Analysis of Thermodynamic Determinants in Helix Propensities of Nonpolar Amino Acids through a Novel Free Energy Calculation

Jing Wang and Enrico O. Purisima*

Contribution from the National Research Council of Canada, Biotechnology Research Institute, 6100 Royalmount Ave., Montreal, Quebec, Canada H4P 2R2, and Montreal Joint Centre for Structural Biology, Montreal, Quebec, Canada H4P 2R2

Received October 2, 1995[⊗]

Abstract: The relative helix propensities of Gly, Ala, Val, Ile, and Leu in the center of a polyalanine helix were calculated using a novel free energy simulation method (Wang *et al*. *J. Mol. Biol.* **1995**, *253*, 473) that permits the decomposition of the free energy into its various thermodynamic components. The calculated relative free energy changes agree well with the recent set of experimental data of Chakrabartty *et al*. (*Protein Science* **1994**, *3*, 843) on alanine-based peptides. The side chain rotamer distributions in the α-helix produced are also consistent with the reports in the literature based on a statistical survey of crystal structures of proteins. A detailed decomposition of the free energy showed that the solvation effect, or hydrophobicity in particular, has little contribution to the helix propensities of the amino acids relative to Gly. The side chain−helical matrix van der Waals interactions are generally favorable and account for a large part of the free energy change relative to Gly upon helix folding. The configurational entropy plays a significant, but not dominant, role in the relative free energy changes. The absolute change of configurational entropy of a central amino acid in folding, which is usually difficult to assess, was also obtained. The entropic cost of restricting the backbone of an amino acid in a helical matrix is about 1.5 kcal/mol at 25 °C, significantly larger than the cost associated with the reduction in side chain entropy in the helix.

Introduction

The α-helix is a major structural element in protein architecture. Understanding the mechanisms of the formation and stability of this structural motif is one of the first steps for clarifying the more complicated protein folding and stability problem. Recently, a number of factors have been put forward as providing the energetic and structural basis for the helix propensities of amino acids. Among the most commonly cited are hydrophobicity, configurational entropy, hydrogen bonding, and electrostatic interactions.[1,2] Creamer and Rose,[3,4] Blaber *et al*.,[5] and Lee *et al*.,[6] quantitatively evaluated the side chain conformational and configurational entropy. Blaber *et al.*[5] semiquantitatively discussed the effect of hydrophobicity. The objective of this paper is the further clarification of the roles of entropy, solvation, hydrophobicity, and van der Waals interactions with respect to the helix propensities of nonpolar amino acids.

We have calculated and analyzed the relative free energies of helix formation for a series of amino acids in the center of a polypeptide helix represented by Ace-Ala_4-Xaa-Ala_4-NMe. The coil state was modeled by the sequence Ace-Ala-Xaa-Ala-NMe. In this study, Xaa was Gly, Ala, Val, Leu, or Ile. We employed the free energy simulation method of Wang *et al*.[7] which permits us to dissect out the entropic and enthalpic components of the free energy. Solvation effects were incorporated using a continuum dielectric boundary element model.[8] For this solvation model, a formalism was developed that partitions the free energy into the solute configurational entropy and an effective solvated conformational energy.

Hermans *et al*.[9] performed free energy simulations on glycine, alanine, valine, proline and four unnatural amino acids. Komeiji *et al*.[10] performed similar simulations on alanine and valine. Both groups used the free energy perturbation method which is a totally different approach from our method. Although they obtained the free energy changes, the other thermodynamic quantities such as the entropic and enthalpic contributions have not been reported.

Theory and Method

The configurational free energy of a system is

$$A = -kT \ln Z \tag{1}$$

where Z is the partition function given by

$$Z = \int_{\Omega} \exp(E(\Omega)/kT)\, d\Omega \tag{2}$$

$E(\Omega)$ is the potential energy and the integration is over all relevant regions of configurational space. For calculations of $\Delta\Delta A$ we make

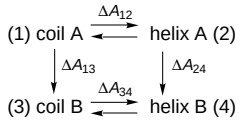
* Address correspondence to this author at the National Research Council of Canada.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1996.

(1) Horovitz, A.; Matthews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560−568.
(2) Blaber, M.; Zhang, X.; Matthews, B. W. *Science* **1993**, *260*, 1637−1640.
(3) Creamer, T. P.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5937−5941.
(4) Creamer, T. P.; Rose, G. D. *Proteins* **1994**, *19*, 85−97.
(5) Blaber, M.; Zhang, X.; Lindstrom, J. D.; Pepiot, S. D.; Baase, W. A.; Matthews, B. W. *J. Mol. Biol.* **1994**, *235*, 600−624.
(6) Lee, K. H.; Xie, D.; Freire, E.; Amzel, L. M. *Proteins* **1994**, *20*, 68−84.
(7) Wang, J.; Szewczuk, Z.; Yue, S.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. *J. Mol. Biol.* **1995**, *253*, 473−492.
(8) Purisima, E. O.; Nilar, S. H. *J. Comp. Chem.* **1995**, *16*, 681−689.
(9) Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry* **1992**, *31*, 5646−5653.
(10) Komeiji, Y.; Honda, N.; Yamato, I. *Biophys. Chem.* **1993**, *47*, 113−121.
(11) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230−252.

