(20) Platzer, K. E. B.; Ananthanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177.
(21) Konishi, Y.; van Nispen, J. W.; Davenport, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1264.
(22) Matheson, R. R., Jr.; Nemenoff, R. A.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1567.
(23) Kobayashi, Y.; Cardinaux, F.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1271.
(24) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 479.
(25) Ananthanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 417.
(26) Sueki, M.; Konishi, Y.; Scheraga, H. A. *Macromolecules*, to be submitted.
(27) Fredrickson, R. A.; Chang, M. C.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 625.
(28) Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 739.
(29) Dygert, M. K.; Taylor, G. T.; Cardinaux, F.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 794.
(30) Barskaya, T. V.; Ptitsyn, O. B. *Biopolymers* **1971**, *10*, 2181.
(31) Hill, D. J. T.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 2447.
(32) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 266.
(33) Van Nispen, J. W.; Hill, D. J.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1587.
(34) Hecht, M. H.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 545.
(35) Nagy, J. A.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1428.
(36) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 23.
(37) Alter, J. E.; Andreatta, R. H.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 564.
(38) Chang, M. C.; Fredrickson, R. A.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 633.
(39) Gō, M.; Scheraga, H. A. *Biopolymers*, to be submitted.
(40) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *246*, 2211.
(41) Fasman, G. D.; Bodenheimer, E.; Lindblow, C. *Biochemistry* **1964**, *3*, 1665.
(42) This reclassification does not affect any of the results of our previous paper[3] because there were no Tyr–Tyr interactions in the proteins considered there. However, this reclassification will be taken into account in future calculations on proteins.
(43) Applequist, J. *J. Chem. Phys.* **1963**, *38*, 934.
(44) Gō, M.; Hesselink, F. T.; Gō, N.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 459.
(45) Finkelstein, A. V.; Ptitsyn, O. B.; Kozitsyn, S. A. *Biopolymers* **1977**, *16*, 497.
(46) Sridhara, S.; Ananthanarayanan, V. S.; Fredrickson, R. A.; Zweifel, B. O.; Taylor, G. T.; Scheraga, H. A. *Biopolymers* **1981**, *20*, 1435.

Computed Conformational States of the 20 Naturally Occurring Amino Acid Residues and of the Prototype Residue α-Aminobutyric Acid[1]

Max Vásquez,[2] George Némethy, and Harold A. Scheraga*

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853.
Received December 1, 1982

ABSTRACT: Conformational energy calculations were carried out on the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids and of α-aminobutyric acid (a useful prototype for conformational studies of backbone–side chain interactions in polypeptides). Use was made of an improved version, ECEPP/2, of the computer program "empirical conformational energy program for peptides", in which some of the parameters describing interatomic energies and residue geometry have been updated. As a result of these changes, better agreement was found between the computed side-chain conformational distributions and those observed in crystalline peptides and proteins and in solutions of oligopeptides, but otherwise there is little change in the conformational properties of the various residues that were reported earlier (Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1). The Boltzmann probabilities of occurrence of various backbone and side-chain conformational states for all residues are tabulated for easy reference, and they are used to analyze trends in conformational preferences that depend on the nature of the residues.

I. Introduction

The conformational behavior of amino acid residues is of fundamental importance for the theoretical conformational analysis of oligo- and polypeptides and proteins. The properties of individual residues constitute information that can be used as the basis of the conformational studies of larger peptides, because short-range effects play an important role in the determination of the preferred conformations of polypeptides and proteins.[3-5] In this laboratory, conformational energy computations have been carried out with a set of empirical potentials that have been incorporated into a computer program named ECEPP (empirical conformational energy program for peptides).[6,7] Soon after the development of ECEPP in 1975, it was used to analyze the low-energy structures of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids,[8] denoted as "terminally blocked single residues".

The results of that study were in general agreement with experimental data and with structural information on peptides available at that time. Recently, however, a discrepancy has been found[9] between the computed distributions[8] and observed distribution in crystalline oligopeptides[9] and in proteins[10] for the dihedral angle χ^1 (describing rotation about the C^α–C^β side-chain bond) in unbranched side chains. The observed data indicate that the g^- conformational state is favored over the other two conformational states (g^+ and t), while the earlier computations indicated a preference for the t state. There was good agreement between observations and computations for the other side-chain dihedral angles and for χ^1 of side chains with β-branching. An analysis of the interatomic interactions for the computed conformations indicated that the discrepancy occurred because the repulsive interactions involving the backbone peptide N atom were too high.[9] The basis for the assignment[6] of the nonbonded interaction energy for this atom was reconsidered, and a modified set of parameters was introduced into a recent revision of ECEPP, named ECEPP/2.[11] In addition, the need arose for minor updating or modification of some other parameters in ECEPP, as a result of new experimental or theoretical

0024-9297/83/2216-1043$01.50/0

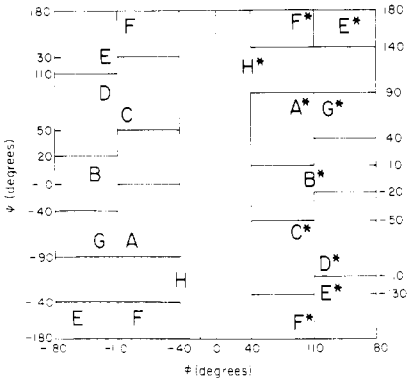

Figure 1. (ϕ,ψ) map, showing the regions used to denote backbone conformational states, as defined by Zimmerman et al.[8]

information that was not available in 1975. The additional modifications involved the nonbonded energy parameters for sulfur and aromatic carbon atoms, the parameters for hydrogen bonds involving oxygen as acceptor, and the adjustment of the length of C–H bonds to a uniform value. The changes are described elsewhere.[11] They have been incorporated into a revised version of ECEPP, ECEPP/2, that has been submitted to the Quantum Chemistry Program Exchange.[12]

