Macromolecules

Volume 10, Number 1 January–February 1977

Conformational Analysis of the 20 Naturally Occurring Amino Acid Residues Using ECEPP[1]

S. Scott Zimmerman,[2] Marcia S. Pottle, George Némethy, and Harold A. Scheraga*

Department of Chemistry, Cornell University, Ithaca, New York 14853.
Received April 14, 1976

ABSTRACT: Conformational energy calculations using ECEPP (Empirical Conformational Energy Program for Peptides) were carried out on the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids. Minimum-energy conformations were located, and the relative conformational energy, librational entropy, and free energy at each minimum were calculated. The effects of intrinsic torsional potentials, intramolecular hydrogen bonds, and librational entropy on relative conformational energies and locations of minima are discussed. The results are categorized most easily by use of a new conformational letter code that is introduced here.

I. Introduction

Over the past several years, a set of empirical potentials and parameters has been developed to calculate the relative conformational energies of peptides, polypeptides, and proteins.[3,4] These parameters were obtained by considering first the *inter*molecular interactions in x-ray crystal structures of model compounds;[4] they were then applied to the analysis of the 20 naturally occurring amino acid residues.[5] The results[5] were in general agreement with experiment[6-9] and encouraged continued refinement of the energy parameters from experimental data on *intra*molecular interactions.[10] The final set of geometric parameters, partial atomic charges, nonbonded interactions, and intrinsic torsional potentials was presented and discussed in ref 10. The computer program which utilizes these potentials, parameters, and geometries to calculate conformational energies of peptides is known as ECEPP (Empirical Conformational Energy Program for Peptides).[11]

Because of the refinements made in the potentials and parameters since the earlier report on single residues by Lewis et al.,[5] and because of the importance of the conformational character of individual amino acid residues in studies of larger peptides, we have analyzed the low-energy structures of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids using ECEPP.[10,11] In addition, we have developed an effective method of searching conformational space for low-energy minima. At each energy minimum, we have calculated the conformational energy, librational entropy, and conformational free energy, assigned a conformational letter code, and checked for the presence of hydrogen bonds.

Although this study does not change the general conclusions of Lewis et al.,[5] it does furnish improvements in the detailed understanding of the conformational properties of single residues.

II. Methods and Definitions

A. Nomenclature. The nomenclature and conventions used here are those adopted by an IUPAC-IUB Commission.[12]

B. Potentials and Energy Parameters. Conformational energy calculations were carried out using ECEPP,[11] which is based on the empirical potential energy functions and energy parameters described by Momany et al.[10] The total conformational energy E is the sum of the electrostatic energy E_{ES}, nonbonded energy E_{NB}, and torsional energy E_{TOR}.[10] The hydrogen bond energy E_{HB} is included in the nonbonded energy component.[10,11] The total energy and the energy components at the ith local minimum are given as ΔE_i, $\Delta E_{ES,i}$, $\Delta E_{NB,i}$, and $\Delta E_{TOR,i}$, where $\Delta E_i = E_i - E_0$, $\Delta E_{ES,i} = E_{ES,i} - E_{ES,0}$, etc., and where the zero subscript refers to the energy at the global minimum.

There are two major differences between the energy parameters[10] used in the present study and those[4] used in the earlier work of Lewis et al.[5] The first is the addition of explicit torsional potentials for aliphatic C–C bonds and C–N single bonds, increasing the barriers to rotation about these bonds from only 0.3 and 0.2 kcal/mol to 3.0 and 2.0 kcal/mol, respectively (see ref 10 for full discussion).

