Differential Helix Propensity of Small Apolar Side Chains Studied by Molecular Dynamics Simulations†

Jan Hermans,* Amil G. Anderson,‡ and R. H. Yun

Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599-7260

Received November 25, 1991; Revised Manuscript Received March 19, 1992

ABSTRACT: A series of oligoalanine molecules with single amino acid replacements in the middle of the chain has been studied by molecular dynamics simulations. Differences in stability of the α-helix (as free energies $\Delta\Delta G°$) were estimated for the following series of residues: α-aminoisobutyric acid, alanine, α-amino-*n*-butyric acid, valine, glycine, D-alanine, *t*-leucine (= α-amino-β,β-dimethyl-*n*-butyric acid), and proline, arranged here in decreasing order of helix-forming potential. (The results for proline and valine had been reported earlier.) No experimental results were available for α-amino-*n*-butyric acid, D-alanine, and *t*-leucine at the time these calculations were done. The values of $\Delta\Delta G°$, including the three predictions, are in striking agreement with recent experimental results. A combination of free dynamics, dynamics with forced conformational change, and dynamics with forced molecular replacement was used. Conformational distributions were calculated for the peptide backbone of the dipeptides and, where appropriate, for the side chains of the dipeptide and the α-helix. The results demonstrate an unexpected level of accuracy for the all-atom model used to represent atomic interactions in the simulations. The simulations permit a detailed analysis of different factors responsible for conformational preferences and differences in stability. These conclusions drawn from this analysis agree with accepted qualitative explanations and allow these explanations to be quantitated to an extent not heretofore possible.

Much effort has been and continues to be devoted to try to understand the relation between amino acid sequence and the conformation of peptides and proteins. Most sequences do not spontaneously assume a unique stable folded structure under physiological conditions. No set of rules has yet been found that determines when a sequence will assume a unique native conformation or what this conformation will be. Not much is known about the quantitative relationship between the stability of a single folded conformation and the amino acid sequence, a notable exception being the α-helix.

The helix is a well-known structural element of globular proteins, and it has long been known that certain amino acids occur with disproportionately high or disproportionately low frequency in helical structures or at the ends of helical segments (Chou & Fasman 1978; Richardson & Richardson, 1988b; Presta & Rose, 1988). Also, high-resolution X-ray structures of proteins indicate strong preferences for certain side-chain conformations of many amino acids and a dependency of these preferences on structural context which is expecially strong for the α-helix (Janin et al., 1978; Richardson & Richardson, 1988a).

Recent experiments with synthetic oligopeptides have produced a set of reference data from which differences in helix stability due to single amino acid replacements can be estimated, which, if not uniquely, at least narrowly determine a set of "canonical" free energy differences (Padmanabhan et al., 1990; Lyu et al., 1990; Merutka et al,. 1990; O'Neill & DeGrado, 1990). It is worth mentioning that these differences are systematically larger than those determined earlier by Scheraga and co-workers, who used random host–guest copolymers, the to-be-tested guest being in a low proportion to the host residue, hydroxypropyl (or butyl) glutamine (Sueki et al., 1984). It is likely that these differences between the copolymer results and the more recent results obtained with oligomers of well-determined length and sequence are real and should be attributed to the difference in the lengths of the host side chains in the copolymer system and the oligopeptides.

Differences in helix propensity have been calculated here with a molecular-mechanics model, using various techniques based on molecular dynamics simulations, in particular free energy calculations. A major technical problem was posed by the need to deal with multiple conformers, i.e., a small number of conformers for the longer side chains in the helical state and a very large number of backbone and side-chain conformers for an oligopeptide in the random coil state. How this problem was explicitly dealt with is described in detail below. Another problem was that of the relatively small changes in helix stability due to most replacements. Except for proline, experimental free energy differences, $\Delta\Delta G°$, amount to at most 5 kJ/mol. We have consequently made an effort to study first those amino acids for which the differences are large, including residue types not commonly found in globular proteins. In addition, we developed methods for assessing the precision of the calculated free energies and ensuring that these were within the 1 kJ/mol limit needed for a meaningful comparison with the experimental results (Hermans et al., 1992).

This work could have led us to either one of two alternative results. At one extreme, it could have been found that the calculations did not reproduce the experimental results. This would of necessity have led to a reassessment of the molecular mechanics model, especially of its parameters; indeed, the earliest results indicated that the central-atom model (in which hydrogens bonded to carbon were not individually represented) was inadequate, and therefore this work was continued with

† Supported by a research grant from the National Science Foundation's Biophysics Program (DMB-8903584) and an instrumentation grant from the National Science Foundation's Biological Instrumentation Program (DIR-8908054).

* Address correspondence to this author.

‡ Present address: Department of Chemistry, Wittenberg University, Springfield, OH 45501.

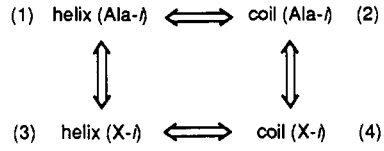
0006-2960/92/0431-5646$03.00/0

a model in which all atoms were represented as attractive and repulsive centers. With the all-atom model, we have come close to obtaining the other extreme result, namely, good agreement between calculated and experimental differences in helix-forming potential for different types of residues. This agreement encouraged us to calculate several predicted differences. (Two of these have been confirmed experimentally more or less simultaneously with the predictions; the third has been confirmed after this paper had been submitted for publication.)

The most important consequence of the good agreement is that it gives license to interpret details of the results of the simulations (details that themselves are not verified experimentally) and so develop a more detailed insight into the structural causes of the free energy differences. The analysis given here confirms the current qualitative understanding of the causes of differences in helix stability and allows a quantitative restatement.