While most of the revisions, with the exception of the changed nonbonded parameter for nitrogen, have very small effects on the computed conformational behavior of the terminally blocked single residues, it was advisable to reanalyze the low-energy structures of these compounds. The results of the analysis are presented here. In the course of this work, the computed distributions of various backbone and side-chain conformational states have been tabulated and are reported here. Data on these distributions were implicit in the computed results reported earlier,[8] but they were not tabulated in that paper.[8] Tables of these distributions, however, are helpful in comparisons of theoretical computations on polypeptide conformations with experimental data obtained from physicochemical and structural studies. In addition to the 20 naturally occurring amino acids, results are presented here for α-aminobutyric acid (Abu), because this amino acid serves as a useful prototype in the analysis of correlations between polypeptide backbone and side-chain conformations, for reasons discussed in section IIIE.

II. Methods

The nomenclature and conventions used follow the recommendations of the IUPAC–IUB Commission on Biochemical Nomenclature.[13] Backbone conformational regions on a (ϕ,ψ) map are denoted in terms of the letter code introduced by Zimmerman et al.[8] and summarized in Figure 1. Side-chain conformational states with respect to a threefold rotational barrier having minima near 60°, 180°, and −60° are denoted by the lower-case letters g^+, t, and g^-, respectively.[14] The ranges of χ indicated by each letter extend over a range of 120°, i.e., g^+ denotes $0° \leq \chi < 120°$, etc.

The computations were carried out with ECEPP/2, the revised version of ECEPP.[11,12] Two algorithms were used for energy minimization, the variable metric algorithm MINOP[15] and the conjugate direction Powell algorithm,[16] in order to ensure convergence. After a minimum had been reached by one method, further minimization was attempted by the other method, until there was no further variation of the energy within a cutoff of 1×10^{-5} kcal/mol.

The computational procedures followed those used and reported earlier,[8] except that the conformational entropy and free energy were not computed. On the other hand, the tables of the minimum-energy conformations include the normalized Boltzmann factor[17] v_i for each conformation, defined as

$$v_i = Q^{-1} \exp(-\Delta E_i/RT) \tag{1}$$

where

$$Q = \sum_{i=1}^{n} \exp(-\Delta E_i/RT) \tag{2}$$

ΔE_i is the relative conformational energy at the ith minimum, and n is the number of minima. Inclusion of this factor facilitates the calculation of the populations of various groups of conformations. It has been calculated for $T = 300$ K. The Boltzmann probabilities, P_J (indicating the approximate relative populations), of various backbone and side-chain conformational states J were obtained as the sum of the v_j values for all minimum-energy conformations j that belong to a particular conformational state J.[18]

The dihedral angles used in the starting conformations for energy minimization were selected as described before,[8] except for Arg. For this residue, all combinations of $(\chi^5,\chi^{6,1},\chi^{6,2})$ listed in ref 8 were used, instead of the more restricted choice described in ref 8.

III. Results and Discussion

A. Conformational Energy Contour Maps. Conformational energy contour maps for terminally blocked Ala and Gly residues, drawn with contour lines at 1 kcal/mol intervals on a (ϕ,ψ) diagram, have the same appearance as those computed by Zimmerman et al. (Figures 2 and 3 of ref 8). Therefore, they are not reproduced here. The steep boundaries of the low-energy regions coincide on the earlier and present maps. Inside the low-energy regions, slight shifts of individual contour lines occur, corresponding to the changes in the values of ΔE of some minimum-energy conformations discussed below. The only noticeable alteration occurs in the "bridge region" (region B), near $\psi = 0°$. The relative energy of this region is lowered. At the saddle point, $\Delta E = 2.14$ and 2.11 kcal/mol for Ala and Gly, respectively, while the earlier calculation[8] gave values of $\Delta E = 3.26$ and 2.99 kcal/mol, respectively.

B. Alanine. The computed minimum-energy conformations (MEC) are listed in Table I. The results generally are similar to those obtained with the earlier set of parameters (Table II of ref 8). The energies of MEC C and A are lowered by about 0.3 and 0.6 kcal/mol, respectively, relative to the other MEC, as can be seen from a comparison of the two sets of values of ΔE. As a result, the energy of MEC A now is lowered below that of D; i.e., the relative positions of these two minima are interchanged. The MEC are located in the same general regions as in the earlier results. There are only very small changes in the values of ϕ, but the absolute values of ψ are lowered considerably for MEC A, D, and A*. It is particularly notable that ψ for the MEC in the A region (α-helical conformation) is changed to −35° from the earlier value of −45°. These changes in ΔE and ψ correspond to the lowering of the energy of the region with small values of $|\psi|$, as described in section IIIA.

