PROTEINS: Structure, Function, and Genetics 25:143–156 (1996)

RESEARCH ARTICLES

The Magnitude of the Backbone Conformational Entropy Change in Protein Folding

J. Alejandro D'Aquino,[1] Javier Gómez,[1] Vincent J. Hilser,[1] Kon Ho Lee,[2,3] L. Mario Amzel,[2] and Ernesto Freire[1]

Department of [1]Biology and Biocalorimetry Center and [3]Physics and Astronomy, The Johns Hopkins University, Baltimore, Maryland 21218; and [2]Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

ABSTRACT **The magnitude of the conformational entropy change experienced by the peptide backbone upon protein folding was investigated experimentally and by computational analysis. Experimentally, two different pairs of mutants of a 33 amino acid peptide corresponding to the leucine zipper region of GCN4 were used for high-sensitivity microcalorimetric analysis. Each pair of mutants differed only by having alanine or glycine at a specific solvent-exposed position under conditions in which the differences in stability could be attributed to differences in the conformational entropy of the unfolded state. The mutants studied were characterized by different stabilities but had identical heat capacity changes of unfolding (ΔC_p), identical solvent-related entropies of unfolding (ΔS_{solv}), and identical enthalpies of unfolding (ΔH) at equivalent temperatures. Accordingly, the differences in stability between the different mutants could be attributed to differences in conformational entropy. The computational studies were aimed at generating the energy profile of backbone conformations as a function of the main chain dihedral angles ϕ and φ. The energy profiles permit a direct calculation of the probability distribution of different conformers and therefore of the conformational entropy of the backbone. The experimental results presented in this paper indicate that the presence of the methyl group in alanine reduces the conformational entropy of the peptide backbone by 2.46 ± 0.2 cal/K · mol with respect to that of glycine, consistent with a 3.4-fold reduction in the number of allowed conformations in the alanine-containing peptides. Similar results were obtained from the energy profiles. The computational analysis also indicates that the addition of further carbon atoms to the side chain had only a small effect as long as the side chains were unbranched at position β. A further reduction with respect to Ala of only 0.61 and 0.81 cal/K · mol in the backbone entropy was obtained for leucine and lysine, respectively. β-branching (Val) produces the largest decrease in conformational entropy (1.92 cal/K · mol less than Ala). Finally, the backbone entropy change associated with the unfolding of an α-helix is 6.51 cal/K · mol for glycine. These and previous results have allowed a complete estimation of the conformational entropy changes associated with protein folding.**

INTRODUCTION

The folding of a protein involves a large reduction in the number of conformations accessible to the backbone and side chains of the polypeptide. This reduction gives rise to a large negative conformational entropy change, ΔS_{conf} and constitutes the main unfavorable force that needs to be overcome for the stabilization of the native state. A precise estimation of the conformational entropy change is critical for accurate stability predictions since relatively small variations in entropy result in large changes in protein stability. For a typical globular protein, an error of ± 5% in the predicted entropy change results in an error of ± 15°C on the predicted denaturation temperature of the molecule. Certainly, structure-based stability calculations with the accuracy needed for protein design require entropy predictions with an error lower than ± 5%. A similar situation occurs in structure-based calculations of the binding affinities of peptides to proteins or protein-protein associations. In these cases, precise estimations of the backbone and side chain con-

Received September 7, 1995; revision accepted November 21, 1995.

Address reprint requests to Ernesto Freire, Department of Biology and Biocalorimetry Center, The Johns Hopkins University, Baltimore, MD 21218.

formational entropy changes are essential for the accurate prediction of binding constants.

The conformational entropy change has remained one of the most elusive thermodynamic parameters to estimate. In the past, two approaches have been used to evaluate ΔS_{conf}: a purely computational approach aimed at estimating the number of conformations accessible to the polypeptide; and an empirical dissection aimed at subtracting the contributions of factors other than conformation from experimental entropy changes.[1-4] Recently, backbone entropies in the unfolded state have also been approximated from the distribution of main chain dihedral angles ϕ and φ in crystallographic structures.[5] This approach, however, might be influenced by specific preferences existing in the native state that drive the ensemble of backbone conformations away from that existing in the unfolded state. The empirical methods have yielded estimates anywhere between 4 and 20 cal/K · mol for the average conformational entropy per residue. The higher estimates[4] are about fivefold larger than those obtained by computational methods.[6-9] To gain direct access to the different contributions to the conformational entropy, in this paper we have examined a unique case in which the main chain conformational entropy is experimentally accessible: the difference in backbone entropy associated with an Ala → Gly replacement in a peptide chain. These experimental results have been compared with computational estimates obtained from calculations of the Boltzmann weighted probability of the different conformers available to the peptide chain.

The conformational entropy change, ΔS_{conf}, for peptide folding/unfolding can be considered as being composed of three different terms: 1) $\Delta S_{bu \rightarrow ex}$, the entropy change associated with the transfer of a side chain that is buried in the interior of the protein to its surface; 2) $\Delta S_{ex \rightarrow u}$, the entropy change gained by a surface-exposed side chain when the peptide backbone unfolds; and 3) ΔS_{bb}, the entropy change gained by the backbone itself upon unfolding.[9-11] In this paper we will be concerned with this third term.

The difference in stabilities between two otherwise identical coiled coils in which a solvent-exposed amino acid position is occupied by either Gly or Ala can be attributed to the difference in backbone entropies of the two peptides since the peptides exhibit no differences in internal modes, no side chain entropies are involved, and the two residues exhibit the same degree of solvation in both the folded and unfolded states. In addition, the change in vibrational entropy for rotation around the C_α–C_β bond is negligible. If these conditions are satisfied, the peptide unfolding transitions are expected to exhibit identical ΔC_p of unfolding and identical ΔH of unfolding when compared at the same temperature. This situation can be experimentally realized and tested using solvent-exposed positions in isolated leucine zippers. For these studies, different mutants of a 33 amino acid peptide corresponding to the leucine zipper region of GCN4 were used. The structure of this region of GCN4 is known at 1.8 Å resolution.[12] It forms a dimeric, parallel, two-stranded coiled coil of α-helices called the leucine zipper. The two strands are packed as in the "knobs-into-holes" model proposed by Crick in 1953. The region is highly conserved in proteins containing leucine zippers and is defined by a heptad repeat, $(abcdefg)_n$ in which the generally hydrophobic residues at positions *a* and *d* are on the same face of each of the helices in apposition to each other in the coiled coil.[12-14] In this paper we have used four different mutants of the 33 amino acid coiled coil forming peptide corresponding to the leucine zipper region of GCN4:

RMKQLEDKVEELL<u>S</u>KNYHLENEV<u>A</u>RLKKLVGER

The mutations were made at two solvent-exposed positions located at different distances from the ends, 14 and 24, which are occupied by a serine and an alanine in the wild-type peptide (underlined in the sequence above). The effects of the mutations on the stability of the helices were studied by experimental and computational methods.

MATERIALS AND METHODS

Peptide Synthesis

The peptides were synthesized using the solid-phase technique[39] on a Milligen model 9050 peptide synthesizer. Purity was verified by reversed-phase high-performance liquid chromatography (HPLC) on a Waters Delta Pak C18 column using a linear gradient of water/acetonitrile, both containing 0.1% trifluoroacetic acid (TFA). Amino acid analysis was performed using the Waters Pico-Tag system. ACES (N-[2-Acetamido]-2-amino ethane sulfonic acid) buffer was obtained from Sigma. Spectra/por 3 dialysis membrane was purchased from Spectrum. Prior to calorimetric experiments the peptide was dissolved in 10 mM ACES, pH 7.0, 0.25 mM EDTA and dialyzed overnight against the same buffer through Spectra/por 3 dialysis membrane. The dimeric nature of the peptide was checked by analytical centrifugation using a Beckman XL-A instrument as described before.[17,18]

Differential Scanning Calorimetry (DSC)

Peptide samples were diluted to the desired concentration, 7×10^{-4} M determined spectrophotometrically at 274 nm,[18,20] and scanned at 60°C h^{-1} in a DASM DS-92 microcalorimeter under a pressure of 17 psi. The calorimeter was interfaced to a PC microcomputer equipped with a Data Translation DT-2801 A/D converter board for instrument control and automatic data collection. The excess heat capacity function was analyzed after scan rate normalization, concentration normalization, and

baseline subtraction using software developed in this laboratory for thermal transitions exhibiting dimer to monomer equilibrium.[19,20] Individual baseline buffer scans were performed for each experiment presented in this paper. In the peptide concentration range (2.8 mg/ml) used in the DSC experiments, the dependence of the transition temperature on peptide concentration that characterizes unfolding coupled to dissociation is already in the plateau region[18,20] and is not expected to change by more than 0.05°C due to experimental uncertainties in concentration determination.