The second important improvement in the present set of parameters is the softening[13] of the repulsive interaction between an amide nitrogen and the amide hydrogen in any other peptide group, in order to lower the energy in the "bridge region" (around $\phi = -90°$, $\psi = 0°$). This lower energy is required since many residues in proteins have conformations in the bridge region.[14-17] For example, Pohl[14] plotted the (ϕ, ψ) values of the residues in crystals of five proteins and observed a relatively high density of conformations around $\psi = 0°$. Karle[15] determined the crystal structure of the Li^+ complex of antamanide, a cyclic decapeptide, and found that four of the ten residues are in the region near $\psi = 0°$. In *N*-acetyl-*N'*-methyl-L-tyrosineamide crystals (ϕ, ψ) = (−80°, −3°),[16] and in *N*-acetyl-*N'*-methyl-L-prolineamide crystals, (ϕ, ψ) = (−76°, −16°);[17] in these crystals, the distance between the hydrogen of the amide end group and the nitrogen of the

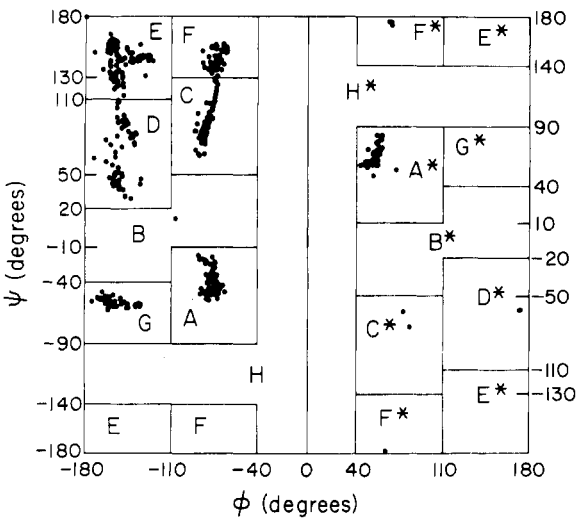

Figure 1. ϕ–ψ map showing the regions defining the conformational letter code. The low-energy ($\Delta E \leq 5$ kcal/mol) minima obtained here for the 20 naturally occurring amino acid residues are plotted on the map. A total of 969 low-energy minima are plotted.

amino acid residue is only 2.45 and 2.40 Å, respectively. Although such conformations may be stabilized by long-range or intermolecular interactions, their occurrence would preclude the strong repulsion of 5–10 kcal/mol between the two atoms used by Lewis et al.[5] and favor the lower energy (2–3 kcal/mol) used in this work.

C. Geometric Parameters. The standard residue and end group geometry of ECEPP[10,11] was used in all cases. All residues were taken in the L configuration. The proline residue was taken with the pyrrolidine ring in the "puckered down" conformation[10] since it is lower in energy than the "puckered up" conformation.[18] The locations and relative energies of minima for *N*-acetyl-*N'*-methyl-prolineamide are essentially the same for both ring structures. All calculations were carried out with ionizable residues in their uncharged state. The histidine residue was examined in both possible uncharged forms, i.e., with a proton on either the N^{δ} or N^{ϵ} atom of the imidazole ring.

Because of the publication of more recent experimental data on the structures of peptides, important changes in the geometric parameters have been made since the 1968 review by Scheraga[19] and several minor changes since the work of Lewis et al.[5] These changes were discussed by Momany et al.[10] and included in ECEPP.

D. Conformational Energy Contour Diagrams. The total conformational energy was calculated at 10° intervals (i.e., at 36 × 36 = 1296 points) over the ϕ–ψ space of Ala and Gly, from which energy contour plots were drawn. The value of χ^1 of Ala was held fixed at 60°. In both Ala and Gly, the torsional angles of peptide bonds and of the two end methyl groups were held fixed at 180°. The conformational energy of proline was determined at 5° intervals from $\psi = -180°$ to $\psi = +180°$, with the peptide bond preceding proline taken in both the trans and cis conformations; from these data, curves of ΔE vs. ψ were drawn.

E. Conformational Letter Code. Conformations are classified in terms of the regions of the ϕ–ψ map in which they occur. These regions are identified here by a convenient letter code which is helpful for the discussion and comparison of low-energy structures when interest is centered on the general features of the conformation and not on the exact values of the dihedral angles.