A new feature of the results presented here is the appearance of a MEC in the F region, where no MEC was obtained in the earlier computation[8] for Ala, although MEC in this region occurred for most other residues and for the hydrated form of Ala.[18] Thus, the results are more self-consistent for Ala and the other residues in the present calculation. The result for Ala also is in better agreement with the occurrence of conformations in the F region in

Table I
Minimum-Energy Conformations of N-Acetyl-N'-methylalaninamide[a]

conformational letter code	ΔE[b]	v	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	dihedral angles,[c] deg ϕ	ψ	χ^1
C	0.000[d]	0.507	0.000	0.000	0.000	−80	76	61
E	0.693[d]	0.159	0.811	−0.118	0.000	−155	157	59
A	0.789	0.135	1.352	−0.567	0.004	−74	−35	62
D	1.088	0.082	1.043	0.044	0.001	−151	46	61
F	1.106	0.079	1.037	0.063	0.006	−75	139	62
G	1.722	0.028	1.425	0.240	0.057	−158	−58	54
A*	2.349	0.010	1.266	0.998	0.085	54	46	67
F*	5.079	0.000	1.713	2.622	0.745	64	−175	81
C*	7.258[d]	0.000	−0.313	6.375	1.195	76	−65	88

[a] All minimum-energy conformations are listed. [b] For each minimum i, $\Delta E_i = E_i - E_0$, where $E_0 = -5.162$ kcal/mol, the energy of the lowest energy minimum. [c] Only variable dihedral angles are listed; all others have a value of 180°. [d] These conformations have a backbone-backbone hydrogen bond, defined as in ref 8, i.e., having an H···O distance < 2.3 Å.

Table II
Minimum-Energy Conformations of N-Acetyl-N'-methylglycinamide[a]

conformational letter code	ΔE[b]	v	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	dihedral angles,[c] deg ϕ	ψ
C*	0.000[d]	0.390	0.000	0.000	0.000	79	−73
C	0.000[d]	0.390	0.000	0.000	0.000	−79	73
A*	1.217	0.051	1.289	−0.073	0.000	73	34
A	1.217	0.051	1.289	−0.073	0.000	−73	−34
E	1.240[d]	0.049	0.837	0.404	0.000	180	180
D*	1.438	0.035	1.095	0.343	0.000	167	−52
D	1.438	0.035	1.095	0.343	0.000	−167	52

[a] All minimum-energy conformations are listed. [b] For each minimum i, $\Delta E_i = E_i - E_0$, where $E_0 = -6.317$ kcal/mol, the energy of the lowest energy minimum. [c] Only variable dihedral angles are listed; all others have a value of 180°. [d] These conformations have a backbone-backbone hydrogen bond, defined as in ref 8, i.e., having an H···O distance < 2.3 Å.

Table III
Minimum-Energy Conformations of N-Acetyl-N'-methylprolinamide[a]

conformational letter code	ΔE[b]	v	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	dihedral angles,[c] deg ω	ψ
trans C	0.000[d]	0.730	0.000	0.000	0.000	180	75
A	0.945[d]	0.150	0.998	−0.053	0.000	−179	−19
F	1.249	0.090	1.304	−0.055	0.000	180	160
cis A	2.024[d]	0.024	1.154	0.975	0.213	−6	−21
F	2.853	0.006	2.060	0.972	0.139	−5	162

[a] All minimum-energy conformations are listed. [b] For each minimum i, $\Delta E_i = E_i - E_0$, where $E_0 = -19.268$ kcal/mol, the energy of the lowest energy minimum. In the cis conformations, 0.318 kcal/mol has been subtracted, in order to account for the internal conformational energy of the Pro residue (see Table I of ref 11). [c] Only variable dihedral angles are listed; $\phi = -75°$; all others have a value of 180°. [d] These conformations have a backbone-backbone hydrogen bond, defined as in ref 8, i.e., having an H···O distance < 2.3 Å.

the observed (ϕ,ψ) distribution for Ala in proteins (Figure 5 of ref 4). The potential energy surface of the (ϕ,ψ) map is quite flat in the F region (see Figure 2 of ref 8); thus, very small changes in interatomic interactions are sufficient to cause the appearance or disappearance of a local minimum in this region.

C. Glycine. The computed MEC are listed in Table II. Comparison with the earlier results (Table III of ref 8) indicates trends very similar to those described above for Ala. The energies of MEC C (and C*) and A (and A*) are lowered by about 0.4 kcal/mol relative to the other MEC. Therefore, the order of the MEC is changed, so that A (and A*) is now the MEC with the second lowest energy. The changes in ϕ and ψ are similar to those seen for Ala.

D. Proline. The computed minimum-energy conformations are listed in Table III, and a plot of ΔE vs. ψ is given in Figure 2, with the peptide bond preceding the proline ring taken in the planar trans ($\omega_0 = 180°$) and cis ($\omega_0 = 0°$) conformations. The corresponding earlier results are contained in Table IV and Figure 4 of ref 8. The most noticeable change in the new results is the shifting of ψ for both the trans and cis A minima, from $\psi = -48°$ in the earlier results to $\psi = -19°$ and $-21°$, respectively, with an accompanying lowering of the energy, just as in the case of Ala and Gly. As a result, the energy of the trans A MEC is lowered below that of the trans F MEC. The computed population of the combined cis minima is 3%, as compared with the earlier computed value of 4%.