Surface Area Calculations

Surface areas calculations were performed as described before[3] using the implementation of Lee and Richard's algorithm[40] in the program ACCESS (Scott R. Presnell, University of California, San Francisco, CA), with a probe radius of 1.4 Å and a slice width of 0.25 Å.

Computational Analysis

For all calculations, dipeptides of the form N-acetyl-(*aminoacyl*)-N'-methylamide were used. The dipeptides were generated in the fully extended conformation using an in-house program.[9] AMBER/OPLS parameters[42] were used for bond lengths, and bond and torsion angles. The main chain dihedral angles ϕ and φ were set initially to the fully extended *trans* conformation. The side chains were set to all *trans* conformations. Coordinates were generated for dipeptides containing Gly, Ala, Val, Leu, and Lys as the aminoacyl residue.

Energies were calculated for all the dipeptides for all combinations of the main chain dihedral angles ϕ and φ between 0° and 360° every 10°. The energy of every ϕ and φ combination was minimized with the program XPLOR (Brunger, 1990)[43] using the CONSTRAINTS DIHEDRAL instruction with a large force constant (400 kcal/mol) to keep ϕ and φ at the preset values. (All other parameters were AMBER/OPLS.) Energies were calculated for the final, minimized conformations without including the additional dihedral term.

Entropies were calculated using the expression

$$S = -R\sum P_i ln P_i$$

where

$$P_i = \exp(-E_i/kT)/\sum \exp(-E_i/kT).$$

The energies E_i are those calculated as described above. Alternatively, the summation can be replaced by an integral over ϕ and φ. To ensure that the changes in entropy between Gly and the other amino acids calculated in this way do not depend on the size of the sampling interval, the sum (or integration) was calculated using different intervals.

The main chain conformational entropy of the folded state was estimated by calculating the energy profile for a nine residue helix. The calculations were carried out for two helices $(Ala)_9$ and $(Gly)_9$. Structures corresponding to all combinations of ϕ and φ within a 20° interval from the helical conformation were generated and their energies minimized following the same procedure used for the dipeptides. The entropy per residue was obtained using the same equations used for the dipeptides.

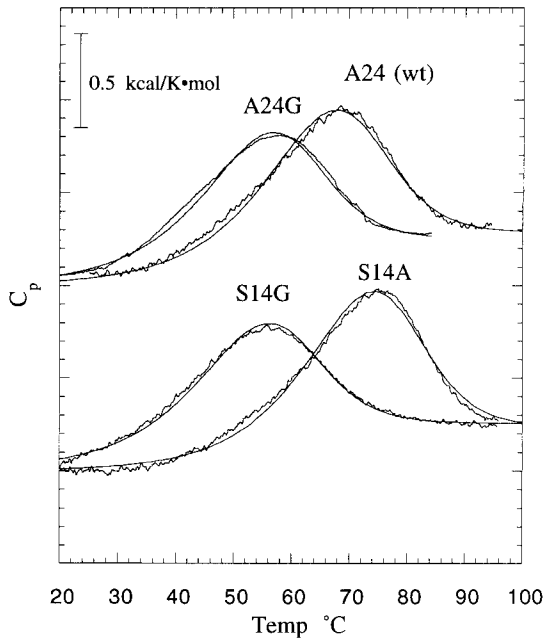


Fig. 1. Excess heat capacity functions of four different mutants of a peptide corresponding to the leucine zipper region of GCN4. The top and bottom pairs of curves correspond to mutants in which the only difference is an Ala or Gly at position 24 or at position 14. The experiments were performed in 10 mM ACES, 0.25 mM EDTA, pH 7.0, using a peptide concentration of 2.8 mg/mL. The solid lines represent the fitting of the experimental data to a two-state dimeric model as described before.[18,20]

Estimation of the Fraction of Conformations Made Inaccessible by Excluded Volume

The number of conformations that are accessible to a chain of n residues in a cubic lattice is equal to 6^n if excluded volume is not taken into consideration. For a chain in which any residue is not allowed to occupy the same volume as the previous one (nearest neighbor excluded volume), the total number of conformations is $6 \times 5^{n-1}$. The total number of allowed conformations considering excluded volume is[22,41]

$$n^{\gamma} \cdot \mu^n$$

where $\gamma = 1/6$ and $\mu \approx 4.68$. The fraction of accessible conformations for a chain of n residues, $F_{acc}(n)$, is

$$F_{acc}(n) = n^{\gamma} \cdot \mu^n/6 \times 5^{n-1}.$$

In the above equation $F_{acc}(n)$ is not the total fraction of allowed conformations but the fraction of allowed conformations within the subensemble in which the

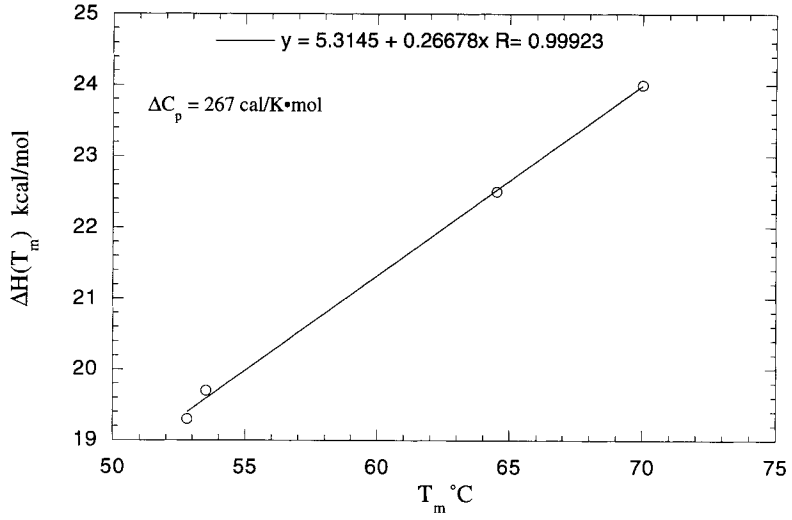


Fig. 2. The temperature dependence of the enthalpy change for the four peptides considered in this study. The four values define a straight line with a slope (heat capacity) of 267 cal/K · mol similar to the value obtained from the individual heat capacity curves.

TABLE I. Thermodynamic Parameters Associated With the Temperature Unfolding of GCN4-33 Coiled Coil Mutants*

Peptide	$T_{m,1/2}$ (°C)	$\Delta H(T_{m,1/2})$ (kcal/mol)	$\Delta S(T_{m,1/2})$ (cal/K · mol)	$\Delta S(65)$ (cal/K · mol)
S14A/S14A	70.0	24.0	63.07	59.25 ± 0.1
S14G/S14G	52.8	19.3	52.35	61.9 ± 0.2
A24/A24 (wild type)	64.5	22.5	59.64	60.03 ± 0.01
A24G/A24G	53.5	19.7	53.33	62.3 ± 0.2

*$T_{m,1/2}$ is defined as the temperature at which the transition is half completed. The enthalpy, entropy, and heat capacity changes are expressed on a per mole of protein monomer basis.

excluded volume of the nearest neighbor is already taken into consideration. The reason for this definition is that the computation for dipeptides already includes the excluded volume of nearest neighbors.