0002-7863/96/1518-0995$12.00/0 Published 1996 by the American Chemical Society

use of the thermodynamic cycle



(3)

The desired relative free energy differences can be calculated from the ratios of the appropriate partition functions.

$$\Delta\Delta A = \Delta A_{34} - \Delta A_{12} = \Delta A_{24} - \Delta A_{13} = -kT \ln(Z_1 Z_4 / Z_2 Z_3) \quad (4)$$

Model System. The helical state is represented by the sequence Ace-Ala$_4$-Xaa-Ala$_4$-NMe where Xaa is the mutated residue flanked by four alanines. Ace and NMe are acetyl and *N*-methyl blocking groups, respectively. The backbone dihedral angles are initially set to α-helical values using the SYBYL program (Tripos Assoc., Inc.). For coil state, we use the truncated sequence Ace-Ala-Xaa-Ala-NME. The assumption is that in the coil state, residues further away in the sequence behave independently of each other. This means that the contribution from the rest of the chain can be factored out of Z_1 and Z_3 and cancel out in the ratio taken in eq 4. The coil state is initially set to an extended conformation.

The method of estimating the partition functions is described in detail in Wang *et al.*[7] A brief summary of the main steps in the method is given below.

Systematic Search. A systematic search is carried out to obtain a database of sterically feasible conformations. The backbone and side chain rotatable bonds of Xaa are varied using 15° increments while keeping all other dihedral angles fixed at their initial values. The van der Waals radii of all atoms are scaled by a factor of 0.85 to take into account the coarseness of the increment. For the helical state, the systematic search of the backbone torsional angles (φ,ψ) is limited to the range of −120 to 0°.

Energy Minimization. The next step is to energy-minimize the conformers in the database. The minimization is done in two stages. First a minimization is performed in which the torsional angles of Xaa are restrained with a harmonic force constant of 0.2 kcal/(mol·deg^2) to be close to the values obtained in the systematic search. This minimization yields an energy for the relaxed starting conformation. The restraint is then released and the minimization continued to a gradient of 0.01 kcal/(mol·Å). The minimizations are done in Cartesian space with all atoms allowed to move. If the database is small enough, all conformers are minimized. Otherwise, a randomly selected subset of conformers is minimized and the results extrapolated as described in Wang *et al.*[7]

The partition function is then constructed from the collection of N_{tot} accepted final energies using the expression

$$Z = C \sum_{j=1}^{N_{tot}} \exp(-E_j/kT) \quad (5)$$

where N_{tot} is the total number of accepted conformers, E_j the final minimized energy of conformer j, and C a coefficient given by[7]

$$C = s^f \Gamma(f/2 + 1)(kT/E_\delta)^{f/2} \quad (6)$$

s is the step size of the systematic search (15°) and f the number of degrees of freedom in the integration (i.e., the number of rotatable bonds used in the systematic search). E_δ is a user-specified energy increment. Conformers for which the energy of the relaxed starting structure is greater than $E_j + E_\delta$ are excluded from the summation. We use a value of $E_\delta = 2$ kcal/mol in our calculations as in Wang *et al.*[7] It should be noted that the coefficient C cancels out in the calculation of free energy changes when ratios of partition functions are taken in eq 4.

Each energy minimum (potential well) contributes to the partition function by an amount proportional to $m_i \exp(-E_i/kT)$, where m_i is the number of conformers belonging to that minimum and whose energy for the relaxed starting conformation is within E_δ of the local minimum energy. In other words, m_i is a measure of the "width" of a potential well measured at a level E_δ from the bottom of the well.

The potential energy fucntion used in the minimizations was the AMBER force field[11] as implemented in SYBYL 6.0 (Tripos Assoc., Inc.). A dielectric constant $\epsilon = 80$ was used to mimic the electrostatic screening provided by the solvent. Calculations were also done using an implicit solvation model. This was accomplished by taking the same energy minimized structures from the *in vacuo* calculation, recalculating the intramolecular energies of the minimized structures with $\epsilon = 2$, and adding a solvation term to be described later.

First we describe the implications of this added solvation term for the proper calculation of the enthalpic and entropic components of the free energy.

Mean-Force Approximation of Solvent. Consider a simulation involving a solute molecule immersed in a box of explicit solvent molecules. The configurational partition function for such a system would be

$$Z = \int_\Omega \int_\eta \exp(-\beta(E_v + E_w))\, d\eta\, d\Omega \quad (7)$$

where $\beta = (kT)^{-1}$. The integrals are evaluated over the solute and solvent configurational space Ω and η, respectively. E_v is the intramolecular energy of a solute and is a function of Ω. E_w is made up of the solute−solvent and solvent−solvent interaction energies and is a function of both Ω and η. Rearranging eq 7 gives,

$$Z = \int_\Omega \exp(-\beta E_v(\Omega))\{\int_\eta \exp(-\beta E_w(\Omega,\eta))\, d\eta\}\, d\Omega \quad (8)$$

We then replace the integral over solvent phase space by a single Boltzmann factor using

$$\int_\eta \exp(-\beta E_w(\Omega,\eta))\, d\eta = \exp(-\beta\Phi(\Omega,T)) \quad (9)$$

where $\Phi(\Omega,T)$ is the solvation free energy for a solute molecule in a given conformation Ω and at a given temperature T. The partition function can then be rewritten as an integral over only the solute configurational space Ω.

$$Z = \int_\Omega \exp(-\beta(E_v + \Phi))\, d\Omega \quad (10)$$

It should be pointed out that unlike E_w which was a temperature-independent potential energy function, Φ is a temperature-dependent *free energy* function. This temperature dependence affects the expressions for the internal energy and entropy of the system.