E. α-Aminobutyric Acid. This amino acid is useful in the analysis of general effects of backbone–side chain interactions, because its small side chain, consisting of a $-C^{\beta}H_2-C^{\gamma}H_3$ group, is a common feature of all amino acids larger than alanine. The conformation of the entire residue can be specified by means of only three dihedral angles, viz., ϕ, ψ, and χ^1, not counting the trivial rotation of the terminal methyl group, specified by χ^2. Thus, it is well suited for the study of general correlations[9,19] between χ^1 and (ϕ,ψ), in the absence of specific effects due to interactions between the backbone and atoms of longer side chains that depend on the values of χ^j ($j > 1$). Therefore, Abu is a general prototype of amino acids with side chains larger than that of alanine.[9,10,19-21]

The computed MEC with $\Delta E < 5.0$ kcal/mol are listed in Table IV. In any given backbone conformational region

Table IV
Minimum-Energy Conformations of *N*-Acetyl-*N'*-methyl-α-aminobutyrylamide

conformational letter code	ΔE[b]	υ	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	dihedral angles,[c] deg ϕ	ψ	χ^1	χ^2
C g^-	0.000[d]	0.348	0.000	0.000	0.000	−81	75	−65	61
C t	0.230[d]	0.237	0.022	0.089	0.119	−79	81	−171	58
A t	0.809	0.090	1.382	−0.613	0.040	−72	−41	−175	56
E g^+	0.877[d]	0.080	0.790	0.130	−0.044	−156	160	60	60
A g^-	0.969	0.069	1.364	−0.408	0.014	−77	−35	−65	61
E t	1.158	0.050	0.732	0.379	0.048	−152	138	−175	55
F g^-	1.278	0.041	1.053	0.224	0.001	−81	142	−65	61
G t	1.574	0.025	1.366	0.207	0.001	−156	−59	−178	55
D g^+	1.708	0.020	1.057	0.612	0.039	−151	38	55	56
E g^-	1.809	0.017	0.920	0.863	0.025	−129	152	−64	64
A* g^-	1.964	0.013	1.290	0.693	−0.020	56	47	−59	64
A g^+	2.598[d]	0.004	1.145	0.976	0.477	−73	−26	75	67
A* t	2.745	0.003	1.425	0.750	0.569	58	55	−161	61
F g^+	2.757	0.003	1.224	1.045	0.488	−70	152	75	68

[a] All minimum-energy conformations with $\Delta E < 5$ kcal/mol are listed. There are, however, no MEC's with ΔE between 2.8 and 5.0 kcal/mol. [b] For each minimum i, $\Delta E_i = E_i - E_0$, where $E_0 = -5.448$ kcal/mol, the energy of the lowest energy minimum. [c] Only variable dihedral angles are listed; all others have a value of 180°. [d] These conformations have a backbone-backbone hydrogen bond, defined as in ref 8, i.e., having an H···O distance < 2.3 Å.

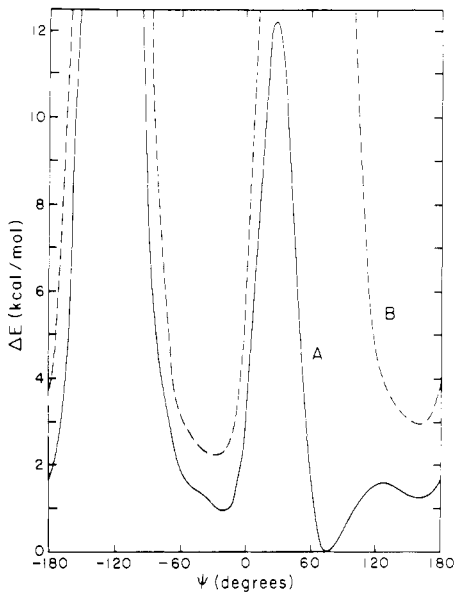

Figure 2. Curve of ΔE vs. ψ for *N*-acetyl-*N'*-methylprolineamide, with the peptide bond preceding the pyrrolidine ring taken (A) in the trans (solid line) and (B) cis (dashed line) conformations. The peptide group following the pyrrolidine ring was kept in the trans conformation.

J, usually there are only one or two low-energy MEC, with the exception of regions E and A, where there are three. In most regions, strong repulsive interactions occur between the backbone and the side-chain $C^\gamma H_3$ group in one or two of the three possible side-chain rotamers. The rotamers with low energy are different in various backbone conformational regions. In most of the low-energy MEC, the values of (ϕ,ψ) differ from those for Ala by at most a few degrees (cf. Tables I and IV). The side chain is nearly staggered in the low-energy MEC: χ^1 deviates from the values for full staggering (viz., $\chi^1 = 60°$, 180°, and −60°) by <10° in all minima with $\Delta E < 3.0$ kcal/mol. These results indicate that the side chain can be accommodated with no repulsive overlaps in at least one rotamer for each backbone conformational state. In these MEC, the interactions of the $C^\gamma H_3$ group with the backbone causes only small changes (usually of the order of 0.2 kcal/mol or less) in the relative energies, ΔE, as compared with similar backbone conformations of Ala. As a result, the relative populations of the various backbone conformational states for Ala and Abu generally are very similar. This can be seen by comparing the first two lines of Table V. Larger deviations in the populations of some conformational states of the other amino acids from those of Ala or Abu (Table V) must therefore be attributed to specific interactions involving side-chain atoms beyond the $C^\gamma H_2$ group or to the effect of side-chain branching.