RESULTS

Differential Scanning Calorimetry

Figure 1 shows the excess heat capacities for the wild type and the mutants S14A, S14G, and A24G measured at pH 7.4 and using the same peptide concentrations. In all cases the two peptide strands that form the coiled coils are identical, i.e., the coiled coils are S14A/S14A, S14G/S14G, A24G/A24G, and A24/A24 (wild type). The four peptides define two pairs of leucine zippers that differ only by an Ala→Gly substitution, (S14A/S14A, S14G/S14G) and (A24G/A24G, A24/A24). It is evident that the two Ala mutants have the highest stability. It must be noted also that the four mutants exhibit similar heat capacity changes (ΔC_p = 255 ± 12 cal/K · mol), as indicated by the difference in the heat capacity values measured after and before their thermal transitions. Since the changes in heat capacity are proportional to the changes in polar and apolar surface areas (ΔASA_{pol} and ΔASA_{ap}) that become accessible to the solvent upon unfolding,[3,15,16] it can be concluded that the four mutants exhibit similar ΔASA changes, independently of the nature of the residue at position 14 or 24. This situation is expected to occur if those residues are already exposed to the solvent in the folded state and therefore their degree of exposure to the solvent does not change measurably upon unfolding.

The enthalpy changes (ΔH) of the four peptides exhibit a temperature dependence consistent with the observed ΔC_p values and are equivalent when evaluated at the same temperature (Table I and Fig. 2). Within the temperature range of the experiments, the enthalpy change of the four peptides as a function of the temperature (T, in °C) is well represented by the equation:

$$\Delta H(T) = (22{,}636 \pm 400)\text{cal/mol} + (267 \pm 7)\text{cal/K·mol·}(T - 65) \quad (1)$$

with a regression coefficient of 0.9994. The value 22,636 cal/mol is the unfolding enthalpy of any of the four peptides at 65°C. The ΔC_p of 267 cal/K · mol obtained from the temperature dependence of the

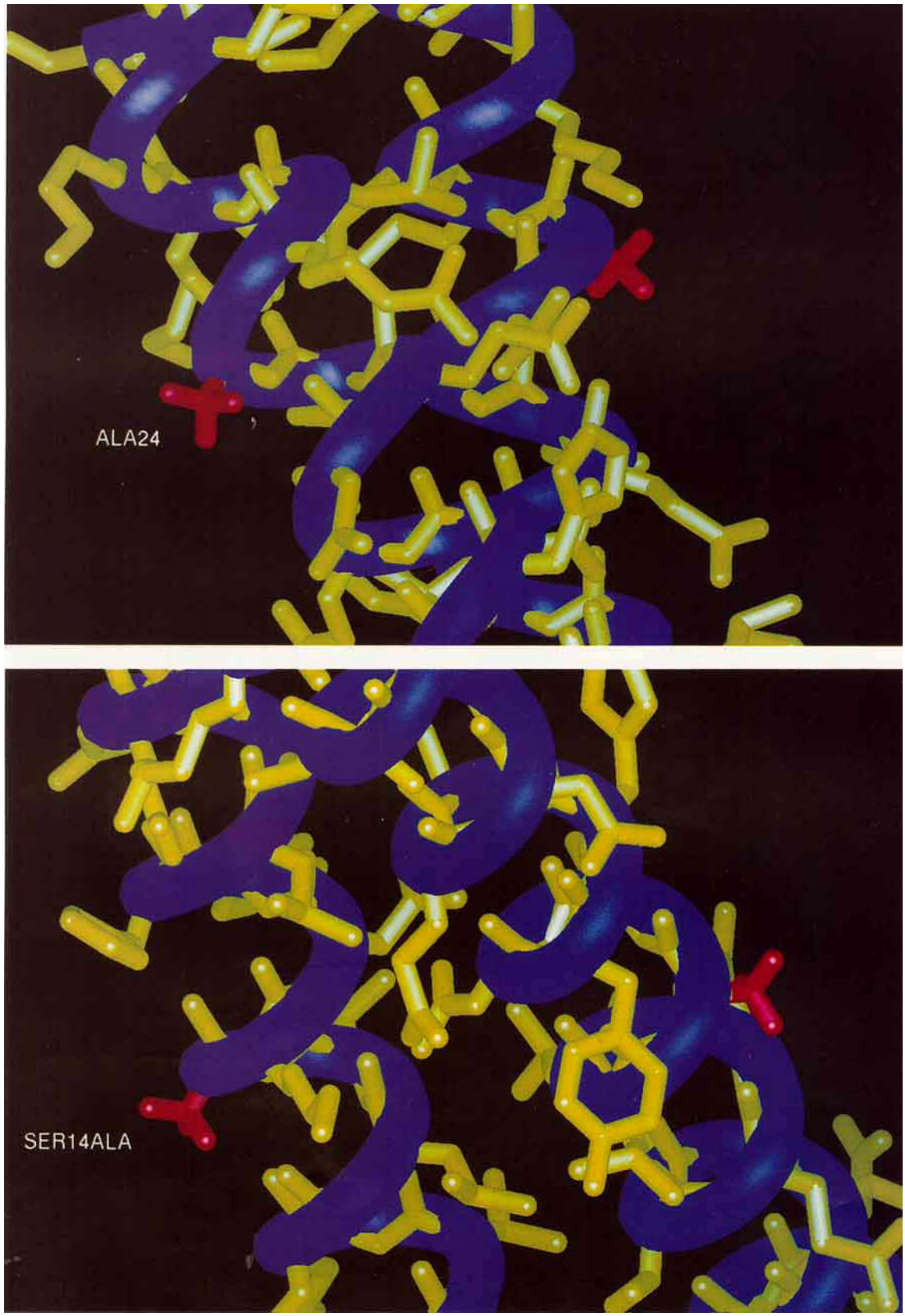
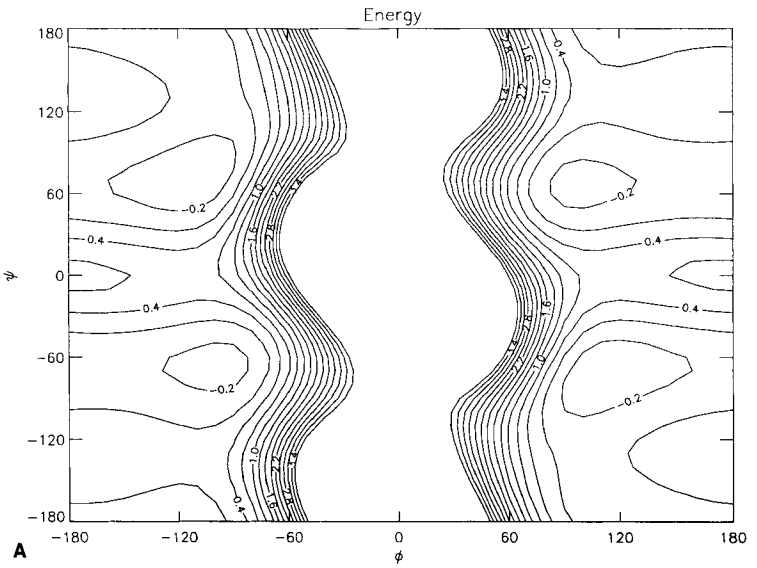
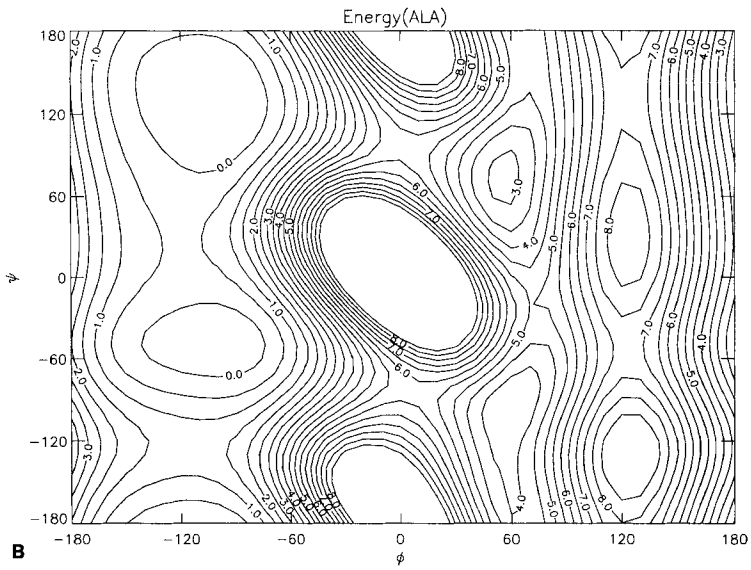


Fig. 3. Graphical representation of the mutation sites at positions 14 and 24. In both cases, these two positions are fully exposed to the solvent in the native state and are not expected to undergo measurable changes in solvation upon unfolding.