The internal energy of the system is given by

$$U = -\frac{\partial \ln Z}{\partial \beta} \quad (11)$$

which when applied to eq 10 gives

$$U = \frac{1}{Z}\int_\Omega E \exp(-\beta E)\, d\Omega - \frac{1}{Z}\int_\Omega T\frac{\partial\Phi}{\partial T}\exp(-\beta E)\, d\Omega = \langle E\rangle - T\langle \partial\Phi/\partial T\rangle \quad (12)$$

where $E = E_v + \Phi$, and $\langle\ \rangle$ indicates an ensemble average. The appearance of the second term on the right hand side of eq 12 is due to the temperature dependence of Φ. Recognizing that the solvation entropy $S_w = -\partial\Phi/\partial T$, we see that the entropic component of Φ in the $\langle E\rangle$ is canceled out giving

$$U = \langle E + TS_w\rangle = \langle E_v + H_w\rangle \quad (13)$$

The internal energy is the ensemble average of the solute conformational energy plus the solvation enthalpy, H_w.

The total entropy can also be derived in a similar way using $S = -\partial A/\partial T$ giving

$$-TS = A - \langle E\rangle + T\langle\partial\Phi/\partial T\rangle = A - \langle E\rangle - T\langle S_w\rangle \quad (14)$$

We note that eqs 12 and 14 satisfy

$$A = U - \mathrm{TS} \tag{15}$$

An interpretation for $A - \langle E \rangle$ and $-T\langle S_w \rangle$ can be obtained by calculating the configurational entropy of the solute which is given by

$$S_v = -k \int_\Omega p \ln p \, d\Omega \tag{16}$$

where p is the normalized Boltzmann probability corresponding to $E(\Omega)$. One can derive from eq 16 that

$$-TS_v = A - \langle E \rangle \tag{17}$$

We see that the entropic component of the free energy expressed in eq 14 can be decomposed into two terms. The first term is the solute configuration entropic contribution $-TS_v$. The second term, $-T\langle S_w \rangle$, has the natural interpretation of being the solvation entropy component of the free energy. This is consistent with the fact that S_w is the "microscopic" solvation entropy corresponding to one conformation of the solute molecule. Thus we have two ways of decomposing the free energy. The first is given by the usual formula, eq 15. The second is given by

$$A = U^* - TS_v \tag{18}$$

where $U^* \equiv \langle E \rangle$.

Equation 18 is a useful decomposition in that the solute configurational entropy is obtained directly from the simulation using just the conformational energy and solvation free energy functions, E and Φ alone. $\langle E \rangle$ is then an effective mean solvated conformational energy. Such a decomposition is interesting in the analysis of folding or binding processes where the net tendency is the result of a balance of changes in configurational entropy and conformational energy. The solvation effects are subsumed into the effective conformational energy $\langle E \rangle$. Evaluation of the total entropy and internal energy using eqs 12 and 14, if desired, then requires an extra step and the introduction of another function, $S_w(\Omega)$.

The nature of the solvation free energy function is described next.

Solvation Energy Function. The solvation free energy is calculated through a continuum dielectric model and contains two terms,

$$\Phi = \Phi_{rf} + \Phi_{surf} \tag{19}$$

The first term is the electrostatic reaction field energy of the solute molecule in solvent, calculated using a boundary element method,[8] and the second term includes the cost of maintaining a cavity in the solvent medium as well as the other non-electrostatic interfacial interactions through the solute−solvent boundary.

In the calculation of the Φ_{rf} term, the dielectric constant of the solute inner region is taken to be 2 and that of the solvent region to be 78.5 at 25 °C and 88.28 at 0 °C which are the dielectric constants of bulk water. The same set of charges and radii of atoms are used in the conformational energy calculations and in the solvation free energy calculations.

The non-electrostatic term was taken to be proportional to the solvent-excluded surface area SA of solute molecule, that is,

$$\Phi_{surf} = \alpha \cdot \mathrm{SA} + \gamma \tag{20}$$

The temperature-dependent parameter a was determined from experimental solvation free energies of hydrocarbons[12,13] and found to be 0.0121 kcal/(mol·Å^2) at 25 °C and 0.00895 at 0 °C with the definition of SA and choice of radii used in our programs. The corresponding values for γ are 0.614 and 0.454. A unique constant α is sufficient for the different chemical species involved in these studies.[8]

The solvation free energy calculation program has been applied to a number of model compounds with different functional groups with good agreement with experimental solvation free energies.[8]

In order to calculate U and TS using eqs 12 and 14, we need to obtain an estimate for S_w, the solvation entropy. The electrostatic component of Φ given in eq 19 is essentially temperature independent. We can thus take $S_w = -\partial\Phi/\partial T = -\partial\Phi_{surf}/\partial T$. As with the solvation free energy, a linear relationship between the solvation entropy versus the solvent-excluded surface area can be determined from the solvation entropies of hydrocarbons.[13,14] We obtain $-TS_w = 0.0371 \cdot \mathrm{SA} + 3.108$ at 25 °C (units in kcal/mol with the area in Å^2).

Table 1. Number of Conformers in the Database Used for the Free Energy Calculation

	coil	helix
Gly	397	11
Ala	162	11
Val	1013	41
Ile[a]	10516	563
Leu[a]	19916	994

[a] For Ile and Leu, the free energies and components of the coil state were extrapolated using 3697 and 3001 randomly chosen entries in the database, respectively.