Materials and Methods

Outline. The change in stability of a short oligoalanine α-helix upon replacement of one of the alanine residues by another residue, X, can be computed as the difference of the free energy of molecular replacement in two simulations, one simulation for each of two different conformations, helix and random coil. This can be seen to follow from the cyclic scheme involving helical and random coil states of the alanine and X variants:



(1)

The condition that the sum of the free energy changes for the entire cycle must be zero, can be rewritten

$$\Delta\Delta G^\circ = \Delta G^\circ_{12} - \Delta G^\circ_{34} = \Delta G^\circ_{13} - \Delta G^\circ_{24} \qquad (2)$$

where the first difference is in terms of experimentally accessible free energies of folding, and the second difference is computed for the model with molecular replacement simulations (Berendsen et al., 1985; Tembe & McCammon, 1984; Mezei & Beveridge, 1986; Beveridge & DiCapua, 1989; van Gunsteren, 1989; Dang et al., 1989).

Multiple Conformation States. Application of molecular replacement calculations is complicated by the fact that peptide molecules assume multiple conformation states separated by energy barriers that, on the time scale of the simulations, are crossed infrequently. As a consequence, it is not practical to calculate directly a difference, such as ΔG°_{24}, with appropriate contributions from all possible conformation states of molecules 2 and 4.

In order to deal with the problem of multiple conformations, we have adopted a new strategy to prevent random transitions between conformation states: these transitions have been eliminated by the introduction of torsional bound potentials, which confine the molecule to a single conformation state and prevent crossings of free energy barriers without perturbing the distribution near the selected free energy minumum. As a result, all free energy differences have been calculated for pairs of systems that are conformationally restricted, by torsional restraints or by torsional bounds, as follows.

A torsional restraint is imposed by use of an extrinsic potential energy term of the form

$$U_f = (U_{f,0}/2)[1 - \cos(\rho - \rho^*)] \qquad (3)$$

with a constant value of ρ^*. The higher the value of the energy parameter $U_{f,0}$, the closer the value of the dihedral angle ρ will remain to the artificial energy minimum ρ^*. Such a "round-bottom" potential obviously perturbs the system considerably.

A much less perturbing restraint can be created with a "flat-bottom" potential, which consists of (torsional) lower and upper bounds. At the lower bound, the potential energy drops smoothly (by $U_{\rho,0}$) over an interval ($\rho_{1,L} < \rho < \rho_{2,L}$), and at the upper bound it increases smoothly by the same amount over another interval ($\rho_{1,U} < \rho < \rho_{2,U}$). The applied torque is zero at the limits of the two energy ramps (ρ_1 and ρ_2).

Torsional bounds should be applied such that the ramps coincide with regions of conformation space where the free energy is high (peaks and saddles), so that the probability distribution near one particular free energy minimum is sampled essentially correctly, and other free energy minima are never sampled. The fact that the extrinsic potential perturbs the saddle point regions does not have a significant effect on equilibrium properties, since the saddle point regions do not usually contribute to these properties.

For comparison with experiment, it is necessary to calculate free energy differences for conformationally unrestricted systems. This is accomplished by calculating restraint free energies that correspond to the extent of torsional confinement and adding these as corrections to the replacement free energies, as follows. The relationship between the free energies for taking system A to system B either with full conformational freedom or else restrained to a particular conformer, ΔG°_{AB} or $\Delta G^\circ_{AB,r}$, can be written as

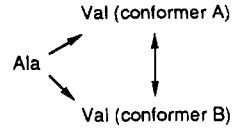
$$\Delta G^\circ_{AB} = \Delta G^\circ_{r,A} + \Delta G^\circ_{AB,r} - \Delta G^\circ_{r,B} \qquad (4)$$

where $\Delta G^\circ_{r,A}$ and $\Delta G^\circ_{r,B}$ represent the free energies for restraining the two systems each to a particular conformer. The restraint free energy for system A is given by

$$\Delta G^\circ_{r,A} = -kT \ln [1 + \sum_j \exp(-\Delta G^\circ_{j,A}/kT)] \qquad (5)$$

where the $\Delta G^\circ_{j,A}$ represent the free energies of the other conformers, relative to the one selected by the restraint, and similarly for B. The correction is small if the selected conformer is the most stable one, in which case all $\Delta G_{j,A}$ are greater than zero.

Free Energy Calculations. Thus, the relative free energies of, in principle, all conformation states of the two systems have been calculated, using either one of two alternative methods. The first of these methods is a direct calculation in which a potential is applied that forces the system across the free energy barrier by internal rotation. The potential has the form of eq 3, but with varying ρ^*. Results obtained thus have been indicated in the tables by "twist". The second method consists of calculating successively the free energies for replacing a simpler molecule (e.g., alanine) with different side-chain conformers of the more complicated one (e.g., valine) and subtracting the results to obtain the free energy difference(s) between the latter's conformers, for example, using the scheme



(6a)

one has

$$\Delta G^\circ_{Val,A\rightarrow B} = \Delta G^\circ_{Ala\rightarrow Val\text{-}B} - \Delta G^\circ_{Ala\rightarrow Val\text{-}A} \qquad (6b)$$

In a molecular replacement calculation, the two alternate residues are both represented, and the energy of the system

is calculated with contributions from each that vary according to the value of a coupling parameter λ. In the Cedar program this is implemented as follows. Where reasonable, equivalent non-hydrogen atoms of each residue are maintained at identical positions; at least one atom in at least one residue is "unique". For example, when replacing alanine with valine, valine has the unique non-hydrogen atoms; accordingly, simulation at $\lambda = 1$ produces the valine-containing form and at $\lambda = 0$ the alanine-containing form. Forces for geometric deformations are multiplied by a coupling parameter (λ for one residue and $1 - \lambda$ for the other) for all terms which involve only shared atoms and are fully counted when any unique atoms are affected. Nonbonded force terms for atom pairs in which at least one atom is part of the residue containing the unique atoms are multiplied by powers of the coupling parameter: λ^3 for the attractive Lennard-Jones term and the electrostatic term and λ^5 for the repulsive Lennard-Jones term (Hermans et al., 1988), while those for the other residue are multiplied by $(1 - \lambda^3)$ and $(1 - \lambda^5)$.[1] At the end points ($\lambda = 0$ and $\lambda = 1$), the inactive side chain's geometry is maintained; the additional mass will perturb the dynamic properties but not the conformational equilibrium properties of the system.