Energy
180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
φ
A


Energy(ALA)
180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
φ
B

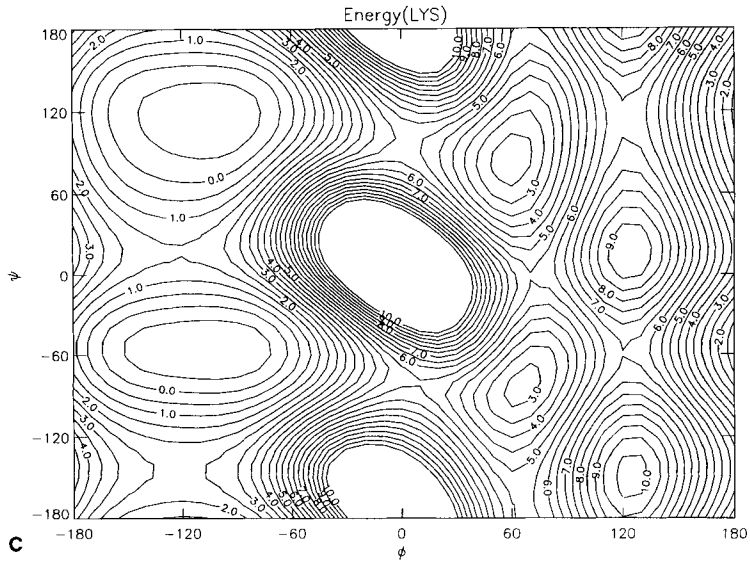

Energy(LYS)
180
120
60
0
-60
-120
-180
ψ
-180 -120 -60 0 60 120 180
φ
C

Fig. 4a–c.

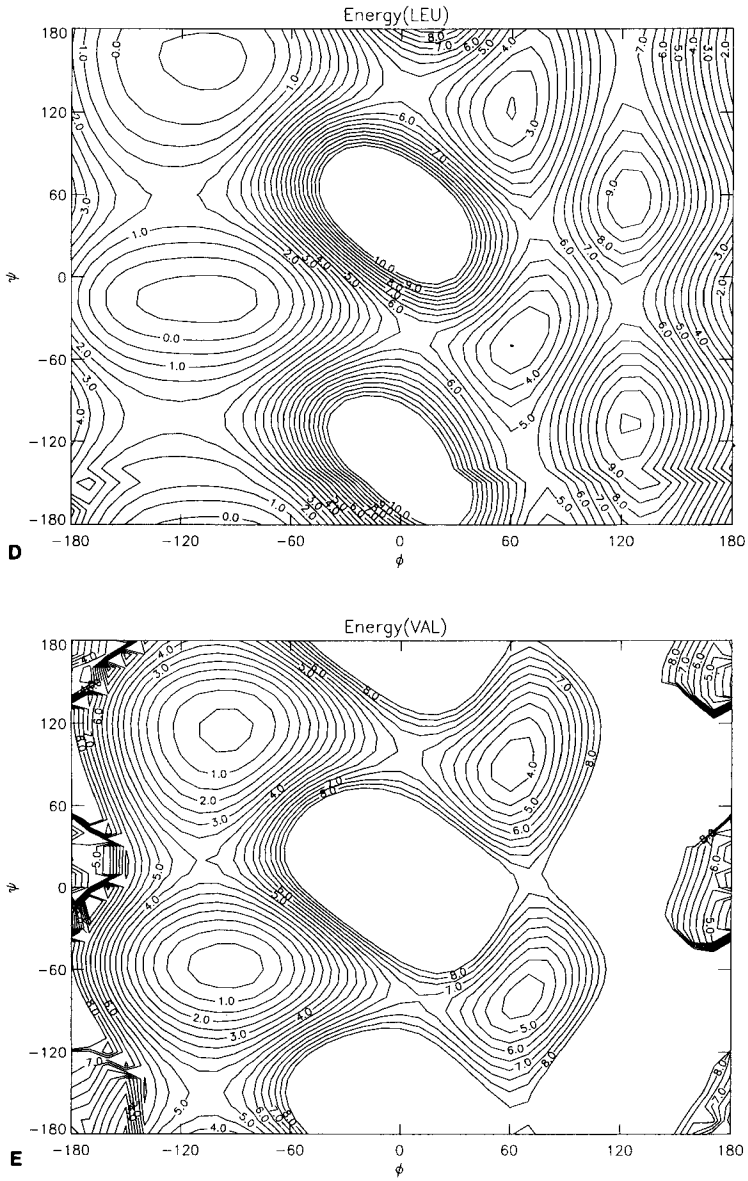


Fig. 4. Backbone energy profiles. The energies of the N-acetyl-(amino acyl)-N′-methyl amides were calculated and energy-minimized as described in Materials and Methods. Points were calculated for all ϕ and φ combinations taken every 10°. Profiles for Gly (**A**), Ala (**B**), Lys (**C**), Leu (**D**), and Val (**E**) are presented.

enthalpy change is very close to the value of 255 cal/K · mol obtained from the individual calorimetric traces. Similar values for the enthalpy and heat capacity changes were obtained in previous studies.[17,18]

That the four peptides show the same enthalpy and heat capacity changes upon unfolding indicates that the observed differences in stability are due to differences in conformational entropy, since solvent-related entropy changes will be reflected in different ΔC_p values.[3,15] The observed equivalences in enthalpies and heat capacities also indicate that the four mutants exhibit the same degree of intramolecular interactions and the same changes in solvation. Changes in the packing of side chains at the dimer interface, for example, would have resulted in a measurable variation in the enthalpy change of the different peptides at any given temperature. The observed behavior is the one expected for Ala→Gly mutations at locations that are fully exposed to the solvent. Since Ala and Gly do not have side chains that can contribute to the conformational entropy change, the conformational entropy differences we measured can be attributed mainly to differences in the backbone entropies in the unfolded state.

Evaluation of the Entropy Change

At any temperature and peptide concentration, $[P_T]$, the population of molecules in the monomeric unfolded state, P_U, can be expressed by the equation[19,20]:

$$P_U = K \cdot [(K^2 + 4)^{1/2} - K)]/2 \quad (2)$$

where

$$K = \exp(-\Delta G°/RT)/(2 \cdot [P_T])^{1/2}$$

and $\Delta G°$ is the free energy of unfolding expressed on a per mole of monomer basis. Rearrangement and evaluation of the above equation at the temperature at which the transition is half completed, $T_{m,1/2}$, yields the following expression for the entropy change:

$$\Delta S(T_{m,1/2}) = \Delta H(T_{m,1/2})/T_{m,1/2} + (R/2) \cdot ln\,([P_T]) \quad (3)$$

where the enthalpy $[\Delta H(T_{m,1/2})]$, entropy $[\Delta S(T_{m,1/2})]$, and protein concentration $([P_T])$ are expressed on a per mole of monomer basis. In this case, the value of $K° = \exp(-\Delta G°/RT)$ when the transition is half completed is equal to $[P_T]^{1/2}$ rather than unity as in the case of a monomeric system. Table I summarizes the experimental thermodynamic data obtained calorimetrically for the four peptides.