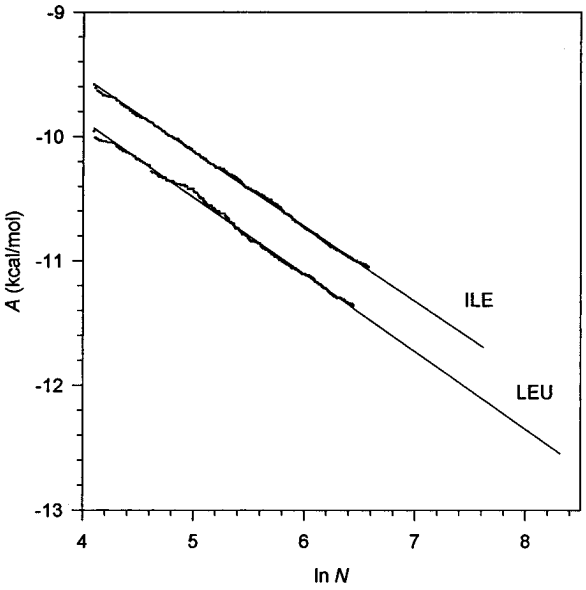


Figure 1. Logarithmic growth of the free energy A versus N, the number of conformers randomly sampled thus far, for the tetrapeptides Ace-Ala-Ile-Ala-NMe (upper curve) and Ace-Ala-Leu-Ala-NMe (lower curve). The straight lines are from least-squares fit and are extrapolated up to the expected total number of accepted conformers in the database.

Results

In presenting the results, we first describe the convergence and numerical properties of our calculation and then describe the structural and thermodynamic results in terms of the effects on α-helix stability of replacement of a central residue by different nonpolar amino acids.

Convergence of Free Energy and Its Components. The number of conformers which are sterically allowed as found in the systematic search are given in Table 1 for the five nonpolar amino acids in the coil and helix states, respectively. All conformers are used in subsequent partition function calculations for all molecules and states except for the coil states of Ile and Leu. Extrapolation was used for those two systems. As shown by Wang *et al.*,[7] an excellent linear correlation exists between A_N, the cumulative free energy, and ln N where N is the number of conformers included in the calculation thus far from a random sampling of the database without repetition. This is illustrated in Figure 1, where A_N is plotted versus ln N for the Ile and Leu random coil tetrapeptides. A linear regression fit of the *in vacuo* calculated free energies A_N *versus* ln N gives, for Ile and Leu

(12) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563−595.

(13) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660−679.

(14) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1993**, *232*, 639−659.

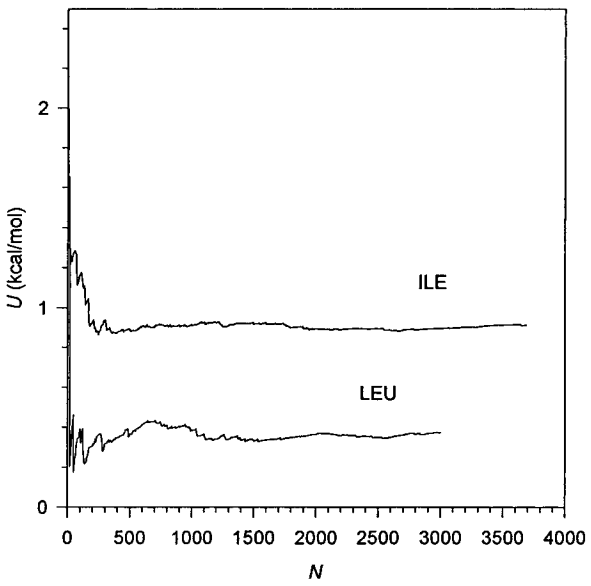


Figure 2. Plots of internal energy U versus N for Ile (upper curve) and Leu (lower curve) in the tetrapeptide coil state. Equilibrium is attained in the early stages of sampling.

Table 2. Calculated and Experimental Relative Free Energies (kcal/mol) for the Coil-to-Helix Transition

	$\Delta\Delta A^a$	$\Delta\Delta A^b$	$\Delta\Delta A^c$	$\Delta\Delta G^d$	$\Delta\Delta A^e$
Leu	−2.12	−1.65	−1.58	−1.60	
Ala	−1.56	−1.26	−1.19	−1.88	−1.2
Ile	−1.62	−1.17	−1.06	−1.18	
Val	−1.33	−0.78	−0.73	−0.83	−0.5
Gly	0	0	0	0	0

[a] Calculated without the solvation term and with dielectric $\epsilon = 80$, 25 °C. [b] Calculated with the solvation term at 25 °C. [c] Calculated with the solvation term at 0 °C. [d] Experimental values measured at 0 °C by Chakrabartty *et al.*[15] [e] Free energy perturbation results of Hermans *et al.*[9]

respectively,

$$A_N = -7.117 - 0.601 \ln N, \quad r = -0.9995 \qquad (21)$$

$$A_N = -7.531 - 0.597 \ln N, \quad r = -0.9984 \qquad (22)$$

We see in Figure 1 that due to log scale, extrapolation can be done with a high degree of confidence. The final free energies of these systems were obtained by taking $N = N_{tot}$ into the above equations. The corresponding entropies were obtained similarly.

The internal energy U and $\langle E \rangle$ also converge rather quickly (fluctuation <0.1 kcal/mol for both Ile and Leu) after a short stage of sampling (Figure 2). Similar properties are also observed for the other quantities which are an average or ensemble average of microscopic states.

Free Energies *in Vacuo* and in Water. The calculated relative free energy changes both *in vacuo* and with solvation are shown in Table 2 along with experimentally determined helix propensities.[15] The results for the *in vacuo* and solvated calculations share the same trend for the helix propensities although the spread in values is smaller in the solvated model. The calculated $\Delta\Delta A$ are in good agreement with the results of Chakrabarrty *et al.*[15] which were based on short alanine-dominant peptides. A notable exception is Ala which is underestimated in our calculations relative to their data.[15] The trend Leu > Ile > Val > Gly that we observe is consistent with what is generally found in various experimental and theoretical studies.