The conformer free energies have also been used in order to combine conformational distributions, computed (sampled) for individual conformations, into a single conformational distribution, as was first done for the alanine dipeptide (Anderson & Hermans, 1988).

Free energy simulations were done with the slow-change method. The free energy change is equated to the work done on the system by the changing potential in a quasistatic process, according to

$$\Delta G^\circ = \int \langle \partial U/\partial \rho^* \rangle \, d\rho^* \tag{7a}$$

or

$$\Delta G^\circ = \int \langle \partial U/\partial \lambda \rangle \, d\lambda \tag{7b}$$

The $\langle\rangle$ signs indicate averages over a Boltzmann distribution (at each value of ρ^* or λ). Each integral is estimated in a molecular dynamics simulation in which the value of ρ^* or λ is changed by the same small increment (or decrement) $\delta\rho^*$ or $\delta\lambda$ after *every* step, as

$$\Delta G^\circ_a = \sum(\partial U/\partial \rho^*) \, \delta\rho^* \tag{8a}$$

or

$$\Delta G^\circ_a = \sum(\partial U/\partial \lambda) \, \delta\lambda \tag{8b}$$

The system is equilibrated at constant ρ^* or λ following this simulation, and then the process is reversed; after another equilibration at constant ρ^* or λ, the cycle may be repeated. In order to eliminate hysteresis, each value has been obtained as the mean of the results of two (or a multiple of two) simulations in opposite directions along the same path connecting the two end states.

The results reported in this paper have been calculated with use of an all-atom representation (in which all atoms, including all hydrogen atoms, are represented as individual attractive and repulsive centers). As in earlier work, peptide molecules contained an acetyl group at the amino terminus and an N-methylamide group at the carboxy terminus and were studied in aqueous environment using explicit water molecules and periodic boundary conditions (Yun et al., 1991; Yun & Hermans, 1991; Hermans et al., 1992). Additional details of the methods used can be found in these three papers from this laboratory.

Model for the Helix. The helical state was represented by polypeptides containing 14 Ala residues, in which residue number 8 was chosen to be replaced with other amino acid residues. [In the study on valine, 14 residues was found to be an adequate length (Yun & Hermans, 1991).] Torsional bound restraints were used to keep the values of all backbone dihedral angles ϕ and ψ in the interval between $-\pi$ and 0. These restraints prevent unfolding of the helix without perturbing the helical state; hence, no corrections are needed for the presence of these restraints. Side-chain dihedral angles of the guest residue were restrained to individual conformers as necessary.

Model for the Random Coil State. The number of conformation states accessible to the random coil state of a peptide of between 10 and 20 residues long is far too great to contemplate doing a calculation of the conformational free energy taking into account contributions from all conformers. In order to deal with the coil state, the assumption is made that the conformational distributions of the individual residues in the coil are independent. This assumption was first made and justified some time ago by Flory and co-workers in a study of the conformation of oligopeptide chains [cf. Brant and Flory (1965)]. In the present context, the resulting agreement between theory and experiment confirms that this assumption is a good approximation for short peptides.

With this assumption, it is possible to estimate the free energy difference in the coil state between Ala and some other residue, X, by studying two short peptides, such as dipeptides, i.e., Ace-Ala-Mam and Ace-X-Mam, and tetrapeptides, i.e., Ace-Ala-Ala-Ala-Mam and Ace-Ala-X-Ala-Mam. To be precise, for an L-amino acid, we calculate relative free energies of all side-chain conformations, for each of the two most stable backbone conformations, β and α_R, of the *di*peptide. Furthermore, we calculate the free energy for interconverting the most stable conformer of the Ala-*tetra*peptide to the most stable conformer of the X-tetrapeptide. To this value is then added a correction for the presence of the restraints that have been used to select these unique conformers; this is calculated using the relative free energies for the *di*peptide conformers, as indicated above (eqs 4 and 5).

Random Errors. Additional methods have been developed in the course of this work to estimate errors in calculated free energy differences. Using the valine dipeptide as a test system, we have shown that, for the kind of transformations studied here (extension of a hydrocarbon side chain and conformation change by internal rotation about a single bond) and for simulation times equal to and exceeding the ones used in this study, (a) the mean free energies from sets of several successive *pairs* of forward and reverse transformations are independent of the simulation time and (b) the mean square deviation of the free energies decreases when the simulation time is increased (Hermans et al., 1992). In the valine study, it was concluded that it is preferable to estimate free energy differences by a *series* of short calculations and thereby obtain an estimate of the variance of the result from the mean-square deviation, instead of by a single very long calculation (as had been our previous method). Most of the results reported in this paper had been calculated from single pairs of long simulations, at a time before this was realized. Extensive calculations of helix stability of the valine residue by both methods

[1] Bond lengths are kept fixed with SHAKE (Ryckaert et al., 1977). The use of λ^3 and λ^5 as multipliers is equivalent to varying both ϵ and σ^3 as λ when using the common $\epsilon - \sigma$ form of the Lennard-Jones equation. With this method, the "volume" of Lennard-Jones particles is proportional to λ. (The multipliers for electrostatic terms are the same as for the attractive Lennard-Jones terms.)

