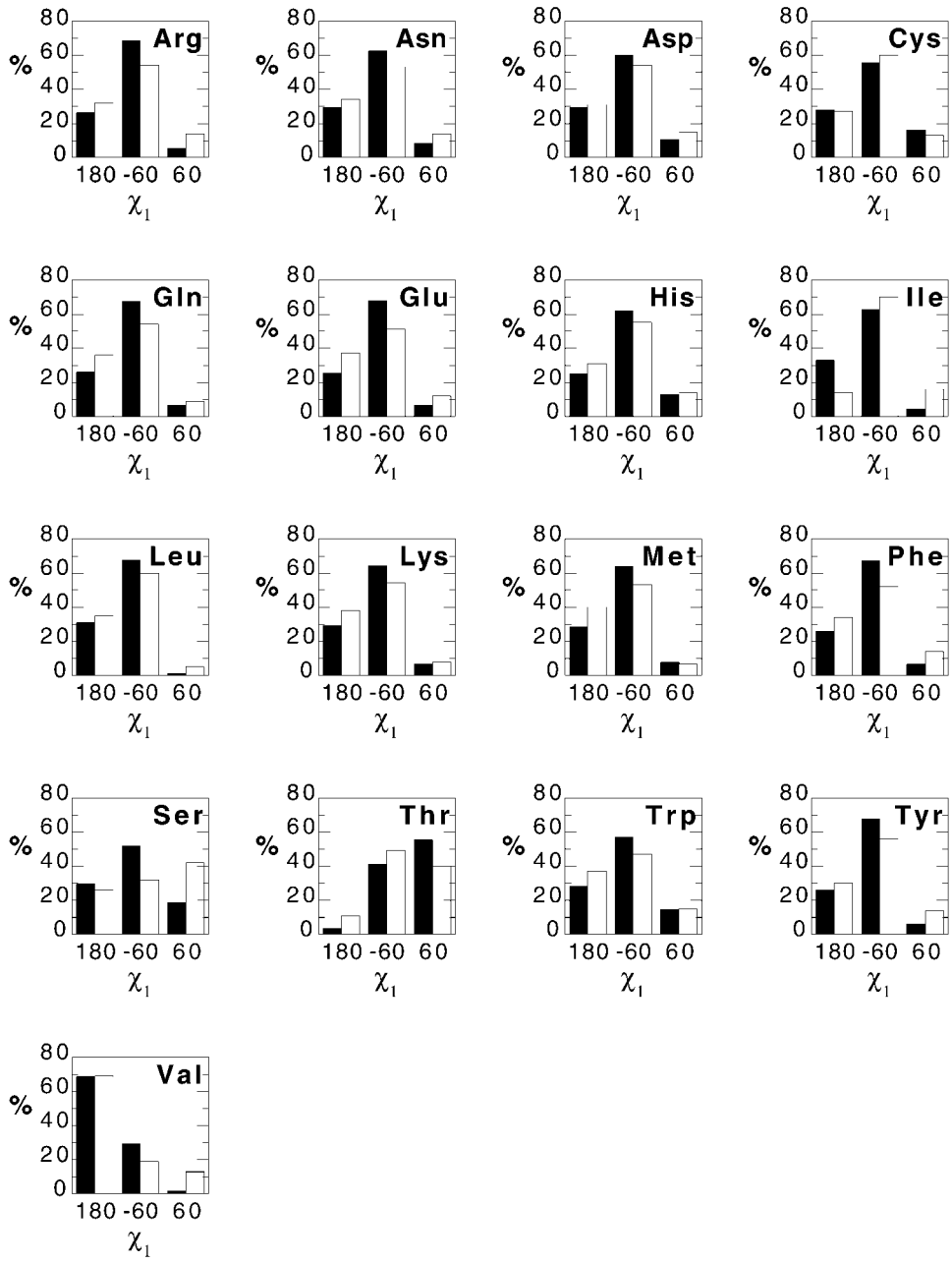

Fig. 2. Side-chain χ_1 rotamer distributions from Monte Carlo computer simulations (black bars) and from a dataset of protein structures surveyed by MacGregor et al.[27] (white bars).