Table 3. Statistical Weights p of Most Probable Conformers of an Amino Acid in a Polyalanine Helix and the Corresponding Backbone (φ, ψ) and Side Chain χ Torsional Angles (in deg)

	p	φ	ψ	χ_1	χ_2
Gly	1.00	−61.2	−48.0		
Ala	1.00	−62.4	−46.0		
Val	0.90	−62.3	−46.0	−69.0	
	0.07	−67.3	−35.3	65.4	
	0.03	−66.7	−37.4	−166.0	
Ile	0.77	−62.3	−47.2	−68.6	163.6
	0.16	−62.7	−45.2	−65.1	−69.6
	0.04	−66.6	−37.4	−165.6	171.7
	0.01	−63.3	−43.1	−88.0	59.3
	0.01	−66.3	−37.3	−171.1	60.2
Leu	0.56	−62.2	−47.6	178.2	60.8
	0.31	−63.0	−44.3	−68.7	164.9
	0.08	−61.1	−50.4	−173.2	152.7
	0.04	−60.9	−51.9	−176.0	−78.1
		−62.8	−43.7	−91.5	61.4

[a] Definition of the torsional angles follows the IUPAC-IUB (1970) convention[26] except for the χ_1 angle of valine which is defined by N−C$^\alpha$−C$^\beta$−C$^{\gamma_2}$ instead of N−C$^\alpha$−C$^\beta$−C$^{\gamma_1}$ for facilitating comparison of conformations which isoleucine, as by Richardson and Richardson.[27]

Table 4. Relative Solute Configurational Entropy and Effective Internal Energy

	in vacuo		solvated	
	$\Delta\Delta U$	$-T\Delta\Delta S_v$	$\Delta\Delta U^*$	$-T\Delta\Delta S_v$
Leu	−2.20	0.07	−1.46	−0.19‘
Ala	−1.42	−0.14	−0.82	−0.44
Ile	−1.87	0.25	−1.30	0.13
Val	−1.49	0.16	−0.80	0.02
Gly	0	0	0	0

[a] All values are in kcal/mol.

Structural Properties. The statistical weights and torsional angles of the most probable conformers of the central residue Xaa in our helix are reported in Table 3. Conformers with probability less than 0.01 are not shown. Both the probabilities and torsional angles are consistent with the results from a statistical survey of protein crystal structures for residues in α-helices.[5] For example, the three conformers of Val have probabilities 0.90, 0.07, and 0.03 in our results as compared to 0.90, 0.06, and 0.04 in the statistical survey. Our two major conformers of Ile have probabilities 0.77 and 0.16 while it is 0.79 and 0.13 in the crystal structures. Our two major conformers of Leu have probabilities 0.55 and 0.31 while it is 0.37 and 0.52 in the survey. The overall agreement with the experimentally observed side-chain rotamer distributions supports the adequacy of the energy functions used in the calculations. Our results for valine are also consistent with those of Yun and Hermans[16] from molecular dynamics simulations with explicit solvent.

Solute Configurational Entropy. The relative free energies can be decomposed into the solute configurational entropy $-T\Delta\Delta\langle S_v \rangle$ and $\Delta\Delta U$ or $\Delta\Delta U^*$, which for the solvated model is an effective internal energy term that includes the average solvation free energy. Table 4 lists these quantities for both the *in vacuo* and solvated model. Solvation alters the magnitudes of the numbers but preserves the trends among the different residues. In Table 5 we list the individual entropic terms for the coil and helical states from which we can calculate the change in configurational entropy of that residue in going

(15) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843−852.
(16) Yun, R. H.; Hermans, J. *Protein Eng.* **1991**, *4*, 761−766.
(17) Stites, W. E.; Pranata, J. *Proteins* **1995**, *22*, 132−140.

Table 5. Configuration Entropy of the Solute in the Coil and Helix States and the Change in the Coil-to-Helix Transition[a]

	in vacuo			solvated		
	$-TS_v^{coil}$	$-TS_v^{helix}$	$-T\Delta S_v^{c\rightarrow h}$	$-TS_v^{coil}$	$-TS_v^{helix}$	$-T\Delta S_v^{c\rightarrow h}$
Leu	−12.88	−11.02	1.86	−12.66	−10.94	1.72
Ala	−6.73	−5.08	1.65	−6.55	−5.08	1.47
Ile	−12.60	−10.57	2.03	−12.33	−10.29	2.04
Val	−9.44	−7.50	1.94	−9.22	−7.29	1.93
Gly	−6.87	−5.08	1.79	−6.99	−5.08	1.91

[a] All quantites are in kcal/mol. $T = 298$ K.

Table 6. Intramolecular and Solvation Contributions to the Free Energy[a]

	$\Delta\Delta\langle E_v\rangle$	$\Delta\Delta\langle H_w\rangle$	$-T\Delta\Delta\langle S_w\rangle$	$\Delta\Delta\langle\Phi_{rf}\rangle$	$\Delta\Delta\langle\Phi_{surrf}\rangle$	$\Delta\Delta\langle\Phi\rangle$
Leu	−1.30	0.47	−0.58	0.08	−0.19	−0.11
Ala	−0.53	0.01	−0.31	−0.20	−0.10	−0.30
Ile	−1.11	0.61	−0.76	0.10	−0.25	−0.15
Val	−0.70	0.39	−0.48	0.06	−0.15	−0.09
Gly	0	0	0	0	0	0