According to the calorimetric results, at equivalent temperatures the difference in ΔS between the S14A and the S14G mutants, $\Delta\Delta S_{14}$, is -2.65 cal/K · mol and that between the wild type and A24G, $\Delta\Delta S_{24}$ is -2.27 cal/K · mol. One way of looking at these $\Delta\Delta S$ values is to consider that they reflect the decrease in conformational degrees of freedom of the peptide backbone in the Ala mutants with respect to the Gly mutants.

$$\Delta\Delta S = \Delta S_{Ala} - \Delta S_{Gly} = R \cdot ln\,(\Omega_{Ala}/\Omega_{Gly}) \quad (4)$$

where Ω_{Ala} and Ω_{Gly} are the number of allowed conformations of the Ala and Gly mutant backbones, respectively. The measured decrease in backbone entropy, which averages 2.46 ± 0.2 cal/K · mol, corresponds to a reduction in the number of allowed conformations by a factor of 3.4 and can be attributed to the conformational restrictions imposed by the presence of the methyl group in Ala.

Structural Analysis

The positions of the two mutation sites on the structure of the GCN4 peptide are shown in Figure 3. Ser-14 is situated near the center of the helix, while Ala-24 is located about one-third the length of the helix from the carboxy terminus. Evaluation of the changes in solvent accessibility of the Ala and Gly residues in the folded and unfolded conformations reveals a difference not greater than 28 Å^2 in ΔASA_{ap} and 32 Å^2 in ΔASA_{pol}. These results predict a difference in ΔC_p between the four peptides not greater than 4 cal/K · mol,[3,16] which is below the limit of experimental detection, in agreement with the experimental observation. Thus, the thermodynamic analysis of the structure is consistent with the experimental results and indicates that differences in solvent accessibility do not contribute measurably to the stability differences between the two pair of mutants, S14A and S14G, and wild type and A24G.

Computational Calculation of Backbone Entropy

Recently, a general method aimed at calculating the energy profile of amino acid side chains was presented.[9] The energy profile as a function of the dihedral angles evaluated for locally relaxed structures permits a direct calculation of the probability distribution of different conformers and therefore of the conformational entropy. In this paper, we have extended this method to the computation of the energy profiles of backbone conformations. The final energy maps as a function of the main chain dihedral angles ϕ and φ for Gly, Ala, Val, Leu, and Lys are presented in Figure 4. The results of the calculations, when expressed as the differences in conformational entropy between Gly and the other amino acids, show that the major change in entropy (-1.96 cal/K · mol) occurs between Gly and Ala, due to the steric restrictions imposed by the presence of the β-carbon. Additional restrictions are introduced by extension of the side chains beyond the β-carbon. Side chains containing one γ-carbon and beyond induce additional but smaller restrictions in the number of allowed conformations. The two side chains tested in this category (Leu and Lys) have similar effects (-2.57 and -2.77 cal/K · mol, respectively). As expected, β-branching (Val) produces the largest decrease in conformational entropy with respect to Gly (-3.88 cal/K · mol).

When an energy similar to kT (0.6 kcal/mol) is used as the cut-off between the allowed and not allowed conformations, the following values are obtained for the fractions of allowed conformations: Gly, 53.8%; Ala, 16.7%; Val, 2%; Leu 13.5%; and Lys, 14.3%. These values are very close to the values 52%, 16%, 4.5%, 11%, and 14% (calculated for norleucine) reported by Leach et al.[21] in 1966 for the same residues. As an alternative method to the one described above, differences in entropy between the unfolded state of a glycine-containing dipeptide and one containing another residue can be calculated based on the fractions allowed, using the following expression:

$$\Delta S_{res \rightarrow Gly} = R\ ln\ F_{res}/F_{Gly} \quad (5)$$

where F_X is the fraction of allowed main chain conformations for residue X. If these values are used to estimate the main chain conformational entropy as suggested by Leach et al.,[21] the difference between Gly and Ala is 2.32 cal/K · mol.

The loss of conformational entropy for a given residue when a peptide goes from unfolded to α-helix can be calculated as:

$$\Delta S_{res,unfolded \rightarrow helix} = S_{conf,helix} - S_{conf,unfolded} \quad (6)$$

The values of $S_{conf,helix}$ and $S_{conf,unfolded}$ were calculated as described in Materials and Methods. For Gly, the value of $S_{conf,unfolded}$ is 13.19 cal/K · mol and the value for $S_{conf,helix}$ is 6.66, giving a value of $\Delta S_{Gly,unfolded \rightarrow helix}$ of −6.51 cal/K · mol. For Ala the values are 11.23 cal/K · mol and 6.68 cal/K · mol, giving a value of $\Delta S_{Ala,unfolded \rightarrow helix}$ of −4.57 cal/K · mol.

It must be noted that the conformational entropies of Ala and Gly residues in an α-helix are extremely similar (6.68 and 6.66 cal/K · mol). These results validate the standard procedure in which the entropic contributions to the helical propensities of these two amino acids are calculated as the differences in the conformational entropies of their unfolded states. Also, the agreement between the computational and experimental values indicates that the computational procedure accurately accounts for the conformational entropy change in solution, which is the quantity of interest.

Effect of Excluded Volume

The above computational analysis was performed for a dipeptide, and as such the results only include the excluded volume effects due to nearest neighbors. As the number of residues (n) in the peptide chain increases, an increasing number of conformations will not be allowed due to excluded volume effects. As the chain length increases the excluded volume effect per residue reaches an asymptotic value. These effects have been studied for many years.[22] The number of conformations, Ω, available to each residue is given by the equation:

$$\Omega = \Omega_{Total} \cdot F_{acc}(n) \quad (7)$$

where Ω_{Total} is the total number of conformations and $F_{acc}(n)$ is the fraction of allowed conformations for a chain of n residues. As a first approximation, one can use the $F_{acc}(n)$ values calculated for a cubic lattice (see Materials and Methods) as reflecting the equivalent fraction in any three-dimensional self-avoiding chain. In the case of a polypeptide chain, when the backbone entropy is calculated as the sum of the entropies of the individual amino acids, the correction for excluded volume restrictions is given by:

$$S_{corr} = (R/n) \cdot ln\ F_{acc}(n). \quad (8)$$

S_{corr} is a slowly varying function of n (Fig. 5) starting at −0.09 for n = 10 and reaching an asymptotic value of −0.128 for n = 440. The value of S_{corr} can be calculated for each particular case, but for typical globular proteins it amounts to −0.121 cal/K · mol per residue. Even though the correction per residue

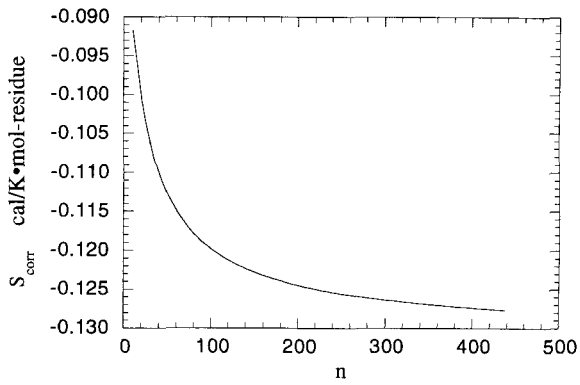


Fig. 5. The dependence on the number of residues of the entropy correction due to excluded volume interactions. As the length of the peptide chain increases, a number of conformations will not be accessible due to the presence of other residues. The resulting entropy correction reaches an asymptotic value as the chain length increases. The range shown in the figure covers the one found in globular proteins.

is small, for a protein of 100 amino acids the total correction is close to −12 cal/K · mol or −3.6 kcal/mol at 25°C.

Backbone Entropy and Polar Hydration

The computational results presented here indicate that the backbone entropy change for the unfolding of a Gly peptide is around 6.5 cal/K · mol of residue or about 6.38 cal/K · mol − res for a peptide of 100 residues. This value sets the entire scale of conformational entropies since all other values are calculated in relation to Gly (Table II). The values presented in Table II allow us to set limits to the magnitude of the entropy of polar hydration, another elusive quantity. It has been known for several years that at 112°C the entropy of apolar hydration is very close to zero.[23] Therefore, at this temperature the remaining entropy, after correction for protonation and other specific effects, is primarily due to the added contributions of the conformational entropy change and the entropy of polar hydration. If the values summarized in Table II are used to calculate the conformational entropy change, the entropy of polar hydration is estimated to be on the order of −0.04 cal/K · mol Å^2 of polar surface at 112°C. This result provides a small correction to our previous assumption[3,9] that the entropy of polar hydration was zero at 112°C. That assumption was necessary at the time because accurate backbone conformational entropy values were not available.