representative. This dataset comprised 61 protein chains. This explanation is probably not the correct one. Although the dataset used by MacGregor et al. is small by current standards, more recent studies of rotamer distributions using larger datasets have not produced significantly different results. For example, Dunbrack and Karplus[28] used a dataset of 132 protein chains and obtained much the same results as MacGregor et al.[27] for side-chains in all structure types. The second possibility for overpopulation of $g-$ is inaccuracies in the forcefield/solvation model employed. The forcefield used here, AMBER/OPLS,[24,25] uses the united atom approximation, and a dielectric continuum model for solvent was used. Use of a forcefield that explicitly models non-polar hydrogens and/or use of a explicit solvent model may alleviate this slight overpopulation.

Comparisons With Previous Estimates

As noted above, the agreement between side-chain conformational entropies estimated in this work and those calculated by others is excellent (see Figure 3 and Table IV). Care was taken to recalculate entropies from this work using the rotamer partitioning employed in the previous scales. Similar agreement is obtained when comparing to scales not included in Figure 3. This level of agreement is encouraging—it indicates that side-chain entropies estimated here are independent of the methods and models employed. This excellent agreement also pro-

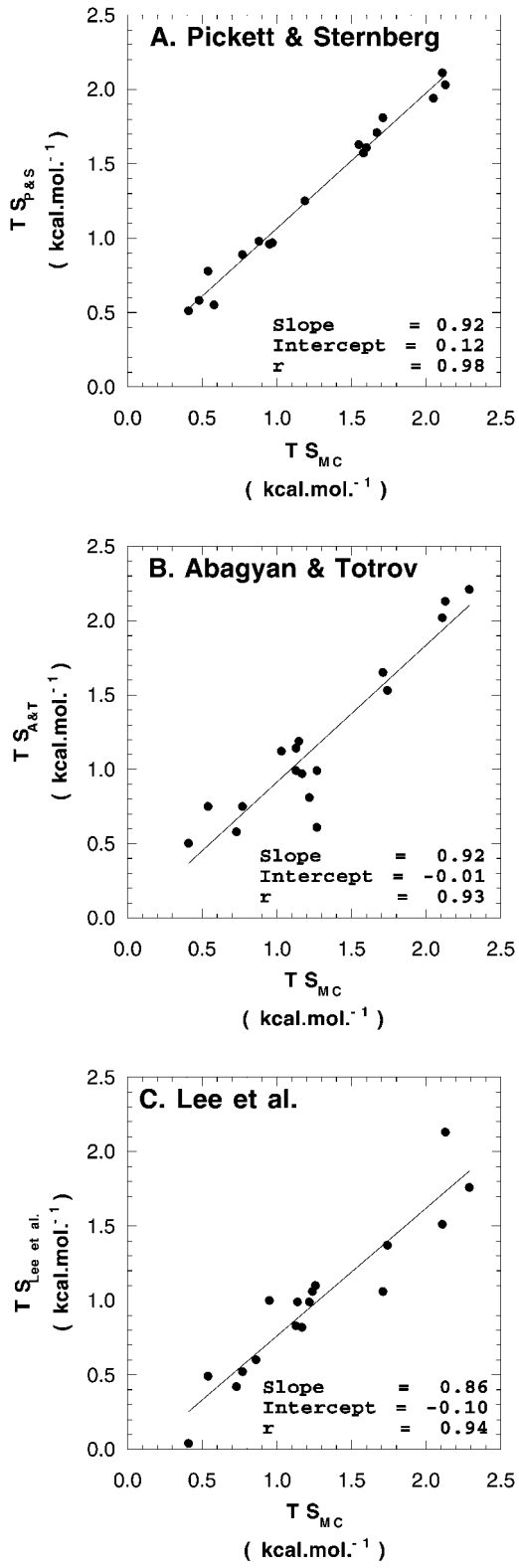

Fig. 3. Comparison of side-chains entropies, *TS* (kcal mol^{-1}) estimated in this work with previous scales: **A:** Pickett and Sternberg,[6] **B:** Abagyan and Totrov,[13] and **C:** Lee et al.[12]