[a] $\Delta\Delta\langle E_v\rangle$, $\Delta\Delta\langle H_w\rangle$, $-T\Delta\Delta\langle S_w\rangle$, $\Delta\Delta\langle\Phi_{rf}\rangle$, $\Delta\Delta\langle\Phi_{surrf}\rangle$, and $\Delta\Delta\langle\Phi\rangle$, are the relative mean intramolecular energy, solvation enthalpy, solvation entropy, reaction field energy, cavitation free energy, and solvation free energy, respectively. All quantites are in kcal/mol.

from a random coil to the middle of a helix. Several observations can be made. (1) The presence of a C^β atom in Ala reduces the entropic cost of helix formation by reducing the backbone conformational space of the coil state relative to that available to Gly. (2) Side chains larger than that of Ala incur an unfavorable entropic cost for helix formation, due to the more restricted space available to the side chain in the helical state as compared to the coil state. (3) β-Branched side chains have an even greater entropic cost for helix formation as compared to their nonbranched isomers. (4) From the results for Ala, we estimate that the backbone entropy change for helix formation is about 0.7 kcal/mol per degree of freedom (i.e., φ and ψ dihedral angles) for the non-glycine amino acids. The backbone entropy cost for Gly is larger at 0.95 kcal/mol per degree of freedom. (5) The cost of freezing a backbone torsional angle is greater than that for restricting a side chain torsion in the helix.

Creamer and Rose[3,4] calculated side chain conformational entropy in α-helix folding and found that the values for Leu, Ile, and Val relative to Ala are 0.15, 0.29, and 0.39 kcal/mol, respectively. Blaber *et al.*[5] derived this term based on rotamer distribution from statistics of protein crystal structures and gave values as 0.10, 0.38, and 0.27 comparatively. Lee *et al.*[6] found a set of values in similar range with variations depending on the systems and approaches used. Our results (0.25, 0.57, and 0.46) are comparable with these data although very different approaches were used. One notable difference is that our calculations included vibrational entropy which was not considered in the calculations of Creamer and Rose.[3,4] Stites and Pranata[17] have recently estimated the relative backbone conformational entropy of different amino acids in the unfolded state based on statistics of crystal structures of proteins. They found a value of 0.72 kcal/mol for $T\Delta S_{conform}$ between Ala and Gly which is somewhat larger than our result (0.44). Yang and Honig,[18] in a theoretical simulation of Zimm−Bragg parameters for polyalanine, estimate the entropic cost per residue of fixing the backbone dihedral angles of Ala to be about 2 kcal/mol at room temperature in good agreement with our calculated value of 1.5 kcal/mol.

Solvation Effects. Table 6 lists the average solute intramolecular energy as well as the various components of the solvation free energy term. Overall, we note that except for Ala, the net solvation contribution, $\Delta\Delta\langle\Phi\rangle$, is minor compared to the relative free energy change $\Delta\Delta A$ reported in Table 2. This is due to a compensation between the entropic and enthalpic components of the solvation free energy. $\Delta\Delta\langle\Phi_{surf}\rangle$ is calculated as a function of solvent-excluded surface area using eq 20. The trend in $\Delta\Delta\langle\Phi_{surf}\rangle$ follows the pattern that one would expect for the removal of nonpolar groups from water. We have Ile < Leu < Val < Ala < Gly = 0. With respect to the electrostatic component of solvation, $\Delta\Delta\langle\Phi_{rf}\rangle$, we find, except for Ala, the opposite trend Ala < Gly = 0 < Val < Leu < Ile. The reason that we have Ala < Gly for the reaction field term is that the presence of a βCH_3 group partially desolvates the CO and NH groups of Ala in the coil state. In the helix, the CO and NH groups are already partially desolvated due to the helical matrix and the additional desolvation due to the βCH_3 of Ala is less significant. Hence the electrostatic desolvation cost for Ala is less than that of Gly upon helix formation. As the side chain size is increased, more of the helix CO and NH groups are desolvated further. However, the coil state is not affected as much due to the diversity of backbone conformations in which the side chain methyl/methylene groups beyond C^β are directed away from the CO and/or NH groups. In Table 6, we see that the electrostatic solvation cost of helix formation increases in step with the amount of desolvated side chain area in the helix.

Table 7. Relative Total Internal Energies and Entropies (kcal/mol)

	$-\Delta\Delta U$	$-T\Delta\Delta S$
Leu	−0.88	−0.77
Ala	−0.51	−0.75
Ile	−0.54	−0.63
Val	−0.32	−0.46
Gly	0	0

Table 7 lists the total relative internal energies and entropies for the different amino acids. In principle, these would be the quantities that are directly related to experimentally measured relative enthalpies and entropies. The enthalpy differences between non-glycyl amino acids are small (variations <0.56 kcal/mol), which is consistent with the suggestion of Scholtz *et al.*[19] based on the experimental observations that the enthalpy changes of unfolding of α-helix of polyalanine, polylysine, and polyglutamate are very similar (1.3, 1.1, and 1.1 kcal/mol per residue for the three systems, respectively).

Discussion

The structural and thermodynamic factors governing the helix propensities of amino acids have been a subject of much interest and controversy. What are the relative contributions of entropy, van der Waals, and hydrophobic interactions? How important is solvation? In our simulation calculations, we sought to obtain some insights into these issues. Using the different decompositions of the free energy, we are able to carry out the analysis of the results at varying levels of detail.

Creamer and Rose[3,4] have proposed that the side chain conformational entropy should be the dominant factor for the helix propensities of nonpolar amino acids since they found that the calculated side chain entropy is correlated, in trend and magnitude, with two sets of experimental free energy differences $\Delta\Delta G$.[20,21] On the other hand, Blaber *et al.*[5] in their study conclude that conformational entropy alone cannot account for

(18) Yang, A.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 351−365.