The entire ϕ–ψ map is subdivided into regions denoted by capital letters, as shown in Figure 1. On the left-hand half of the map ($\phi \leq 0°$), six of the regions (A, C to G) comprise the distinct ϕ–ψ areas in which energy minima are found for various amino acid residues, while one region (B) is defined around the moderate-energy bridge region. On the right-hand half of the map ($\phi > 0°$), regions are defined by inversion of the left-hand half around the center of the map, i.e, around the point (0°, 0°), and an asterisk is appended to the letters.

In defining the letter code, the boundaries of the regions were selected so that all related minima would fall within the same region. The boundary lines run along fixed values of either ϕ or ψ so as to simplify the calculation of code assignments. On the left-hand side of the map, boundary lines to the left and below each region are included within the region (the opposite is the case for the right-hand side). For example, region A is defined to have $-110° \leq \phi < -40°$ and $-90° \leq \psi < -10°$.

The letters denoting each region were chosen so that they are easy to remember: A denotes the region which contains the right-handed α-helical conformation, B is the bridge region, C contains the C_7^{eq} ring (a seven-membered hydrogen-bonded ring, with the side chain in the equatorial position), E contains the extended conformations (e.g., C_5, a five-membered hydrogen-bonded ring), and H is the high-energy region. D, F, and G were assigned to the remaining regions to indicate contiguity. Regions A, C, D, E, and G contain the five lowest-energy minima of alanine (see Results). No minimum of alanine occurs in region F, although the energy there is low, but several other amino acids have minima in that region, as indicated in Figure 1.

Side-chain conformations can be classified in a similar fashion. The orientations near the staggered conformations for side-chain rotations with threefold minima are coded by single letters. Following the established nomenclature for polymer chains,[20] χ values in the range of 180 ± 30° are denoted as t, those in the range 60 ± 30° as g^+, and those in the range −60 ± 30° as g^-.

F. Minimization of the Total Conformational Energy. One of the main purposes of this study has been to locate all low-energy minima (i.e., minima with $\Delta E < 5$ kcal/mol) of the 20 naturally occurring amino acid residues. The earlier method[5] approached this problem by (a) selecting a relatively large number of structures (at least 21 backbone conformations in combination with at least 45 side-chain conformations for residues with two or more side-chain single bonds about which rotation can take place) and then (b) rejecting many of those structures as starting points based on their relative energies (before minimization).[5] Thus, a preliminary list of about 1200 conformations for each residue was reduced to 25–100 starting conformations. In this study, we have chosen an alternative approach which involves (a) selecting fewer structures and (b) minimizing the energy of each of these conformations. Our list of starting points includes from 81 to 1600 conformations for each residue with more than one variable side-chain dihedral angle.

The details of our method for selecting these starting conformations are given below.

1. Selection of Backbone Dihedral Angles. Backbone starting conformations for the *N*-acetyl-*N'*-methylamides of Ala, Gly, and Pro were selected from the regions of low energy in the ϕ–ψ contour plots of Ala and Gly and the ΔE vs. ψ curve of Pro, respectively. For the other residues, the backbone dihedral angles included one point in each of the eight regions containing an Ala minimum, and an additional point in region F. Thus, the nine (ϕ, ψ) values used as starting points were (−85°, 80°), (−155°, 155°), (−150°, 70°), (−75°, −45°), (−160°, −60°), (55°, 60°), (65°, 180°), (80°, −65°), and (−85°, 155°). In residues with branching at the C^{β} atom, the last ϕ–ψ starting point was modified to (−60°, 140°) when χ^1 had a

Table I
Side-Chain Dihedral Angles Used as Starting Points for Energy Minimization[a]