vides support for use of the simple models employed here. Estimated entropies agree well with those derived from protein structure datasets.[6] Side-chains in such scales are in environments that differ significantly from a simple Ala-based tripeptide. Excellent agreement is also achieved with scales derived from computational methods.[12,13] Abagyan and Totrov[13] used the ECEPP energy function[29] with modifications to the treatment of electrostatics and solvation. Agreement of the entropies calculated here with the Abagyan and Totrov scale indicates that the estimates are not significantly dependent upon the force field and solvation model used. Of course one would hope that different force fields give similar results since the force fields themselves are not remarkably different from one another. Further advancements in accuracy using these kinds of calculations are likely to be small at best.

Effective Number of Rotamers

Pickett and Sternberg[6] defined a quantity called the effective number of rotamers, r_{Eff}. This is obtained from estimated side-chain entropies using the relation

$$r_{Eff} = e^{S/R} \tag{3}$$

where S is the entropy of the side-chain and R is the gas constant. This is essentially a measure of the number of rotamers populated by the side-chain assuming that the rotamers are all populated equally. The effective number of rotamers for each side-chain simulated in this work are given in Table V.

It is instructive to compare r_{Eff} to the total possible number of rotamers, r_{Total} (Table V). This comparison can be facilitated by considering the quantity f_r, the ratio of effective to total rotamers (r_{Eff}/r_{Total}). This ratio is given in Table V and plotted against the number of rotatable bonds in each side-chain in Figure 4. Some side-chains behave much as would be expected. For example, of the four residues with just the one side-chain rotatable bond, Val and Thr have the lowest f_r ratios, indicative of the steric interference experienced by β-branched residues when rotating about χ_1.[7,8,30,31] Cys and Ser, on the other hand, do not appear to be significantly affected by steric interactions and have f_r ratios almost equal to one. Leu and Ile have low values of f_r, consistent with the observation that these side-chains each predominantly populate just two rotamer classes.[27,32] One should not, however, assume that r_{Eff} is equal to the number of rotamer classes populated. For example, eight of the nine possible rotamer classes are populated by Leu (r_{Eff} = 2.5), and all 27 possible rotamer classes are populated by the Met side-chain (r_{Eff} = 14.9).

Strikingly, as the number of rotatable bonds increases, there is a decrease in f_r: the fraction of significantly populated rotamer classes decreases as the potential number of degrees of freedom increases (Fig. 4). This indicates that even rotations about χ_3 and χ_4 of the longer side-chains are affected by interactions with the tripeptide host. This leads to the somewhat surprising observation that shorter side-chains possess most of the entropy theoretically available to them in unfolded states, while

TABLE IV. Unfolded Side Chain Conformational Entropies (kcal mol^{-1})