The distribution of the three side-chain rotamers (summed over all backbone conformational states) is g^+:t:g^- = 11%:40%:49%; i.e., the g^- conformational state is preferred over the other two, in agreement with the observed distributions of side-chain conformations of unbranched amino acids in peptides[9] and proteins.[10,22]

In our earlier study,[8] no computations were carried out for Abu. The present results can be compared, however, with another investigation in which the old, 1975 version of ECEPP was used.[21] The comparison is approximate, because the energies of various conformations were calculated in ref 21 at 10° intervals of ϕ and ψ, instead of using energy minimization. Nevertheless, the trends in the backbone dihedral angles and the energies of the various backbone conformational states parallel closely those described above for Ala (section IIIB). In terms of side-chain rotamers, the introduction of the revised ECEPP/2 parameters resulted in a marked increase of the population of the g^- conformation (from 28% to 49%), at the expense of the t conformation.

F. Other Residues. For each of the 20 naturally occurring amino acid residues and for Abu, tables similar to Tables I–IV, containing minima with $\Delta E < 5.0$ kcal/mol ($\Delta E < 3.0$ kcal/mol in the case of Lys and Arg) are given in the supplementary material.[23] For each MEC, the tables give (a) the conformational letter code for the backbone and χ^1, (b) the number of hydrogen bonds, (c) the relative energy ΔE_i, (d) the normalized Boltzmann factor υ_i, (e) the electrostatic, nonbonded, and torsional components of the energy, ΔE_{ES}, ΔE_{NB}, and ΔE_{TOR}, and (f) the dihedral angles.

The results shown in those tables are summarized here in Tables V and VI, expressed as Boltzmann probabilities of the various backbone and side-chain conformational states. Table V also includes the number of computed minima, for comparison with Table V of ref 8.

The relative population of various backbone conformational states is shown in Table V for each residue.[24] Generally, the most favored state is C. This state has the highest Boltzmann probability in the case of 16 of the 22

Table V
Summary of the Results of Energy Minimization for the 20 Naturally Occurring Amino Acids and for Abu: Number of Low-Energy Local Minima and the Relative Populations of Various Backbone Conformational States[a]

Residue	No. of minima with		Boltzmann Probability, P_J						
	ΔE<3 kcal/mol	3≤ΔE<5 kcal/mol	A	C	D	E	F	G	A*
Abu	14	0	0.163	0.585*[b]	0.020	0.147	0.044	0.025	0.016
Ala	7	0	0.135	0.507*	0.082	0.159	0.079	0.028	0.010
Arg	210	c	0.296	0.459*	0.008	0.194	0.023	0.014	0.007
Asn	39	10	0.196	0.287	0.032	0.385*	0.076	0.018	0.006
Asp	35	8	0.289*	0.223	0.043	0.283	0.140	0.015	0.006
CysH	37	6	0.159	0.592*	0.019	0.187	0.011	0.021	0.011
Gln	76	27	0.236	0.383	0.011	0.236*	0.092	0.024	0.017
Glu	74	30	0.184	0.445*	0.015	0.206	0.112	0.019	0.018
Gly[d]	7	0	0.051	0.390*	0.035	0.049	-	-	0.051
His(N^{δ})[e]	43	5	0.169	0.300*	0.126	0.262	0.094	0.039	0.010
His(N^{ϵ})[e]	37	5	0.277*	0.319	0.057	0.204	0.095	0.039	0.009
Ile	14	7	0.173	0.111	0.098	0.581*	0.036	-	0.001
Leu	17	10	0.215	0.619*	0.007	0.083	0.001	0.057	0.019
Lys	178	c	0.170	0.559*	0.034	0.159	0.044	0.022	0.012
Met	53	34	0.239	0.431*	0.018	0.198	0.077	0.026	0.010
Phe	14	4	0.087	0.077	0.059	0.682*	0.072	0.021	0.002
Pro[f]	5	0	0.150	0.730*	-	-	0.090	-	-
Ser[g]	9	35	0.010	0.689*	0.087	0.150	0.003	0.043	0.015
Thr	15	11	0.100	0.785*	0.013	0.074	0.011	0.014	0.003
Trp	30	4	0.166	0.160	0.070	0.463*	0.110	0.027	0.004
Tyr	26	4	0.066	0.060	0.079	0.737*	0.039	0.018	0.001
Val	9	0	0.285	0.356*	0.034	0.301	0.021	-	0.003

[a] Expressed as the Boltzmann probability P_J for conformational state J, where J = A, C, D, E, F, G, or A*. [b] For each residue, an asterisk indicates the conformational state containing the conformation of lowest energy. [c] Only the minima with $\Delta E < 3$ kcal/mol are listed in the tables of the supplementary material. [d] Gly has minima in the A*, C*, and D* regions that have the same Boltzmann probability as those listed in the table for the corresponding regions A, C, and D, respectively. [e] The two forms of the His residue listed have a hydrogen attached to the N^{δ} or the N^{ϵ}, respectively, of the neutral imidazole ring. [f] The Boltzmann probabilities listed in the table are those for the trans conformation of the peptide bond preceding the Pro residue. In the cis conformation, the Boltzmann probabilities are 0.024 for the minimum in the A region and 0.006 for the minimum in the F region. [g] Ser has an additional minimum in the F* region, with a Boltzmann probability of 0.003.

residues listed, and it contains the lowest energy conformation for 14 of the residues. The only systematic deviation from this general finding occurs for the three aromatic residues (Phe, Trp, and Tyr), for which state E is the most stable, presumably because it is easier to accommodate the bulky inflexible side chain on a nearly extended peptide backbone. The strong preference of Ile for the E over the C state may be attributed to a similar reason. Generally, the E state has the second highest population, followed by the A state. The populations of the other four states (D, F, G, and A*) generally are much lower than those of the three mentioned.