Recently, some authors have arrived at rather large conformational entropy values for protein unfolding (9–20 cal/K · mol).[2,4,24] Invariably, however, those high estimates are a consequence of equally large but negative values for the entropy of polar hydration. In the approach of Makhatadze and Privalov,[4,24] the conformational entropy is estimated by subtracting the calculated hydration entropy from the experimental entropy. Accordingly, a

TABLE II. Conformational Entropies for Amino Acids*

Amino acid	$\Delta S_{bu\rightarrow ex}$ (cal/K · mol)	$\Delta S_{ex\rightarrow u}$ (cal/K · mol)	ΔS_{bb} (cal/K · mol)
ALA	0.00	0.00	4.1
ARG	7.11	−0.84	3.4
ASN	3.29	2.24	3.4
ASP	2.00	2.16	3.4
CYS	3.55	0.61	3.4
GLN	5.02	2.12	3.4
GLU	3.53	2.27	3.4
GLY	0.00	0.00	6.5
HIS	3.44	0.79	3.4
ILE	1.74	0.67	2.18
LEU	1.63	0.25	3.4
LYS	5.86	1.02	3.4
MET	4.55	0.58	3.4
PHE	1.40	2.89	3.4
SER	3.68	0.55	3.4
THR	3.31	0.48	3.4
TRP	2.74	1.15	3.4
TYR	2.78	3.12	3.4
VAL	0.12	1.29	2.18

*$\Delta S_{bu\rightarrow ex}$ and $\Delta S_{ex\rightarrow u}$ are from Lee et al.[9] For the purpose of approximating ΔS_{bb}, the amino acids have been divided into four groups: (1) no β-carbon (Gly); (2) β-carbon only (Ala); (3) β-branching (Val and Ile); and (4) γ-carbon (remaining amino acids).

large negative polar hydration entropy is immediately reflected in a correspondingly large and positive conformational entropy. In those papers the entropy of polar hydration is calculated from thermodynamic data obtained from the transfer of small molecules from the gas phase to water. The protein values are then calculated by assuming group additivity and that the polar hydration entropy scales up in terms of accessible surface areas.

As a representative example of the differences between different approaches, the magnitude of the various contributions to the unfolding entropy of hen egg white lysozyme have been plotted in Figure 6A as a function of temperature. It is clear that the conformational entropy values reported here (Table II) in conjunction with our previously published parametrization[3,16,25,26] accurately predict the experimental entropy change. Similar results are obtained for other proteins. The "revised" values of Makhatadze and Privalov[24] are very close to our apolar entropy values, suggesting that for apolar groups the assumption of group additivity is valid, i.e., the same parameter values are obtained from small molecules or from the analysis of the protein thermodynamic database. This type of comparison can be used as an empirical test of the validity of the additivity assumption. The situation is clearly different for polar hydration. In this case, the results obtained from small molecules are different from those obtained from the analysis of the protein thermodynamic database.

It is also apparent from Figure 6A that the conformational entropy and the entropy of polar hydration are correlated (see the original and revised values of Makhatadze and Privalov[4,24]). A large negative polar hydration coupled to a large conformational entropy accounts equally well for the protein data and vice versa. In fact, an analysis of the available data for protein unfolding indicates that the backbone entropy for Gly [ΔS_{bb}(Gly)] and the entropy of polar hydration per mole of Å^2 evaluated at 112°C [ΔS_{pol}(112)] are linked by the equation:

$$\Delta S_{pol}(112) = 0.13 - 0.027 \cdot \Delta S_{bb}(\text{Gly}). \quad (9)$$

Any combination of values that satisfies the above equation will quantitatively account for protein unfolding data. In the case of Privalov and Makhatadze[4] the conformational entropy and polar hydration entropy at 112°C are on the order of 20 cal/K · mol and −0.4 cal/K · mol-Å^2, respectively, which satisfies the above equation. In the case of Ooi and Oobatake,[2] the conformational entropy and polar hydration entropy at 112°C are on the order of 9 cal/K · mol and −0.1 cal/K · mol-Å^2 respectively, which also satisfies the above equation. In our own previous work,[27] ΔS_{bb}(Gly) was set equal to 4.8 and $\Delta S_{pol}(112) = 0$, which also satisfies Equation (9).

The above discussion indicates that protein folding/unfolding data alone cannot be used to separate the polar hydration entropy from the backbone conformational entropy. Also, from a structural point of view, most of the polar hydration is due to the peptide backbone. The two terms can be uncoupled, however, if data for the binding of unstructured peptides to proteins, and data for protein-protein association reactions are used together with folding/unfolding data. As mentioned before, protein folding/unfolding for monomeric proteins involves three conformational entropy terms: $\Delta S_{bu\rightarrow ex}$, $\Delta S_{ex\rightarrow u}$, and ΔS_{bb}. The binding of an unstructured peptide to a protein involves the ΔS_{bb} term only for the peptide since the protein backbone is already in a fixed conformation, and the binding of two already folded proteins involves no ΔS_{bb} term. Therefore, data for these different situations can be used to uncouple ΔS_{bb} and $\Delta S_{pol}(112)$.

Figure 6B illustrates the situation by plotting the dependence of the Gibbs free energy on different combinations of ΔS_{bb} and $\Delta S_{pol}(112)$ for the stability of hen egg white lysozyme, the binding of the peptide pepstatin A to endothiapepsin,[27] and the binding of hen egg white lysozyme to the D1.3 antibody.[28] In all cases Equation (9) was used to calculate the polar hydration entropy that corresponds to each ΔS_{bb} value. The data for the stability of hen egg white lysozyme is representative of the situation encountered with other folding/unfolding data; the data for pepstatin A represents the binding of an unstructured peptide to a protein; and the data for the D1.3 antibody represents the situation for the