Side chain	TS_{MC} (kcal mol^{-1})	Pickett and Sternberg[b]		Abagyan and Totrov[c]		Lee et al.[d]	
		TS_{MC}[e]	$TS_{P\&S}$	TS_{MC}[e]	$TS_{A\&T}$	TS_{MC}[e]	TS_{Lee}
Arg	2.13 ± 0.01	2.13	2.03	2.13	2.13	2.13	2.13
Asn	1.22 ± 0.01	1.58	1.57	1.22	0.81	1.22	0.99
Asp[a]	0.86 ± 0.01	1.19	1.25	1.27	0.61	0.86	0.60
Cys	0.58 ± 0.01	0.58	0.55	1.13	1.14	1.24	1.06
Gln	1.79 ± 0.01	2.11	2.11	2.11	2.02	2.11	1.51
Glu[a]	1.39 ± 0.01	1.71	1.81	1.71	1.65	1.71	1.06
His	1.27 ± 0.01	1.27	0.96	1.27	0.99	0.95	1.00
Ile	0.77 ± 0.01	0.95	0.89	0.77	0.75	0.77	0.52
Leu	0.54 ± 0.01	0.54	0.78	0.54	0.75	0.54	0.49
Lys	2.06 ± 0.01	2.06	1.94	2.29	2.21	2.29	1.76
Met	1.60 ± 0.01	1.60	1.61	1.74	1.53	1.74	1.37
Phe[a]	0.73 ± 0.01	0.48	0.58	0.73	0.58	0.73	0.42
Ser	0.60 ± 0.01	1.67	1.71	1.15	1.19	1.26	1.10
Thr	0.48 ± 0.01	1.55	1.63	1.03	1.12	1.14	0.99
Trp	1.17 ± 0.01	0.97	0.97	1.17	0.97	1.17	0.82
Tyr[a]	0.72 ± 0.01	0.88	0.98	1.13	0.99	1.13	0.83
Val	0.41 ± 0.01	0.41	0.51	0.41	0.50	0.41	0.04

[a]Corrected for symmetry.
[b]Pickett & Sternberg.[6]
[c]Abagyan & Totrov.[13]
[d]Lee et al.[12]
[e]Entropies from this work were recalculated using the appropriate rotamer partitioning for comparison to the other scales.

TABLE V. Effective Number of Rotamers

Side chain	r_{Eff}	r_{Total}	f_r	Number of rotatable bonds
Arg	36.5	81	0.45	4
Asn	7.8	12	0.65	2
Asp	4.3	6	0.72	2
Cys	2.7	3	0.90	1
Gln	20.6	36	0.57	3
Glu	10.4	18	0.58	3
His	8.5	12	0.71	2
Ile	3.7	9	0.41	2
Leu	2.5	9	0.28	2
Lys	32.4	81	0.40	4
Met	14.9	27	0.55	3
Phe	3.4	6	0.57	2
Ser	2.7	3	0.90	1
Thr	2.2	3	0.73	1
Trp	7.2	12	0.60	2
Tyr	3.4	6	0.57	2
Val	2.0	3	0.67	1

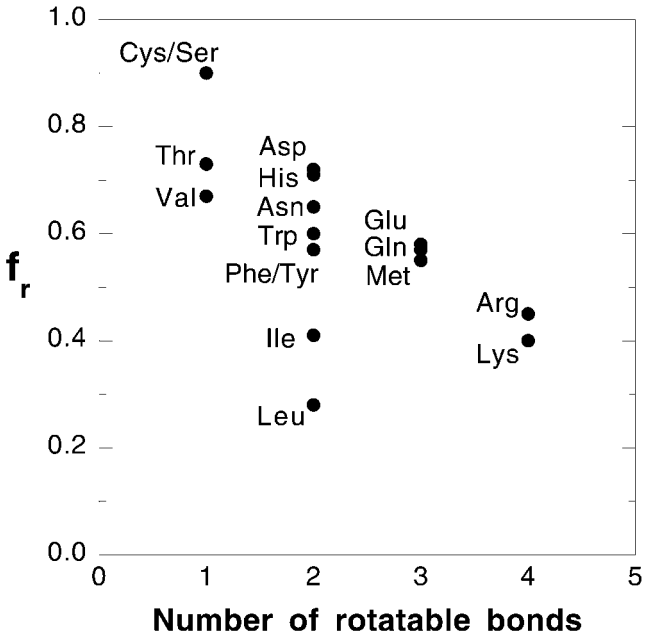

Fig. 4. Plot of the ratio of effective to total number of rotamers, f_r, against number of rotatable bonds in the side-chain.

longer side-chains possess a smaller fraction of the theoretical maximum amount of entropy.

In previous scales it has often been assumed that rotations about side-chain torsions after χ_2 are unaffected by the polypeptide backbone.[5,6,13] In these previous scales additional terms for χ_3 and χ_4 are added to the estimated entropies. Figure 4 demonstrates that such assumptions are incorrect: rotations about χ_3 and χ_4 are clearly affected by interactions with the tripeptide host. These interactions include steric clashes, hydrophobic interactions with the methyl side-chains of the adjacent Ala residues, and potential hydrogen bonds to the backbone. The effect of such interactions upon long side-chains cannot be ignored.