Table I: Conformational Free Energy Differences for Glycine and Alanine

final state	reference state	method	data set[a] (ps)	$\Delta G°$ (kJ/mol)	rmsd[b]
dipeptide					
Ala dipeptide, α_R	β conformation	twist	4 × 200	4.3	1.5
tetrapeptide[c]					
Gly tetrapeptide, β	D-Ala tetrapeptide	replace	800	−6.3	
Gly tetrapeptide, β	L-Ala tetrapeptide	replace	400	−5.1	
correction for conformational restraint[d]				−3.3	
net $\Delta G°_{13}$, D-Ala to Gly [= −6.3 + (−3.3)]				−9.6	
net $\Delta G°_{13}$, L-Ala to Gly [= −5.1 + (−3.3)]				−8.4	
α-helix					
Ala–Gly helix	all-L-Ala helix	replace	400	−3.8	
$\Delta G°_{24}$, L-Ala to Gly				−3.8	
$\Delta\Delta G°$, L-Ala to Gly [= −3.8 − (−8.4)]				**4.6**	
Ala–Gly helix	Ala–D-Ala helix	replace	400	−10.0	
$\Delta G°_{24}$, D-Ala to Gly				−10.0	
$\Delta\Delta G°$, D-Ala to Gly [= −10.0 − (−9.6)]				**−0.4**	
$\Delta\Delta G°$, L-Ala to D-Ala [= 4.6 − (−0.4)]				**5.0**	
Ala–D-Ala helix	all-L-Ala helix	replace	3 × 80	6.1	0.3
$\Delta\Delta G°$, L-Ala to D-Ala [= 6.1 − [−5.1 − (−6.3)]]				**4.9**	

[a] Number of runs (if greater than 1) and simulation time (ps) for each half of the free energy calculation. [b] For multiple results only, root-mean-square deviation. [c] The two states indicated by α_R and β refer to the equilibrium *ensembles* of conformations near the two principal free energy minima of the dipeptide [Figure 1 of Anderson and Hermans (1988)]. [d] Free energy for imposition and removal of conformational restraints on ϕ, ψ, and/or χ.

have shown the precision of these long simulations to be of the order 1 kJ/mol.

Results

L-*Alanine*

The local conformation distributions and the free energy difference for the β and α_R conformers of the alanine dipeptide have been recalculated using the all-atom potential that is the basis of the work reported in this paper. The distributions are very similar to those reported for the central-atom potential; however, the free energy difference is significantly less (4 kJ/mol rather than 6 kJ/mol; cf. Table I).

Glycine

Dipeptide. The glycine dipeptide was studied in long simulations, in earlier work with the central-atom model (Hermans & Anderson, 1990) and in this work with the all-atom model, with very similar results. In contrast to what was found for alanine, and for other amino acids, the readily accessible conformation space is barely divided into minima (Figure 1). The dipeptide readily changes conformation within the low-energy space; thus, this distribution was calculated from a single long free dynamics simulation.

Tetrapeptide and α-Helix. Free energy differences for substituting L-alanine with glycine in the tetrapeptide and the α-helix are given in Table I. As was explained under Materials and Methods, molecular replacement calculations were done for systems in which backbone and side-chain conformations are restrained. In this case, the backbone of the tetrapeptide was restrained to conformations for which ϕ lies in the range between $-\pi$ and 0, and ψ in the range between 0 and $+\pi$. This range includes the most stable conformations of both the L-alanine and the glycine residues. In order to obtain the free energy difference for the unrestrained molecules, we have applied a correction, equal to the difference of the free energies for imposing the restraint on the two residues. Because the distribution for the glycine dipeptide does not neatly separate into a set of conformer distributions, the restraint correction for this residue has not been calculated with use of eqs 4 and 5, but by applying the equivalent equation 8 of Hermans et al. (1992) to the entire conformation distribution (Figure 1). For the alanine residue, eqs 4 and 5 were applied to the two most stable conformers, β and α_R (cf. Table I).

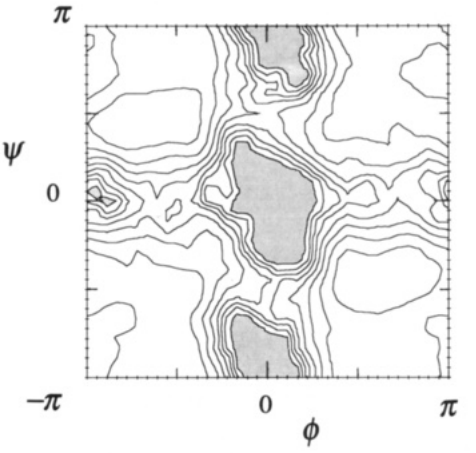


FIGURE 1: Free energy map for the backbone conformation of the glycine dipeptide based on unrestrained simulations over a total simulation time of 2.4 ns, without symmetry averaging. Contours are drawn at intervals of 2 kJ/mol. The minima are near −90°,+90° and +90°,−90°.

Clearly the "correction" term makes a large contribution to the difference in helix potential of L-alanine and glycine. The restraint free energy is smaller for alanine than for glycine, and the difference is, of course, due to the well-known difference in conformational freedom of the backbones of the alanine and glycine residues. However, as the correction term depends on the exact form of the restraint potential, it is not a good measure of the difference in conformational freedom.

One may uniquely define the conformational free energy of a residue in terms of the likelihood of all conformations, relative to the most stable conformation. In terms of probabilities, P, this gives

$$\Delta G°_{conf} = -kT \ln \left(\sum_j P_j / P_{max}\right) \quad (9)$$

Subtracting two such free energies, for different types of residues, X and Y, and noting that by definition $\sum_j P_j = 1$, one has

$$\Delta G°_{conf,XY} = -kT \ln \left(P_{max,Y} / P_{max,X}\right) \quad (10)$$

which can be easily obtained from the $\phi - \psi$ distributions for the backbone.