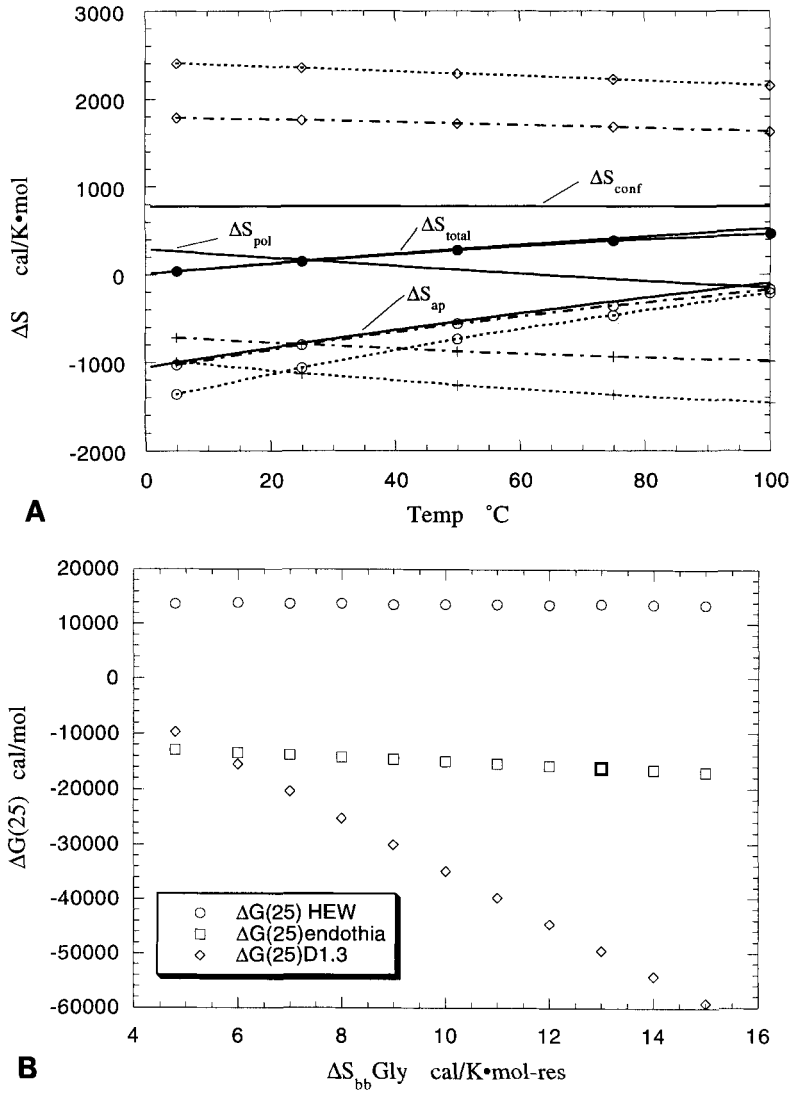


Fig. 6. **A:** Different contributions to the unfolding entropy of hen egg white (HEW) lysozyme. The experimental entropy values are represented by solid circles. The solid lines represent the values calculated with the parametrization developed here. The entropy of apolar hydration was calculated with the formula $\Delta S_{ap} = 0.45 \cdot \Delta ASA_{ap} \cdot ln(T/385.15)$; the entropy of polar hydration was calculated with the formula $\Delta S_{pol} = -0.26 \cdot \Delta ASA_{pol} \cdot ln(T/335.15)$; the conformational entropy (ΔS_{conf}) was calculated with the parameters in Table II; and the total entropy as $\Delta S_{total} = \Delta S_{conf} + \Delta S_{ap} + \Delta S_{pol}$. The dotted lines represent the original values published by Privalov and Makhatadze[4] and the dashed lines the revised values published by the same authors[24] ($\diamond$ ΔS_{conf}; $\circ$ ΔS_{ap}; $+$ ΔS_{pol}). **B:** The backbone entropy and the entropy of polar hydration are correlated for protein unfolding. As indicated for HEW, the free energy of stabilization is independent of the backbone entropy (expressed here in terms of $\Delta S_{bb,Gly}$) when the polar hydration entropy is calculated by Equation 9. The situation is different for association reactions, especially those between folded proteins in which the backbones already exist in a unique conformation before binding. This figure also shows the binding Gibbs energy for the association of the peptide pepstatin A to endothiapepsin[27] and that of HEW to the specific antibody D1.3[28] calculated using the polar hydration entropy that corresponds to the indicated backbone entropy, as dictated by equation 9.

association of two already folded proteins in which no backbone entropy changes take place. As shown in the figure, the folding/unfolding data for hen egg white lysozyme is independent of ΔS_{bb} because this term is completely correlated with the polar hydration term. The situation is different for those cases in which no backbone entropy change is involved. In these cases unreasonably high ΔG values are obtained for $\Delta S_{pol}(112)$ values more negative than -0.05 cal/K · mol-Å^2, which correspond to ΔS_{bb} values larger than 6.5 cal/K · mol. These correlations set an upper limit to ΔS_{bb} that is very close to the value obtained computationally. With the precision available now, ΔS_{bb} values ranging between 4.8 and 6.5 are consistent with the experimental data provided that the corresponding polar hydration entropy satisfies Equation (9).

The results presented here indicate that the average entropy of polar hydration when normalized per unit area is significantly smaller for proteins than for small organic molecules. In fact, for proteins the entropy of polar hydration appears to be very small

at room temperature and increasingly positive at lower temperatures. In this respect, Spolar and Record[29] have obtained reasonable results assuming that the entropy of polar hydration is zero. These results argue against the validity of the additivity assumption for polar hydration. In fact, deviations from additivity are already observed for small molecules that contain two or more polar groups.[30]

The anomalously large conformational entropy changes resulting from the analysis of Makhatadze and Privalov[4,24] have been noticed before by other groups.[7,31,32] There are several reasons for this overestimation: experimental errors in their measured thermodynamic parameters, as noted by several authors,[31,33,34] and more fundamental problems arising from the assumption of additivity for the polar hydration entropy and with the validity of the assumption that the transfer of small organic molecules from the gas phase to water actually represents the hydration process associated with protein unfolding. Similar problems exist in their analysis of the enthalpy change of polar hydration, as discussed previously by Lazaridis et al.[35]

DISCUSSION

The experimental results presented in this paper indicate that the presence of the methyl group in alanine reduces the conformational entropy of the peptide backbone by 2.46 ± 0.2 cal/K · mol with respect to that of glycine, consistent with a 3.4-fold reduction in the number of allowed conformations in the Ala-containing peptide. This value is close to the value of 1.96 cal/K · mol obtained from the probability distribution of backbone conformations and the value 2.32 cal/K · mol obtained from the fraction of allowed conformations with an energy smaller than kT. The agreement between the calculated results and the direct thermodynamic measurements provides support to the correctness of the assumptions involved in the computational algorithms and validates the use of this approach in other situations. In particular, the computational approach reproduces well the conformational entropy change in solution, which is the important quantity for understanding the stability of proteins.

About 30 years ago Leach et al.[21] presented a computation of the effects of side chains on the sterically allowed conformation of peptides. Their method used a "step function" to decide if a given conformation was or not allowed: if any two atoms separated by more than three bonds were closer than the sum of their van der Waals radii, the conformation was counted as disallowed. For glycine these authors found that the allowed conformations covered about 52% of the total area of the steric map. The presence of a β-carbon (Gly–Ala) restricted the number of allowed conformations to about 16% of the total area, i.e., by a factor of 3.25, consistent with a conformational entropy difference of 2.34 cal/K · mol, which is very close to the values reported here. It is a tribute to the intuition of the authors that such a crude method (necessary at the time because of computing power) gave such a good result.

The computational analysis presented here indicates that the addition of further C atoms to the side chain has only a small effect as long as the side chains are unbranched at position β. Further reductions with respect to Ala of only 0.61 and 0.81 cal/K · mol in the backbone entropy were obtained for leucine and lysine, respectively. For β-branching, as in valine, the entropy was found to be 1.92 cal/K · mol less than Ala. Branching at further distances has a smaller effect. These conclusions are also similar to those obtained by Leach et al.[21] Interestingly, the distribution of the main chain dihedral angles ϕ and φ in crystallographic structures yields a value for Ala close to the one found here[5]; however, for other residues this approach grossly overestimates the backbone conformational entropy.

As far back as 1955, Schellman[36] estimated that the loss of conformational entropy of the peptide backbone itself (i.e., glycine) upon α-helix formation must lie anywhere between 3.0 and 7.18 cal/K · mol. Nemethy and Scheraga[37] calculated that for a Gly residue that is part of a dipeptide 21 conformations are allowed, which is equivalent to a conformational entropy of 6.05 cal/K · mol, very close to 6.51 cal/K · mol, the value calculated in this paper for $\Delta S_{Gly,helix \rightarrow unfolded}$. If the allowed conformations of each residue in the peptide chain were independent of the other residues, the total number of allowed conformations will be the product of the allowed conformations for each residue and the backbone entropy would be perfectly additive. In actuality, there are chain length-dependent excluded volume effects that will restrict the number of total conformations. These additional effects have been estimated and are expected to be on the order of −0.121 cal/K · mol-residue for a typical globular protein.