Residue	Side-chain dihedral angles, deg			
	χ^1	$\chi^{2\,b}$	$\chi^{3\,c}$	$\chi^{4\,d}$
Ala	60			
Arg	±60, 180	±60, 180	±60, 180	±60, 180[e]
Asn	±60, 180	±30, ±90, ±150	180	
Asp	±60, 180	±30, ±90, ±150	0, 180	
Cys	±60, 180	±60, 180		
Gln	±60, 180	±60, 180	±30, ±90, ±150	180
Glu	±60, 180	±60, 180	±30, ±90, ±150	0, 180
Gly	*f*			
His[g]	±60, 180	±30, ±90, ±150		
Ile	±60, 180	±60, 180	60	60
Leu	±60, 180	±60, 180	60	60
Lys	±60, 180	±60, 180	±60, 180	±60, 180[h]
Met	±60, 180	±60, 180	±60, 180	60
Phe	±60, 180	90, ±30		
Pro	*f*			
Ser	±60, 180	±60, 180		
Thr	±60, 180	±60, 180	60	
Trp	±60, 180	±30, ±90, ±150		
Tyr	±60, 180	±30, ±90, ±150	0	
Val	±60, 180	60	60	

[a] All possible combinations of these sets of dihedral angles were used with the (ϕ, ψ) values given in the text. [b] This angle is $\chi^{2,1}$ in Ile, Thr, and Val. [c] This angle is $\chi^{2,2}$ in Ile, Thr, and Val; $\chi^{3,2}$ in Asp and Asn; $\chi^{3,1}$ in Leu; and χ^6 in Tyr. [d] This angle is χ^3 in Ile; $\chi^{4,2}$ in Glu and Gln; and $\chi^{3,2}$ in Leu. [e] χ^5, $\chi^{6,1}$, and $\chi^{6,2}$ of Arg were not allowed to vary during minimization (see text). [f] These residues have no variable side-chain dihedral angles. [g] These values for the side-chain dihedral angles of His were used both with the proton on the N^δ and with the proton on the N^ϵ atoms. [h] χ^5 of Lys was also allowed to vary during minimization, with starting values of ±60°, and in certain cases 180° (see text).

value of 60°. This was necessary because, in some cases, when the starting point was (−85°, 155°), the conformation moved out of the F region during minimization.

2. Selection of Side-Chain Dihedral Angles for Starting Conformations. Table I lists the side-chain dihedral angles selected as starting conformations for subsequent energy minimization. Where applicable, the dihedral angles chosen are those at which the intrinsic torsional potential has a minimum.[10] All possible combinations of these side-chain conformations and the nine backbone conformations made up the starting points for each residue. To limit further the amount of computing, energy minimization of Lys conformations was carried out in two steps. First, χ^5 was taken at +60° or −60° (but not at 180°) and, in combination with the starting values for the other torsional angles, the energy was minimized. In the second step, the value of χ^5 at each minimum obtained in the first step was replaced by $\chi^5 = 180°$, and the resulting conformations (with redundancies removed) were used as starting points. Thus, minimization with $\chi^5 = 180°$ was carried out from only 82 points rather than from 729 points. A similar procedure was used for Arg, minimizing first with (χ^5, $\chi^{6,1}$, $\chi^{6,2}$) held fixed at (0°, 180°, 180°) and at (180°, 180°, 180°), and then replacing (χ^5, $\chi^{6,1}$, $\chi^{6,2}$) in the resulting minima with (0°, 0°, 180°) and (180°, 0°, 180°) to obtain a second set of starting conformations.

The total number of starting points for all 20 residues was about 6500, whereas the total number selected by Lewis et al.[5] was about 700 (after considering the energy of some 16 000 conformations prior to minimization).

3. Procedure for Energy Minimization. Minimization was carried out using the Powell algorithm,[21] with a convergence criterion of 0.002 kcal/mol. Termination also occurred when each of the variable dihedral angles changed by less than 0.02° between any two iterations.