In the earlier computation, similar trends were found in the populations of the various states.[8] The effect of changing to the new set of parameters is generally an increase of the population of state A for most residues (with an average increase of about 9% in the Boltzmann probability for this state), with a corresponding general decrease of the populations of the C and E states, while the changes for the other states are small. These shifts are a result of the change of the potential surface in and near the bridge region, as discussed in section IIIA.

The relative populations of the various side-chain rotamers around each single bond with a threefold rotational barrier are shown in Table VI for each residue. In the distributions of χ^1, the g^- rotamer has the highest probability for unbranched side chains (Abu, Arg, Cys, Gln, Glu, Lys, and Met) while the g^+ rotamer is least favored, in agreement with observed distributions.[9,10,22] The low probability of the g^+ rotamer is due to repulsive interactions between the $C^{\gamma}H_2$ group (or $C^{\gamma}H_3$ in Abu) with the NH and C′O groups of the same residue, because the C^{γ} atom is in the gauche position with respect to both the N and the C′ atoms.[9,22] In the other two rotamers, the C^{γ} is gauche to one heavy atom and the H^{α}, and hence the repulsive interactions are decreased. The relative probabilities of the g^- and t rotamers do not result from any unique interaction but from the overall balance of the favorable and unfavorable interactions involving the atoms of the $C^{\gamma}H_2$ group and atoms of the neighboring peptide groups.

The preference of Ser for the g^+ rotamer, observed[9] and found in the computations, is due to the presence of a strong side chain–backbone hydrogen bond in many low-energy conformations.[18] The preference of Asn and Asp for the t rotamer and of His for the g^- rotamer also is due to hydrogen bonding in some low-energy conformations, while the somewhat lower probability of the g^- rotamer in

Table VI
Summary of the Results of Energy Minimization for the 20 Naturally Occurring Amino Acids and for Abu: Relative Populations of Various Side-Chain Rotamers[a]

	Boltzmann Probability,[b] P_J											
	χ^1			χ^2			χ^3			χ^4		
Residue	g^+	t	g^-	g^+	t	g^-	g^+	t	g^-	g^+	t	g^-
Abu	0.107	0.405	0.488									
Arg	0.101	0.332	0.567	0.012	0.970	0.018	0.251	0.497	0.252	0.350	0.310	0.340
Asn	0.300	0.417	0.283									
Asp	0.535	0.258	0.207									
CysH	0.159	0.389	0.452	0.306	0.398	0.296						
Gln	0.085	0.457	0.458	0.321	0.538	0.141						
Glu	0.073	0.444	0.483	0.384	0.472	0.144						
His(N^δ)[c]	0.417	0.374	0.209									
His(N^ϵ)[c]	0.513	0.258	0.229									
Ile	0.248[d]	0.527[d]	0.225[d]	0.094	0.894	0.012						
Leu	0.020	0.572	0.408	0.615[e]	0.381[e]	0.004[e]						
Lys	0.119	0.410	0.471	0.008	0.982	0.011	0.033	0.936	0.031	0.214	0.580	0.206
Met	0.079	0.377	0.544	0.070	0.840	0.090	0.118	0.762	0.120			
Phe	0.346	0.365	0.289									
Ser	0.714	0.091	0.195	0.874	0.016	0.110						
Thr	0.807[d]	0.055[d]	0.138[d]	0.771	0.206	0.023						
Trp	0.433	0.354	0.213									
Tyr	0.357	0.336	0.307									
Val	0.286[e]	0.610[e]	0.104[e]									

[a] Expressed as the Boltzmann probability P_J for rotamer J around the bond j (specified by χ^j). Each entry is the sum of the Boltzmann factors v_i for all conformations that belong to a given χ^j rotamer (i.e., summing the v_i over all backbone conformational states and all side-chain rotamers around bonds other than bond j). [b] Listed only for bonds with a threefold rotational barrier. [c] The two forms of the His residue have a hydrogen attached to the N^δ or the N^ϵ, respectively, of the neutral imidazole ring. [d] The dihedral angle χ^1 is defined in Ile by the position of the longer side-chain branch and in Thr by the O^γ atom.[13,25] Consequently, $\chi^1 \approx 60°$ corresponds to a rotamer in which the two branches of the side chain are in positions g^- and g^+; $\chi^1 \approx 180°$ to g^+ and t; $\chi^1 \approx -60°$ to t and g^-. [e] For χ^2 of Leu and χ^1 of Val, $\chi \approx 60°$ corresponds to the two terminal methyl groups being in positions g^+ and t, $\chi \approx 180°$ to t and g^-, and $\chi \approx -60°$ to g^- and g^+.

Leu, Phe, Trp, and Tyr is due to the balance of many small interactions involving the rigid aromatic rings and the backbone.