The ratio of the maximum probabilities in the $\phi - \psi$ distributions of the glycine and alanine dipeptide is found to be

Table II: Conformational Free Energy Differences for α-Amino-*n*-butyric Acid

final state	reference state	method	data set (ps)	$\Delta G°$ (kJ/mol)	$\Delta G°$ (relative)[a]
Abu dipeptide					
β conformation, χ = −60	Ala dipeptide (β)	replace	400	7.4	0.0
β conformation, χ = 180	Ala dipeptide (β)	replace	400	8.9	1.5
β conformation, χ = +60	Ala dipeptide (β)	replace	400	11.0	3.6
α_R conformation, χ = +60	β conformation χ = +60	twist	400	4.2	7.8
α_R conformation, χ = −60	Ala dipeptide (α_R)	replace	400	7.6	3.8
α_R conformation, χ = 180	Ala dipeptide (α_R)	replace	400	9.2	5.4
α_R conformation, χ = +60	Ala dipeptide (α_R)	replace	400	11.6	7.8
Abu tetrapeptide					
β conformation, χ = −60	Ala tetrapeptide	replace	400	7.0	
correction for conformational restraint				−1.5	
net $\Delta G°_{13}$ [= 7.0 + (−1.5)]				5.5	
Ala–Abu helix					
Ala–Abu helix, χ = −60	all-Ala helix	replace	400	8.1	0.0
Ala–Abu helix, χ = 180	all-Ala helix	replace	400	8.7	0.6
Ala–Abu helix, χ = +60	ala-Ala helix	replace	400	17.9	9.8
correction for conformational restraint				−1.5	
net $\Delta G°_{24}$ [= 8.1 + (−1.5)]				6.6	
$\Delta\Delta G°$ (Ala to Abu [= 6.6 − 5.5]				1.1	

[a] Relative conformer free energies for dipeptide and for helix.

equal to 5.2. The conformational free energy of the glycine residue in the coil state is then found to be lower than that of the alanine residue by 4 kJ/mol. This difference is almost the same as the difference in free energy of helix stabilization of 4.6 kJ/mol.

While it is not unreasonable to think of the difference between glycine and alanine residues in terms of an entropy difference, strictly speaking, there is not information available to determine to what extent the free energy difference of 4 kJ/mol in fact corresponds to an entropy difference. (The same thing is true of differences in conformational free energy related to conformational freedom of side chains.)

D-*Alanine*

The effect of having a single D-alanine in a right-handed α-helix otherwise consisting of L-alanine residues was computed by two different paths. The first calculation was made by reference to an all-L-alanine helix, to obtain $\Delta G°_{24}$ (−3.8). Also required was a comparison of tetrapeptides with L-alanine and D-alanine as central residues, in order to obtain $\Delta G°_{13}$ (for the coil state). Since D-alanine and L-alanine do not share a low-energy conformation in the coil state, $\Delta G°_{13}$ was calculated by subtracting the $\Delta G°_{13}$'s for D-alanine and L-alanine related to glycine, which gave 1.2. By subtracting $\Delta G°_{24}$ from $\Delta G°_{13}$, a value of 5.0 was obtained for $\Delta\Delta G°$ for D-Ala relative to L-Ala.

The second calculation was made by calculating $\Delta\Delta G°$ for D-alanine relative to glycine and adding this to the $\Delta\Delta G°$ for glycine relative to L-alanine. This second calculations gave a value of 4.9 for $\Delta\Delta G°$, i.e., essentially the same answer as the first (Table I).

α-*Amino-n-butyric Acid.*

For α-amino-*n*-butyric acid, the free energies of six conformers of the dipeptide (three side-chain conformations each with the backbone in the β and again in the α_R conformation) and of three side-chain conformations in the helix were calculated by molecular replacement starting from alanine (Table II). The resulting value of $\Delta\Delta G°$ for this residue is very close to zero.

Free dynamics were used to sample the distributions of the dihedral angles in each of these states; the distributions of the side-chain dihedral angle χ_1 in the helix, the dipeptide in the α_R conformation, and the dipeptide in the β conformation are shown (top to bottom) in Figure 2. One sees that in the dipeptide the three side-chain conformers are close to equally probable, while in the helix the side-chain conformer having χ_1 near 60° is excluded. This difference in side-chain conformational freedom between the two states corresponds to a free energy difference of circa 1 kJ/mol, by which α-amino-*n*-butyric acid should be a less good helix former than alanine (eq 10).

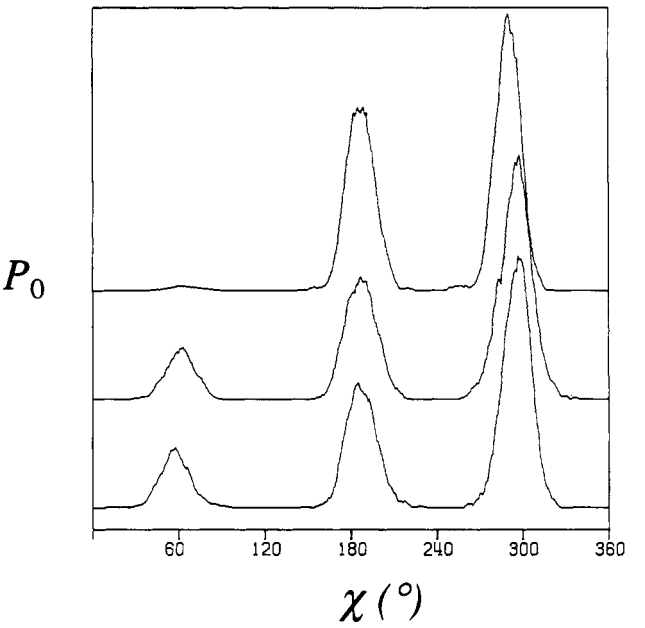


FIGURE 2: Conformational distributions of the Abu side chain in (top to bottom) the α-helix, the Abu dipeptide in the α_R conformation, and the Abu dipeptide in the β conformation (Abu = α-amino-*n*-butyric acid). Each is based on three local distributions obtained in 20 ps, scaled together using separately computed free energy differences.