During minimization, ϕ, ψ, and all side-chain dihedral angles (except χ^5, $\chi^{6,1}$, and $\chi^{6,2}$ of Arg; see above) were allowed to vary. Peptide bonds were held fixed at 180°, except for the one preceding the pyrrolidine ring of Pro, which was taken at 0° and at 180° and allowed to vary during minimization. Torsional angles of the end methyl groups were held fixed at 180°.

G. Conformational Entropy and Free Energy. The normalized statistical weight w_i, which expresses the probability of occurrence of the ith conformation, in which there are small fluctuations in the dihedral angles about the energy minimum, is given by[22]

$$w_i = [(2\pi RT)^{m/2}(\det \mathbf{F}_i)^{-1/2} \exp(-\Delta E_i/RT)]/Z \quad (1)$$

where ΔE_i is the conformational energy at the ith minimum, R is the gas constant, T is the temperature (taken as 300 K throughout this study), m is the number of variable dihedral angles (degrees of freedom), and $\mathbf{F}_i$ is the matrix of second derivatives[22] at the ith minimum. The partition function Z is given by

$$Z = (2\pi RT)^{m/2} \sum_{i=1}^{n} (\det \mathbf{F}_i)^{-1/2} \exp(-\Delta E_i/RT) \quad (2)$$

where n is the number of low-energy minima. The $\mathbf{F}_i$ matrix was computed numerically using a double precision version of ECEPP. The second derivatives at the minimum can be calculated accurately only if the minimum is located more precisely than is possible by the minimization procedure described above; therefore, prior to evaluating the $\mathbf{F}_i$ matrix, further minimization was carried out by the Newton–Raphson method.[23]

The conformational free energy G_i at the ith minimum is defined as

$$G_i = -RT \ln w_i \quad (3)$$

The relative free energy is given by

$$\Delta G = G - G_0 \quad (4)$$

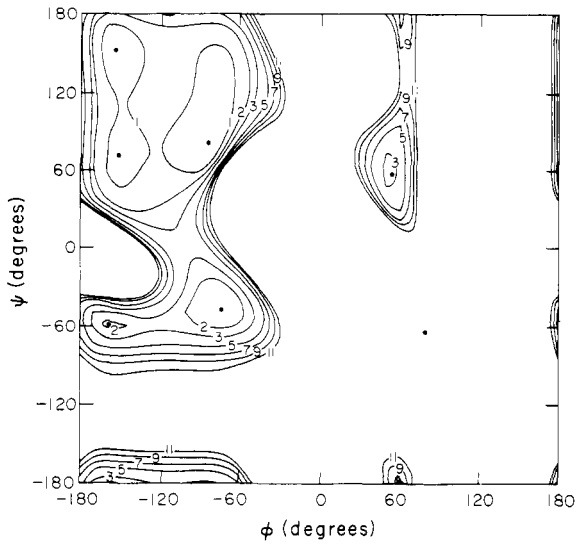

Figure 2. Conformational energy contour map of *N*-acetyl-*N'*-methyl-alanineamide, for $\chi^1 = 60°$. Locations of minima are indicated by the filled circles (also see Table II). The contour lines are labeled with energy in kcal/mol above the minimum-energy point at $(\phi, \psi) = (-84°, 79°)$.