Previously, we and others[6,8,9] have estimated the conformational entropies, $\Delta S_{ex \rightarrow u}$, associated with the unfolding of solvent-exposed side chains. These studies have indicated that $\Delta S_{ex \rightarrow u}$ is small in magnitude. Lee et al.[9] estimated $\Delta S_{ex \rightarrow u}$ for all 20 amino acids by computing the entire energy profile of each side chain. $\Delta S_{ex \rightarrow u}$ ranges between −0.84 and 2.27 cal/K · mol, averaging 0.98 cal/K · mol or 0.82 cal/K · mol if the frequency of each amino acid in globular proteins[38] is used to calculate a weighted average. Creamer and Rose[6] and Blaber et al.[8] obtained similar results. Lee et al.[9] also estimated $\Delta S_{bu \rightarrow ex}$, the conformational entropy associated with the transfer of a side chain that is buried in the interior of the protein to its surface. $\Delta S_{bu \rightarrow ex}$ averages 3.08 cal/K · mol or 2.62 cal/K · mol for the weighted average. Together with the ΔS_{bb} values presented in this paper, these results indicate that the overall conformational entropy change for the unfolding of a globular

protein is on the order of 6 cal/K · mol-res. This value is obtained for the unfolding of globular proteins for which high-resolution structures are available, taking into consideration their amino acid sequence and the degree of exposure to the solvent of each amino acid. This value is about 1.2 cal/K · mol-res larger than estimates made previously under the assumption that the polar entropy of hydration is zero at 112°C.[9] The entropy of polar hydration at 112°C consistent with the conformational entropy values in Table II is on the order of −0.04 cal/K · mol-Å^2. This value is significantly smaller than the value estimated from the polar hydration entropies of small molecules assuming group additivity, a finding that calls into question the validity of this assumption for polar interactions. Table II summarizes our current estimates for the conformational entropies of 19 amino acids. These values can be used to perform structure-based thermodynamic calculations of protein stability and binding affinities.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (RR04328 to E.F. and GM44692 to L.M.A.) and the National Science Foundation (MCB-9118687 to E.F.).

REFERENCES

1. Privalov, P.L., Khechinashvili, N.N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. J. Mol. Biol. 86:665–684, 1974.
2. Oobatake, M., Ooi, T. Hydration and heat stability effects on protein unfolding. Prog. Biophys. Mol. Biol. 59:237–284, 1992.
3. Murphy, K.P., Bhakuni, V., Xie, D., Freire, E. The molecular basis of co-operativity in protein folding III: Identification of cooperative folding units and folding intermediates. J. Mol. Biol. 227:293–306, 1992.
4. Privalov, P.L., Makhatadze, G.I. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. J. Mol. Biol. 232:660–679, 1993.
5. Stites, W.E., Pranata, J. Empirical evaluation of the influence of side chains on the conformational entropy of the polypeptide backbone. Proteins 22:132–140, 1995.
6. Creamer, T.P., Rose, G.D. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. Proc. Natl. Acad. Sci. USA 89: 5937–5941, 1992.
7. Pickett, S.D., Sternberg, M.J.E. Empirical scale of side-chain conformational entropy in protein folding. J. Mol. Biol. 231:825–839, 1993.
8. Blaber, M., Zhang, X.J., Lindstrom, J.L., Pepsot, S.D., Baase, W.A., and Matthews, B.W. Determination of alpha-helix propensity within the context of a folded protein: Sites 44 and 131 in bacteriophage T4 lysozyme. J. Mol. Biol. 235:600–624, 1994.
9. Lee, K.H., Xie, D., Freire, E., Amzel, L.M. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. Proteins 20:68–84, 1994.
10. Freire, E., Haynie, D.T., Xie, D. Molecular basis of cooperativity in protein folding. IV. CORE: A general cooperative folding model. Proteins 17:111–123, 1993.
11. Murphy, K.P., Xie, D., Thompson, K., Amzel, L.M., Freire, E. Entropy in biological binding processes: Estimation of translational entropy loss. Proteins 18:63–67, 1994.
12. O'Shea, E.K., Klemm, J.D., Kim, P.S., Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science 254:539–544, 1991.
13. O'Shea, E.K., Rutkowski, R., Kim, P.S. Evidence that the leucine zipper is a coiled coil. Science 243:538–542, 1989.
14. Oas, T.G., McIntosh, L.P., O'Shea, E.K., Dahlquist, F.W., Kim, P.S. Secondary structure of a leucine zipper determined by nuclear magnetic resonance spectroscopy. Biochemistry 29:2891–2894, 1990.
15. Spolar, R.S., Livingstone, J.R., Record, M.T. Jr. Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. Biochemistry 31:3947–3955, 1992.
16. Gomez, J., Hilser, J.V., Xie, D., Freire, E. The heat capacity of proteins. Proteins 22:404–412, 1995.
17. Kenar, K.T. "Thermodynamic Characterization of the Structural Stability of the GCN4 Leucine Zipper Dimer." Baltimore: The Johns Hopkins University Press, 1994.
18. Kenar, K.T., García-Moreno, B., Freire, E. A calorimetric characterization of the salt dependence of the stability of the GCN4 leucine zipper. Protein Sci. 4:1934–1938, 1995.
19. Freire, E. Statistical thermodynamic analysis of the heat capacity function associated with protein folding-unfolding transitions. Comments Mol. Cell. Biophys. 6:123–140, 1989.
20. Thompson, K., Vinson, C., Freire, E. Thermodynamic characterization of the structural stability of the coiled-coil region of the bZIP transcription factor GCN4. Biochemistry 32:5491–5496, 1993.
21. Leach, S.J., Nemethy, G., Scheraga, H.A. Computation of the sterically allowed conformations of peptides. Biopolymers 4:369–407, 1966.
22. Fisher, M.E. Effect of excluded volume on phase transitions in biopolymers. J. Chem. Phys. 45:1469–1473, 1966.
23. Baldwin, R.L. Temperature dependence of the hydrophobic interaction in protein folding. Proc. Natl. Acad. Sci. USA 83:8069–8072, 1986.
24. Makhatadze, G.I., Privalov, P.L. Hydration effects in protein unfolding. Biophys. Chem. 51:291–309, 1994.
25. Xie, D., Freire, E. Structure based prediction of protein folding intermediates. J. Mol. Biol. 242:62–80, 1994.
26. Xie, D., Freire, E. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins 19:291–301, 1994.
27. Gomez, J., Freire, E. Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. J. Mol. Biol. 252:337–350, 1995.
28. Schwarz, F.P., Tello, D., Goldbaum, F.A., Mariuzza, R.A., Poljak, R.J. Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies. Eur. J. Biochem. 228:388–394, 1995.
29. Spolar, R.S., Record, M.T. Coupling of local folding to site specific binding of proteins to DNA. Science 263:777–784, 1994.
30. Cabani, S., Gianni, P., Mollica, V., Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. J. Solution Chem. 10:563–595, 1981.
31. Murphy, K.P. Hydration and convergence temperatures: on the use and interpretation of correlation plots. Biophys. Chem. 51:311–326, 1994.
32. Khechinashvili, N.N., Janin, J., Rodier, F. Thermodynamics of the temperature-induced unfolding of globular proteins. Protein Sci. 4:1315–1324, 1995.
33. Vogl, T., Hinz, H.-J., Hedwig, G.R. Partial molar heat capacities and volumes of Gly-X-Gly tripeptides in aqueous solution: Model studies for the rationalization of thermodynamic parameters of proteins. Biophys. Chem. 54:261–269, 1995.
34. Habermann, S.M., Murphy, K.P. Energetics of hydrogen bonding in proteins: A model compound study. Nature Struct. Biol. 1995 (submitted).
35. Lazaridis, T., Archontis, G., and Karplus, M. Enthalpic contributions to protein stability: Atom-based calculations and statistical mechanics. Adv. Prot. Chem. 47:231–306, 1995.
36. Schellman, J.A. The stability of hydrogen bonded structures in aqueous solution. CR Trav. Carlsburg, Ser. Chim. 29:230–259, 1955.

37. Nemethy, G., Scheraga, H.A. Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. Biopolymers 3:155–184, 1965.
38. Klapper, M.H. The independent distribution of amino acid near neighbor pairs into polypeptides. Biochem. Biophys. Res. Commun. 78:1018–1024, 1977.
39. Kent, S.B.H. Chemical synthesis of peptides and proteins. Annu. Rev. Biochem. 57:957–989, 1988.
40. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55: 379–400, 1971.
41. Chan, H.S., Dill, K.A. Polymer principles in protein structure and stability. Annu. Rev. Biophys. Biophys. Chem. 20:447–490, 1991.
42. Jorgensen, W.L. and Tirado-Rives, J. The OPLS potential function for proteins. Energy minimization for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110:1657–1666, 1988.
43. Brunger, A.T. "X-PLOR, version 3-1A. A system for crystallography and NMR," New Haven, Yale University Press 1992.