t-Leucine (α-Amino-β,β'-dimethyl-n-butyric Acid) and α-Aminoisobutyric Acid

Free energy differences for t-Leu and Aib are given in Table III. For neither residue is it necessary to consider multiple side-chain conformers. The *t*-Leu residue is seen to be predicted to be a poor helix former, in agreement with subsequently obtained experimental results.

The Aib dipeptide is seen to be as stable in the α_R conformation as in the β conformation. This residue is found to be a slightly better helix former than alanine, in agreement with the experiment. The dipeptide free energy map for Aib is shown in Figure 3.

Table III: Conformational Free Energy Differences for *t*-Leucine and α-Aminoisobutyric Acid

final state	reference state	method	data set	$\Delta G°$ (kJ/mol)
t-Leu dipeptide				
α_R conformation	β conformation	twist	800	11.0
t-Leu tetrapeptide				
α_R conformation	Ala tetrapeptide	replace	800	44.4
correction for conformational restraint				0.4
net $\Delta G°_{13}$				44.8
Ala–*t*-Leu helix				
Ala–*t*-Leu helix	all-Ala helix	replace	400	52.4
$\Delta\Delta G°$ (Ala to *t*-Leu) [= 52.4 − 44.8]				7.6
Aib dipeptide				
α_R conformation	β conformation	twist	400	−0.3
Aib tetrapeptide				
α_R conformation	Ala tetrapeptide	replace	400	20.8
correction for conformational restraint				3.7
net $\Delta G°_{13}$				24.5
Ala–Aib helix				
Ala–Aib helix	all-Ala helix	replace	400	22.1
$\Delta\Delta G°$ (Ala to Aib) [= 22.1 − 24.5]				−2.4

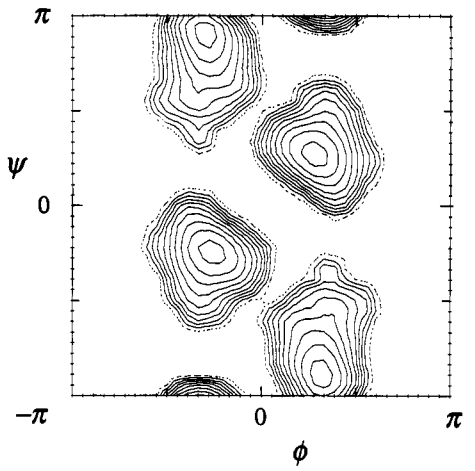


FIGURE 3: Free energy map for the backbone conformation of the Aib dipeptide (Aib = α-aminoisobutyric acid) based on two local distributions obtained in 20 ps each, scaled together using a separately computed free energy difference, and expanded by symmetry. Contours are drawn at intervals of 2 kJ/mol. The four minima are at nearly equal energy.

Table IV: Calculated and Experimental Values of $\Delta\Delta G°$ for Coil to α-Helix Conformation Change (kJ/mol)[a]

	theory	experiment[b,c,d,e]
α-aminoisobutyric acid	−2	−2[e]
L-alanine (= reference)	(0)	(0)
α-amino-*n*-butyric acid	1	0![d]
L-valine[e]	3	(0[b]) 3[c,d,e]
glycine	5	(2[b]) 4[c,d,e]
D-alanine	5	5[f]
t-leucine	8	4–5![d]
L-proline (internal)	14	(6[g]) 11–14[e,h]
L-proline (N2)[i]	6	?
L-proline (N1)	−4	?

[a] Exclamation points indicate experimental values were lacking at the time of the simulations. Question marks indicate that experimental information is still lacking. [b] Sueki et al. (1984). [c] Padmanabhan et al. (1990). [d] Lyu et al. (1990, 1991). [e] O'Neill and DeGrado (1990). [f] Fairman et al. (1992). That reference gives 4.0 and 3.2 kJ/mol for D-alanine and glycine, respectively, in the presence of urea. The values 5 (assumed by analogy) and 4 kJ/mol and other experimental values in this table cited from the work of O'Neill and DeGrado (1990) correspond to the conditions of their Table I (no urea). [g] Altmann et al. (1991). [h] Strehlow et al. (1991). [i] cf. Presta and Rose (1988) for definition of N1 and N2.

Valine and Proline

Results for valine and proline have been described in detail in earlier papers from this laboratory (Yun & Hermans, 1991; Yun et al., 1991).

Discussion

Comparison of Theory and Experiment. For the sake of comparison, calculated and experimental free energy differences have been summarized in Table IV. The experimental values obtained by different authors are not always the same, and we next summarize these differences.

As has already been mentioned, the experimental data from the Scheraga laboratory [summarized in Sueki et al. (1984) and Altmann et al. (1990)] present the smallest range of differences $\Delta\Delta G°$ (calculated as the free energies corresponding to the ratios of helix growth parameters, *s*). Several recent reports from different laboratories appear to agree rather closely on differences for valine and glycine, as well as several other residues, all differences consistently being higher than those in the former set. We consider as the "preferred" experimental data those from the work of O'Neill and DeGrado (1990). This preference is based on advantages inherent in the experimental design used by these authors. Briefly, in the experiments of O'Neill and DeGrado, the observed effects are amplified since the differences are caused by an effective double substitution, one in each of two identical chains that can associate to form a complex of two identical helices; furthermore, the experimental behavior fits closely the expected behavior of a monomer–dimer equilibrium, and thus there is no need to invoke the complex theory of helix–coil transitions. Several independent experimental studies with single helices have given numbers that agree well with those obtained by O'Neill and DeGrado with the two-helix system.