Use of Scale

It should be emphasized that the entropy scale presented here should be used in conjunction with folded state entropy estimates calculated using the same rotamer partitioning scheme. These are not absolute entropies—

they will vary depending upon rotamer definitions. This can be seen simply by comparing previous scales to one another as well as to the scale presented here (Table IV). If used correctly, this scale will prove a valuable tool for understanding and dissecting the energetics of protein folding, in ligand binding and design, and could prove useful in protein engineering.

ACKNOWLEDGMENTS

The author wishes to thank Ed Lattman for helpful suggestions.

REFERENCES

1. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
2. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
3. D'Aquino JA, Gomez 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.
4. Bromberg S, Dill KA. Side-chain entropy and packing in proteins. Protein Sci 1994;3:997–1009.
5. Doig AJ, Sternberg MJE. Side-chain conformational entropy in protein folding. Protein Sci 1995;4:2247–2251.
6. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
7. Creamer TP, Rose GD. α-Helix-forming propensities in peptides and proteins. Proteins 1994;19:85–97.
8. 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.
9. Yang D, Mok YK, Forman-Kay JD, Farrow NA, Kay LE. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. J Mol Biol 1997;272: 790–804.
10. Lee AL, Kinnear SA, Wand AJ. Redistribution and loss of side-chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 2000;7:72–77.
11. Creamer TP, Rose GD. Simple force field for study of peptide and protein conformational properties. Methods Enzymol 1995;259: 576–589.
12. 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.
13. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994;235:983–1002.
14. 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.
15. Blaber M, Zhang Xj, Lindstrom JD, Pepoit SD, Baase WA, Matthews 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.
16. Sternberg MJE, Chickos JS. Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales. Protein Eng 1994;7:149–155.
17. Némethy G, Leach SJ, Scheraga HA. The influence of amino acid side-chains on the free energy of helix-coil transition. J Phys Chem 1966;70:998–1004.
18. Luque I, Mayorga OL, Freire E. Structure-based thermodynamic scale of α-helix propensities in amino acids. Biochemistry 1996;35: 13681–13688.
19. Luque I, Todd MJ, Gomez J, Semo N, Freire E. Molecular basis of resistance to HIV-1 protease inhibition: a plausible hypothesis. Biochemistry 1998;37:5791–5797.
20. Luque I, Gomez J, Semo N, Freire E. Structure-based thermodynamic design of peptide ligands: application to peptide inhibitors of the aspartic protease endothiapepsin. Proteins 1998;30:74–85.
21. Doig AJ, Gardner M, Searle MS, Williams DH. Thermodynamics of side-chain internal rotations—effects on protein structure and stability. Tech Protein Chem 1993;IV:557–566.
22. IUPAC-IUB. IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of the conformation of polypeptide chains. J Mol Biol 1970;52:1–17.
23. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087–1092.
24. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for proteins. Energy minimization for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
25. Weiner SJ, Kollman PA, Case DA, et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.
26. Kolafa J. Autocorrelations and subseries averages in Monte Carlo simulations. Mol Phys 1986;59:1035–1042.
27. MacGregor MJ, Islam SA, Sternberg MJE. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. J Mol Biol 1987;198:295–310.
28. Dunbrack Jr. RL, Karplus M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 1993;230:543–574.
29. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 1975;79:2361–2381.
30. Walther D, Argos P. Intrahelical side-chain-side-chain contacts: the consequences of restricted rotameric states and implications for helix engineering and design. Protein Eng 1996;9:471–478.
31. Schrauber H, Eisenhaber F, Argos P. Rotamers: to be or not to be? An analysis of amino acid side-chain conformation in globular proteins. J Mol Biol 1993;230:592–611.
32. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193:775–791.