(19) Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Santoro, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2854−2858.

(20) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646−651.

(21) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. R. *Science* **1990**, *250*, 669−673.

Table 8. Comparison of Theoretical and Experimental Effective Conformational Energy $\Delta\Delta U^*$ with Differential Interaction Energy $\Delta\langle E_{inter}\rangle$ between the Side Chain of a Central Residue and the Rest of the Molecule in a Polyalanine α-helix[a]

	$\Delta\Delta U^*$	$\Delta\Delta U^*_{expt}$	$D\langle E_{inter}\rangle$
Leu	−1.46	−1.43	−1.80
Ala	−0.82	−1.48	−0.90
Ile	−1.30	−1.30	−1.50
Val	−0.80	−0.85	−1.07
Gly	0	0	0

[a] All values in kcal/mol.

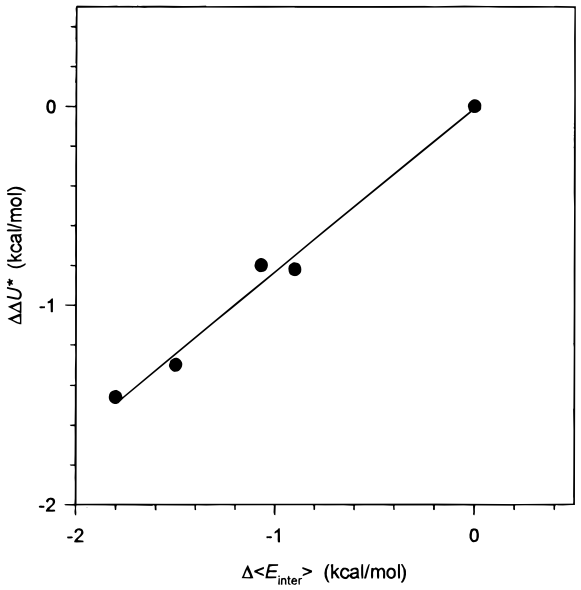


Figure 3. Plot of calculated effective conformational energy $\Delta\Delta U^*$ versus the side chain−helix interaction energy $\Delta\langle E_{inter}\rangle$ relative to Gly. The straight line is from the least-squares fit.

the $\Delta\Delta G$ that they observe, and that hydrophobicity should play a major role in determining helix propensity. Recent results from Chakrabartty *et al.*[15] also give free energy changes whose magnitudes are significantly larger than the entropies of Creamer and Rose[3,4] or the experimental values of Lyu *et al.*[21] Our results indicate that although the solute configurational entropy modulates the helix propensity of an amino acid, it is not by itself a dominant component in the calculated free energy changes.

$\Delta\Delta U^*$ contains the free energy components other than the solute configurational entropy. Since this is the effective solvated conformational energy, we explored the correlation of $\Delta\Delta U^*$ with $\langle E_{inter}\rangle$, the average interaction energy of a guest side chain with the helical matrix for the various side chains. The value of $\langle E_{inter}\rangle$ relative to Gly is reported in Table 8, together with $\Delta\Delta U^*$. A semiempirical $\Delta\Delta U^*_{expt}$ was also obtained using the experimental[15] $\Delta\Delta G$ plus the calculated TS_v. Figures 3 and 4 show that $\Delta\Delta U^*$ parallels the interaction energy $\Delta\langle E_{inter}\rangle$ with a linear correlation as,

$$\Delta\Delta U^* = -0.007 + 0.82\Delta\langle E_{inter}\rangle, \quad r = 0.993 \quad (23)$$

$$\Delta\Delta U^*_{expt} = -0.203 + 0.76\Delta\langle E_{inter}\rangle, \quad r = 0.85 \quad (24)$$

The poorer correlation with $\Delta\Delta U^*_{expt}$ is due to the discrepancy between the theoretical and experimental helix propensities for Ala. The fitted straight lines have intercept near zero and slope fairly close to unity. The calculated $\Delta\langle E_{inter}\rangle$ are larger than the $\Delta\Delta U^*$. This is probably because some of the side chain−helix interactions that are included in $\langle E_{inter}\rangle$ are also present in the coil state and hence their contribution to the free energy is

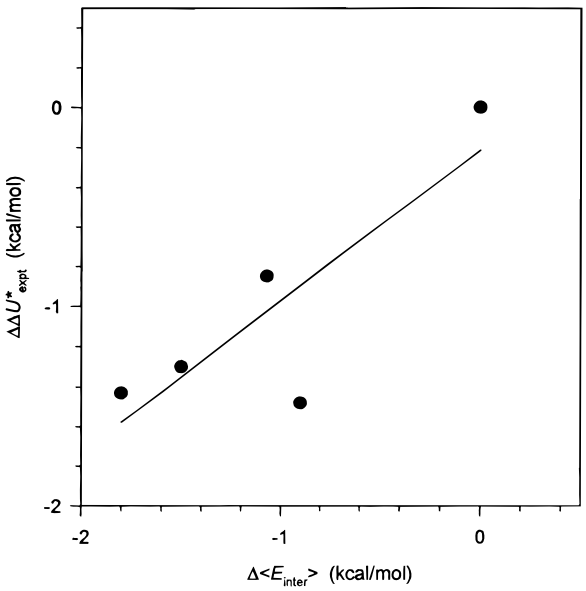


Figure 4. Similar to Figure 3 but with a semiempirical $\Delta\Delta U^*$ obtained from subtracting the theoretical $-T\Delta\Delta S_v$ from the experimental $\Delta\Delta G$.

overestimated. The correlation suggests that side chain−helix van der Waals interactions, which form the bulk of $\langle E_{inter}\rangle$ for these nonpolar side chains, are a major factor affecting the relative helix propensities of the amino acids in addition to the solute configurational entropy. Note that these interactions do not include specific side chain−side chain interactions since a polyalanine host is used in our calculations.