In the three residues with branching on the C^β atom (Ile, Thr, and Val),[25] the presence of two bulky side-chain branches increases repulsive interactions in some of the rotamers for χ^1, and hence the populations are very sensitive to details of the side chain–backbone interactions such as the substitution in Thr of an $O^\gamma H$ for $C^\gamma H_2$ (or $C\gamma H_3$)[18,22] or the presence of a side chain–backbone hydrogen bond in Thr.[18] It is notable, in comparing Val and Ile, that the addition of a $C^\delta H_3$ group can cause large differences in the probability of the various χ^1 rotamers, as well as of backbone conformational states (Table V). This difference in the interactions involving Val and Ile side chains can have large effects on the stability of regular polypeptide structures such as α-helices and β-sheets.[26,27] The computed distribution of the χ^1 rotamers for these two residues agrees well with the observed distribution in peptides.[9]

For the other side-chain dihedral angles (χ^j, j = 2–4) with a threefold rotational barrier, the t rotamer is strongly favored (Table VI), in agreement with observed distributions.[9,10] This is due to the same reason as the preference for the t rotamer in alkyl chains in general, viz., the balance of the interaction energies between the substituents on the atoms next to the bond about which rotation takes place. The strong preference for the t rotamer is lowered in the presence of γ-branching (χ^2 of Leu) or when there is a planar polar substituent group separated by one bond from the bond about which rotation takes place (χ^4 of Arg, χ^2 of Gln and Glu), because of repulsive interactions. The distribution is different for the rotation of OH groups as well (χ^2 of Ser and Thr), because steric effects are not important in this case and hydrogen bonding dominates the conformational preferences.

In both Asp and Glu, a conformation of the carboxyl group is overwhelmingly preferred in which the O–H group is syn to the C=O bond (i.e., $\chi^{3,2}$ for Asp and $\chi^{4,2}$ for Glu $\approx$180°, respectively). This result agrees with recent ab initio calculations on carboxylic acids.[28,29]

In the computations using the older ECEPP parameters,[8] generally there was a lower probability for the g^- rotamer and a higher probability of the t rotamer for χ^1, resulting in poor agreement with observed distributions in crystals.[9,10] In fact, this discrepancy was one of the reasons for reexamining the basis for the choice (and hence modification) of some of the ECEPP parameters,[9,11,12] as discussed in section I. For χ^j (j = 2–4), the rotamer distributions were changed very little as a result of the modifications in ECEPP/2, with one exception. The change in the nonbonded energy parameters for sulfur caused a

strong increase in the probability of the t rotamer for χ^2 in Met, bringing it close to the observed distribution for amino acids with unbranched side chains.[9,11]

Very little experimental information is available on preferential side-chain conformations of peptides in solution. Proton nuclear magnetic resonance measurements on terminally blocked amino acids and oligopeptides, involving mostly aromatic residues, indicate a preference for the g^- rotamer around the C^α–C^β bond,[30-37] in agreement with the general conclusions presented here. Interactions with the solvent may modify to some extent the numerical values of the relative populations of the various rotamers, but it is unlikely that they cause large changes in their relative preferences. The relative insensitivity of averaged rotameric preferences to the environment of the residue has been seen in comparisons of side-chain conformational distributions in peptide crystals and in globular proteins.[9,10] The solvent may cause larger changes in conformational distributions of polar side chains that may be involved in hydrogen bonding.[18]

IV. Conclusions

The updating to ECEPP/2 parameters generally does not cause much change in the computed backbone conformational properties of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids, and hence the conclusions presented in an earlier study[8] remain valid. The change in the nonbonded energy parameters for nitrogen caused some systematic changes in the preferred conformations. In agreement with observed distributions, the g^- rotamer for χ^1 is computed to be favored over the other two, while the g^+ rotamer is least favored, in general. In terms of backbone (ϕ,ψ) conformational states, the revision of the parameters increased somewhat the probability of the A state for most residues, at the expense of the C and E states. The revision of the nonbonded energy parameters for sulfur improved the agreement between computed and observed side-chain distributions for Met.

The conformational properties of Abu parallel the general trends seen for the other amino acids. Thus, the utility of this residue as a prototype in general investigations of conformational properties is established.

Acknowledgment. We thank M. S. Pottle and S. Rumsey for computing help and for helpful discussions and Drs. E. Benedetti and S. S. Zimmerman for useful discussions.

Registry No. Ac-Ala-NHMe, 19701-83-8; Ac-Gly-NHMe, 7606-79-3; Ac-Pro-NHMe, 19701-85-0; Ac-Abu-NHMe, 35023-27-9; Ac-Arg-NHMe, 34276-26-1; Ac-Asn-NHMe, 33878-41-0; Ac-Asp-NHMe, 33067-37-7; Ac-Cys-NHMe, 10061-65-1; Ac-Gln-NHMe, 33878-42-1; Ac-Glu-NHMe, 33878-44-3; Ac-His-NHMe, 6367-11-9; Ac-Ile-NHMe, 32483-16-2; Ac-Leu-NHMe, 32483-15-1; Ac-Lys-NHMe, 6367-10-8; Ac-Met-NHMe, 29744-03-4; Ac-Phe-NHMe, 17186-60-6; Ac-Ser-NHMe, 6367-12-0; Ac-Thr-NHMe, 19746-35-1; Ac-Trp-NHMe, 6367-17-5; Ac-Tyr-NHMe, 6367-14-2; Ac-Val-NHMe, 19701-84-9; Abu, 80-60-4; Ala, 56-41-7; Arg, 74-79-3; Asn, 70-47-3; Asp, 56-84-8; CysH, 52-90-4; Gln, 56-85-9; Glu, 56-86-0; Gly, 56-40-6; His, 71-00-1; Ile, 73-32-5; Leu, 61-90-5; Lys, 56-87-1; Met, 63-68-3; Phe, 63-91-2; Pro, 147-85-3; Ser, 56-45-1; Thr, 72-19-5; Trp, 73-22-3; Tyr, 60-18-4; Val, 72-18-4.