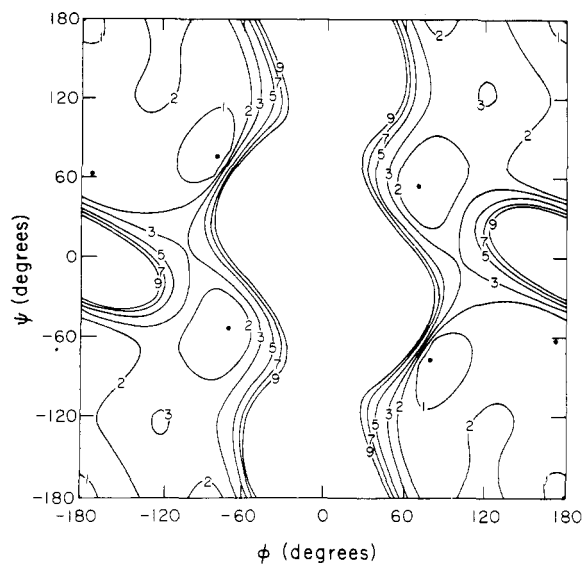

Figure 3. Conformational energy contour map of *N*-acetyl-*N'*-methyl-glycineamide. Locations of minima are indicated by the filled circles (also see Table III). The contour lines are labeled with energy in kcal/mol above the minimum energy points at $(\phi, \psi) = (-83°, 76°)$ and $(83°, -76°)$.

where G_0 is the free energy of the conformation of lowest energy (i.e., the one at $\Delta E = 0$). Since these calculations were carried out in the gas phase at constant pressure and volume, $\Delta H = \Delta E$ and $\Delta G = \Delta E - T\Delta S$. Therefore, the relative entropy is given by

$$\Delta S = (1/T)(\Delta E - \Delta G) \quad (5)$$

This is equivalent to the definition of the librational entropy given by Gō et al.[24]

H. Hydrogen Bonds. Each minimum-energy structure was analyzed for backbone–backbone and backbone–side chain hydrogen bonds. A computer algorithm was employed to calculate all H···A distances, where H is a polar hydrogen atom and A is a proton acceptor (an oxygen or nitrogen atom). If the H···A distance was less than or equal to 2.3 Å, the two atoms were considered to be involved in a hydrogen bond.

III. Results

A. Conformational Energy Contour Maps. Figures 2 and 3 give the conformational energy contour maps of the blocked single residues Ala and Gly, respectively. The locations of minimum-energy conformations (see Tables II and III) are also marked. The energies at the locations of the Ala minima marked on Figure 2 may not correspond exactly to the energies given in Table II because, in determining the contour map, χ^1 was held fixed at 60°, whereas during minimization it was allowed to vary.

Figure 4 gives a plot of ΔE vs. ψ for Pro, with the peptide bond preceding the proline ring taken in the trans (180°) and in the cis (0°) conformations. The locations of minima are essentially the same as those listed for L-Pro in Table II of ref 25.

B. Low-Energy Conformations of Single Residues. Tables II, III, and IV contain data on all minima for each of the amino acid residues Ala, Gly, and Pro. Similar tables, containing minima with $\Delta E < 5$ kcal/mol (or $\Delta E < 3$ kcal/mol in the case of Lys and Arg) for each of the 20 naturally occurring amino acid residues, are given in the supplementary material.[26] For each minimum-energy conformation, the tables give (a) the conformational letter code, (b) a notation of the presence of any hydrogen bonds, (c) the relative energy ΔE, (d) the relative free energy ΔG and its entropic contribution $-T\Delta S$, (e) the components of energy, ΔE_{ES}, ΔE_{NB}, and ΔE_{TOR}, and (f) the dihedral angles.

Each amino acid residue has one or more low-energy ($\Delta E \leq 5$ kcal/mol) minima in each of the seven regions C, E, F, D,

Table II
Minimum-Energy Conformations of *N*-Acetyl-*N'*-methyl-alanineamide[a]

							Dihedral angles, deg[d]		
Code	ΔE^b	ΔG^c	$-T\Delta S$	ΔE_{ES}	ΔE_{ES}	ΔE_{TOR}	ϕ	ψ	χ^1
C	0.0[e]	0.0	0.0	0.0	0.0	0.0	−84	79	61
E	0.382[e]	0.028	−0.354	0.667	−0.285	0.000	−154	153	60
D	0.719	−0.086	−0.805	0.823	−0.108	0.003	−150	72	59
A	1.131	0.947	−0.184	1.374	−0.243	0.000	−74	−45	61
G	1.581	1.875	0.294	1.291	0.216	0.074	−158	−58	54
A*	2.348	2.646	0.298	1.348	0.951	0.049	54	57	65
F*	5.027			1.607	2.719	0.702	64	−178	80
C*	8.778[e]			−0.411	8.029	1.160	78	−64	87