An exception to this rather general agreement is a recent study by Chakrabartty et al. (1991) which reports results that indicate a value for $\Delta\Delta G°$ for glycine (relative to alanine) of 10 kJ/mol or greater. The reason for this difference is not known; it may be the unexpected result of oversimplifications introduced by application of the theory of the helix–coil transition by Chakrabartty et al. That this may be the case is indicated by the exceptionally large helix propensity (*s* value of 1.6 to 1.8) for alanine that results from this analysis, whereas results obtained by Kemp et al. (1991) indicate an *s* value much closer to 1.0. On the other hand, the same analysis applied to substitutions of proline for alanine in another peptide (Strehlow et al., 1991) results in a $\Delta\Delta G°$ for proline that agrees quite well with O'Neill and DeGrado's value.

The calculated values agree very well with the set of values determined experimentally by O'Neill and DeGrado and the

value for D-alanine very recently obtained in the same laboratory by Fairman et al. (Table IV). In two cases not studied by O'Neill and DeGrado (Abu and *t*-Leu), the calculated values agree reasonably well with the experimental values obtained simultaneously by Lyu et al. (1991). It is useful to add that experimental values from that laboratory again agree rather well with O'Neill and DeGrado's values for all amino acids that have been studied by both groups (Lyu et al., 1990). This theoretical study very clearly does not produce the large $\Delta\Delta G°$ for glycine relative to alanine that has been found experimentally by Chakrabartty et al.

Finally, Strehlow et al. (1991) were unable to detect a high helix propensity for a proline residue at the extreme amino-terminal end of a helix, as predicted for a proline residue in position N1 (Yun et al., 1991).

Backbone and Side-Chain Statistical Factors. Differences in conformational freedom have long been identified as a potential source of differences in helix propensity. [This has been clearly summarized by O'Neill and DeGrado (1990)]. The effect is largest for glycine. The results obtained here establish that the difference in helical stability for alanine and glycine residues can be regarded as entirely due to the difference in backbone conformational freedom of these residues in the coil state and that interaction of alanine's β-methyl group with the helix makes no (net) contribution. A similar effect, but in the opposite direction, has been postulated for proline, the more greatly restrained backbone making a contribution that favors folded conformations. We have calculated the corresponding contribution to $\Delta\Delta G°$ to be −2 kJ/mol (Yun et al., 1991). This is seen to correspond to one-half of the small extent to which a proline, rather than an alanine, residue in position N1 stabilizes the helix in the model.

Smaller effects have been attributed to differences in conformational freedom of side chains in the coil and in the helix. For two amino acids (Abu and Val) for which theoretical $\Delta\Delta G°$s have been calculated in this laboratory (Table IV), the side chain's conformational freedom has been determined by calculation of complete probability distributions of the side-chain torsion angle χ_1 in the helix and in the dipeptide. The most salient aspect of the distributions for Abu (Figure 2) is the relative inaccessibility of one of the three side-chain conformers in the helix. The corresponding small loss of side-chain conformational freedom of Abu accompanying the transition from coil to helix (a free energy difference of 1 kJ/mol) is equal to the computed and experimental difference in $\Delta\Delta G°$.

The distributions for Val are given in Figure 1 of an earlier paper (Yun & Hermans, 1991). For Val two out of three side-chain conformers are inaccessible; in each of these conformers, one of the two γ-methyl groups occupies the inaccessible position identified in Abu. The correspondingly larger loss of side-chain conformational freedom of Val (a free energy difference of 2 kJ/mol) is less than the computed and experimental difference in $\Delta\Delta G°$; the remainder has been attributed to unfavorable contacts in the helix, which are also responsible for shifting the probability maximum from χ_1 = 300° to 290°, as is also observed in precisely determined crystal structures of proteins (Yun & Hermans, 1991; Richardson & Richardson, 1988a).

For the remaining amino acid residues studied here (D-Ala, *t*-Leu, Aib, and Pro in other positions than N1) considerations of conformational freedom are unimportant or irrelevant.

Unfavorable Side Chain–Helix Interactions. A prediction of low helix propensity for *t*-leucine follows easily from the observations for Abu and Val. For a residue of Abu in a helix, one of three positions is unfavorable for the single γ-methyl group; in Val the same position is unfavorable for both γ-methyl groups, which leaves a single preferred conformation in which the unfavorable position is unoccupied; in *t*-Leu one of the three γ-methyl groups unavoidably occupies the unfavorable position. The same effect is much less striking in the dipeptide, regardless of whether it has the β or the α_R conformation, which means that the bad contact is between the γ-methyl group and the other residues in the helix.

An estimate of the unfavorable interactions of the second and third γ-methyl groups of *t*-Leu can be based on the different free energies of the side-chain conformers of Abu, by adding the free energies required to bring the side chain of Abu from the most favorable into each of the two less favorable conformations. Doing this for both helix and coil (i.e., dipeptide) states (Table II) and taking the difference, one obtains 5 kJ/mol, a reasonable first estimate of the $\Delta\Delta G°$ of 8 kJ/mol that was actually calculated for *t*-Leu.