Supplementary Material Available: Tables 1–22 containing data for the low-energy minima of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids and Abu (41 pages). Ordering information is given on any current masthead page.

References and Notes

(1) This work was supported by research grants from the National Science Foundation (PCM79-20279 and PCM79-18336), from the National Institute of General Medical Sciences (GM-14312 and GM-25138) and the National Institute on Aging (AG-00322) of the National Institutes of Health, U.S. Public Health Service, and from the Mobil Foundation.
(2) Recipient of a fellowship from the Universidad de Costa Rica and of a Sage Graduate Fellowship from Cornell University.
(3) Scheraga, H. A. *Pure Appl. Chem.* **1973**, *36*, 1.
(4) Némethy, G.; Scheraga, H. A. *Q. Rev. Biophys.* **1977**, *10*, 239.
(5) Scheraga, H. A. *Biopolymers* **1983**, *22*, 1.
(6) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
(7) The original version of ECEPP is Program No. QCPE 286. This version is obsolete and should not be used. See ref 12 for revised version.
(8) Zimmerman, S. S.; Pottle, M. S.; Némethy, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1.
(9) Benedetti, E.; Morelli, G.; Némethy, G.; Scheraga, H. A. *Int. J. Pept. Protein Res.*, in press.
(10) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357.
(11) Némethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883.
(12) The ECEPP/2 computer program, its description, and all associated geometric and energy parameters are available on magnetic tape as Program QCPE No. 454 from the Quantum Chemistry Program Exchange, Chemistry Department, Room 204, Indiana University, Bloomington, IN 47401.
(13) IUPAC–IUB Commission on Biochemical Nomenclature *Biochemistry* **1970**, *9*, 3471.
(14) Flory, P. J. "Statistical Mechanics of Chain Molecules"; Interscience: New York, 1969; p 56.
(15) Dennis, J. E.; Mei, H. H. W. Technical Report No. 75-246, Department of Computer Science, Cornell University, Ithaca, NY.
(16) Powell, M. J. D. *Comput. J.* **1964**, *7*, 155.
(17) Zimmerman, S. S.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 811.
(18) Hodes, Z. I.; Némethy, G.; Scheraga, H. A. *Biopolymers* **1979**, *18*, 1565.
(19) Némethy, G.; Scheraga, H. A. *Biopolymers* **1982**, *21*, 1535.
(20) Ponnuswamy, P. K.; Sasisekharan, V. *Biopolymers* **1971**, *10*, 565.
(21) Paterson, Y.; Leach, S. J. *Macromolecules* **1978**, *11*, 409.
(22) Bhat, T. N.; Sasisekharan, V.; Vijayan, M. *Int. J. Pept. Protein Res.* **1979**, *13*, 170.
(23) See paragraph at end of paper regarding supplementary material.
(24) No such comparative summary table was presented in ref 8. Data from ref 8, in the format of the present Table V, are contained in Table VI of ref 18.
(25) It should be noted that the same relative orientation of the two branches on the C^β atom corresponds to different values of χ^1 for Ile and Thr, on the one hand, and for Val, on the other hand, because χ^1 is defined with respect to differently placed branches in the IUB–IUPAC convention.[13] Thus, for example, when the two branches are in the g^- and t positions (with respect to the N–C^α bond), this corresponds to $\chi^1 \approx 180°$ in Val but to $\chi^1 \approx -60°$ in Ile and Thr.
(26) Gō, M.; Scheraga, H. A. *Biopolymers*, in preparation.
(27) Chou, K.-C.; Némethy, G.; Scheraga, H. A. *J. Mol. Biol.*, in press.
(28) Petersen, M. R.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1979**, *101*, 1076.
(29) Schafer, L.; van Alsenoy, C.; Scarsdale, J. N.; Sellers, H. L.; Pinegar, J. F. *J. Mol. Struct.* **1982**, *86*, 267.
(30) Bundi, A.; Grathwohl, C.; Hochmann, J.; Keller, R. M.; Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1975**, *18*, 191.
(31) Wüthrich, K.; de Marco, A. *Helv. Chim Acta* **1976**, *59*, 2228.
(32) Kobayashi, J.; Nagai, U. *Biopolymers* **1978**, *17*, 2265.
(33) Niccolai, N.; Tiezzi, E. *J. Phys. Chem.* **1979**, *83*, 3249.
(34) Skrabal, P.; Rizzo, V.; Baici, A.; Bangerter, F.; Luisi, P. L. *Biopolymers* **1979**, *18*, 995.
(35) Baici, A.; Rizzo, V.; Skrabal, P.; Luisi, P. L. *J. Am. Chem. Soc.* **1979**, *101*, 5170.
(36) Jäckle, H.; Luisi, P. L. *Biopolymers* **1981**, *20*, 65.
(37) Inoue, Y.; Okuda, T.; Miyata, Y. *J. Am. Chem. Soc.* **1981**, *103*, 7393.