On the issue of solvation, there is no doubt that solvation has an important effect on the absolute helix propensity of an amino acid. However, on the specific question of the role of solvation in the incremental change in helix propensity due to variations in the structure of nonpolar side chains, our calculations show that solvation has only a marginal effect. The only significant effect observed was in examining the effect of having or not having a side chain, i.e., Ala versus Gly. In that case, we saw that the presence of the methyl affects the electrostatic interactions of the coil state by modifying the solvent exposure of the nearby amide polar groups. However, for groups beyond the C^β, the solvation effect is minimal.

Avbelj and Moult[22] have proposed that the electrostatic interactions of the backbone CO and NH with the solvent and/or protein environment is a major determinant of the conformational preference of an amino acid. The effect of the side chain was thought to be the modulation of the electrostatic screening of these interactions by the environment. In our results, we do observe that the nature of the side chain modulates the calculated reaction field energy. However, the magnitude of our calculated relative changes in the reaction field energy for different side chains is small compared to the total relative free energy changes for the coil-to-helix transition. Thus, our calculations do not support the proposal of main chain electrostatic screening as a major determinant in helix propensities.

(22) Avbelj, F.; Moult, J. *Biochemistry* **1995**, *34*, 755−764.
(23) Lyu, P. C.; Sherman, J. C.; Chen, A.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5317−5320.
(24) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1990**, *344*, 268−270.
(25) Wójcik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121−134.
(26) IUPAC-IUB Commission on Biochemical Nomenclature *J. Mol. Biol.* **1970**, *52*, 1−17.
(27) Richardson, J. S.; Richardson, D. C. In *Prediction of Protein Structure and the Principles of Protein Conformation*; Fasman, G., Ed.; Plenum Press: New York, 1988; pp 1−98.

The picture that emerges from our study is as follows. For a pure polypeptide backbone (i.e., polyglycine), we can infer from the low helix propensity of glycine that the enthalpic gain from hydrogen bonds in the helix is insufficient to overcome the loss in backbone entropy. However, the introduction of a C^{β} atom has three helix-inducing effects. First, it destabilizes the coil state by partially desolvating the polar groups in the neighboring amide bonds. Second, the coil state is further destabilized by the reduction in entropy due to the more restricted conformational freedom of the backbone. Third, side chain−helical matrix van der Waals interactions stabilize the helix. Beyond the C^{β} atom, further additions to the side chain have no large net solvation contribution nor do they impose significant extra restrictions on the backbone conformation in the coil state. Longer side chains in the helix do incur an unfavorable entropic cost due to the more restricted space available to them. This is partially offset by van der Waals interactions with the helix and burial of hydrophobic surfaces. Comparatively, the configurational entropy change due to the reduction of side chain entropy is much less than that for the reduction of backbone entropy upon helix formation. This is not unexpected since the side chains in the α-helix retain a fair amount of conformational freedom. In a study[7] of the binding of thrombin inhibitors where the side chains of the ligand are restricted to a single conformation in the complex, the calculated configurational entropy loss due to the side chains was about 0.48 kcal/mol per degree of freedom frozen, which is still smaller than the backbone entropy cost of 0.7 to 0.9 kcal/mol per degree of freedom frozen found here. This is due to the greater vibrational entropy of the side chain even when restricted to a single potential energy well. Our simulations also suggest that although solute configurational entropy is important, it is not dominant. Side chain−helical matrix interactions provide an even greater contribution to the helix propensity.

Our calculated free energies can be broken down into four components: $\Delta\Delta\langle E_v\rangle$, $\Delta\Delta\langle H_w\rangle$, $-T\Delta\Delta S_v$, and $-T\Delta\Delta\langle S_w\rangle$. We see in Tables 4 and 6 that the total relative free energy is the sum of terms of similar magnitudes, some of which have opposite signs. No single factor dominates the free energy and the final outcome is a delicate balance among the different factors. Clearly, any one of these factors can be altered by the choice of model system and conditions used in either an experimental measurement or theoretical simulation. It is quite understandable then that using different approaches one gets a diverse set of helix propensities that differ in both magnitude and rank order for the amino acids.[5,15,20,23−25] In our simulations, as in those by others,[3,4,9] we used a polyalanine host sequence in order to minimize the effects of the specific sequence on the helix propensity of the guest residue of interest.

Conclusion. The "intrinsic" helix propensity of amino acids is an elusive quantity to measure experimentally or to calculate theoretically. The delicate balance among the contributing factors and the small variation in helix propensities (sub-kcal/mol differences in free energies of helix formation among the different non-glycine and non-proline amino acids) make the choice of the proper experimental model difficult. With simulation methods one has somewhat greater control in defining, constructing, and manipulating the model system. The trade-off is uncertainty over the realism of the potential function used to describe the system. In this paper, we have used a novel free energy method[7] to estimate relative helix propensities for a set of nonpolar amino acids. More importantly, we have dissected out the various contributing factors to helix propensity. We hope that some insights into the process of helix formation have been gained.

Acknowledgment. We wish to thank Hervé Hogues for useful discussions and programming assistance. This is National Research Council of Canada publication No. 38563.

JA953347O