In a somewhat analogous manner, a lower limit of the magnitude of $\Delta\Delta G°$ for D-alanine is predictable from the free energy of the alanine dipeptide, which is circa 5 kJ/mol lower for the α_R conformation than for the α_L conformation (Anderson & Hermans, 1988); energetically, this conversion is equivalent to the conversion of L-Ala to D-Ala in the α_R conformation. This estimate is a lower limit of the entire difference in $\Delta\Delta G°$, since it accounts only for unfavorable interactions of the β-methyl group with the backbone of the *same* residue (as represented in the dipeptide). As this estimate is equal to the calculated $\Delta\Delta G°$, interactions within the residue must be responsible for the low helix propensity of D-Ala.

The small helix stabilization caused by replacement of Ala with Aib is echoed by a preference of the backbone of the Aib dipeptide to adopt the α_R conformation (Figure 3). As in the case of D-alanine, interactions within the Aib dipeptide determine the difference in helix stability.

Proline lowers the stability of the α-helix by two related, disrupting causes: (1) the presence of a proline residue interrupts the helical hydrogen-bond pattern and (2) the size of the group replacing the peptide hydrogen atom forces the helix into a kinked conformation (Yun et al., 1991).

Effects Not Yet Studied. Several factors that have the potential of stabilizing or destabilizing the helical conformation of peptides or, for that matter, the folded conformation of globular proteins have not been addressed by the differences in helix propensity studied in this paper. Among these are dipole–dipole, charge–dipole, and charge–charge interactions and hydrophobic interactions. It will be valuable to determine the extent, if any, to which dipole–dipole interactions determine, for example, the low helix propensity of histidine measured by O'Neill and DeGrado (1990). In order to study hydrophobic interactions, one will have to consider structures more compact than the isolated α-helix. The small magnitude of the measured $\Delta\Delta G°$ for the remaining amino acids is somewhat of an obstacle to a complete study of *all* residues that have been studied by O'Neill and DeGrado. The ability to reproduce all of these small changes with high accuracy would provide a strong validation of the model; however, new insight into the problem of helix stability is not to be expected in each and every case.

ACKNOWLEDGMENTS

Computations were done on the University's Convex C240, the North Carolina Supercomputing Center's Cray YMP, and on IBM RS/6000 workstations. We thank Dr. Neville R. Kallenbach for informing us of ongoing measurements of

$\Delta\Delta G°$ for Abu and *t*-Leu and encouraging us to provide independent theoretical estimates.

References

Altmann, K.-H., Wôjcik, M., Vâsquez, M., & Scheraga, H. A. (1990) *Biopolymers 30*, 107–120.

Anderson, A. G., & Hermans, J. (1988) *Proteins 3*, 262–265.

Berendsen, H. J. C., Postma, J. P. M., & van Gunsteren, W. F. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 43–46, Polycrystal Book Service, Western Springs, IL.

Beveridge, D. L., & DiCapua, F. M. (1989) in *Computer Simulation of Biomolecular Systems* (Beveridge, D. L., & van Gunsteren, W. F., Eds.) pp 1–26, Escom, Leiden.

Brant, D. A., & Flory, P. J. (1965) *J. Am. Chem. Soc. 87*, 2791–2800.

Chakrabartty, A., Schellman, J. A., & Baldwin, R. L. (1991) *Nature 351*, 586–588.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45.

Dang, L. X., Merz, K. M., & Kollman, P. A. (1989) *J. Am. Chem. Soc. 111*, 8505–8508.

Fairman, R., Anthony-Cahill, S. J., & DeGrado, W. F. (1992) *J. Am. Chem. Soc.* (submitted).

Hermans, J., & Anderson, A. G. (1990) in *Theoretical Biochemistry and Molecular Biophysics, A Comprehensive Survey* (Beveridge, D. L., & Lavery, R., Ed.) pp 45–61, Adenine Press, Schenectady, NY.

Hermans, J., Pathiaseril, A., & Anderson, A. G. (1988) *J. Am. Chem. Soc. 110*, 5982–5986.

Hermans, J., Yun, R. H., & Anderson, A. G. (1992) *J. Comput. Chem.* (in press).

Janin, J., Wodak, S., Levitt, M., & Maigret, B. (1978) *J. Mol. Biol. 125*, 357–386.

Kemp, D. S., Boyd, J. G., & Muendel, C. C. (1991) *Nature 352*, 451–454.

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

Lyu, P. C., Sherman, J. C., Chen, A., & Kallenbach, N. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5317–5320.

Merutka, G., Lipton, W., Shalongo, W., Park, S.-H., & Stellwagen, E. (1990) *Biochemistry 29*, 7511–7515.

Mezei, M., & Beveridge, D. L. (1986) *Ann. N.Y. Acad. Sci. 482*, 1–23.

O'Neill, K., & DeGrado, W. L. (1990) *Science 250*, 646–651.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) *Nature 344*, 268–270.

Presta, L. G., & Rose, G. D. (1988) *Science 240*, 1632–1641.

Richardson, J. S., & Richardson, D. C. (1988a) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G., Ed.) pp 1–98, Plenum Press, New York.

Richardson, J. S., & Richardson, D. C. (1988b) *Science 240*, 1648–1652.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–341.

Strehlow, K. G., Robertson, A. D., & Baldwin, R. L. (1991) *Biochemistry 30*, 5810–5814.

Sueki, M., Lee, S., Powers, S. P., Denton, J. B., Konishi, Y., & Scheraga, H. A. (1984) *Macromolecules 17*, 148–155.

Tembe, T. L., & McCammon, J. A. (1984) *Comput. Chem. 8*, 281.

van Gunsteren, W. F. (1989) in *Computer Simulation of Biomolecular Systems* (Beveridge, D. L., & van Gunsteren, W. F., Eds.) pp 27–59, Escom, Leiden.

Yun, R. H., & Hermans, J. (1991) *Protein Eng. 4*, 761–766.

Yun, R. H., Anderson, A. G., & Hermans, J. (1991) *Proteins 10*, 219–228.