[a] All minimum-energy conformations are listed. ΔG and ΔS were calculated only for minima with $\Delta E < 3$ kcal/mol. [b] $E_0 = -3.187$ kcal/mol. [c] $G_0 = -0.724$ kcal/mol. Values of ΔG and ΔS are given for $T = 300.0$ K. [d] Only variable dihedral angles are listed; all others have a value of 180°. [e] These conformations have a backbone–backbone hydrogen bond.

Table III
Minimum-Energy Conformations of *N*-Acetyl-*N'*-methyl-glycyineamide[a]

Code	ΔE[b]	ΔG[c]	$-T\Delta S$	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	Dihedral angles, deg[d] ϕ	ψ
C*	0.0[e]	0.0	0.0	0.0	0.0	0.0	83	−76
C	0.0[e]	0.0	0.0	0.0	0.0	0.0	−83	76
E	0.816[e]	−0.135	−0.952	0.713	0.103	0.0	180	180
D*	1.024	−0.009	−1.033	1.008	0.016	0.0	173	−62
D	1.024	−0.009	−1.033	1.008	0.016	0.0	−173	62
A*	1.181	0.553	−0.628	1.474	−0.293	0.0	72	53
A	1.181	0.553	−0.628	1.474	−0.293	0.0	−72	−53

[a] All minimum-energy conformations are listed. [b] $E_0 = -4.708$ kcal/mol. [c] $G_0 = -1.076$ kcal/mol. The values of ΔG and ΔS are given for $T = 300.0$ K. [d] Only variable dihedral angles are listed; all others have a value of 180°. [e] These conformations have a backbone–backbone hydrogen bond.

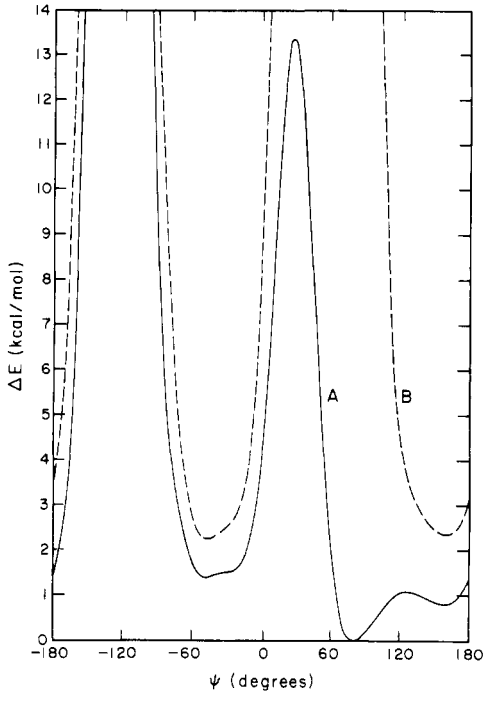

Figure 4. Curve of ΔE vs. ψ for *N*-acetyl-*N'*-methyl-prolineamide, with the peptide bond preceding the pyrrolidine ring taken (a) in the trans (solid line) and (b) in the cis (dashed line) conformations. The peptide group following the pyrrolidine ring was kept in the trans conformation.

A, G, and A*, with the following exceptions: no minimum occurs in region G for glycine, isoleucine, and valine, and no minimum occurs in region F for glycine and alanine. The energy of minima in region A* is always high; the A* conformation of lowest energy for each of the residues occurs with ΔE between 1.2 and 3.6 kcal/mol. Only three amino acids (alanine, histidine, and serine) have one or more local minima in region F*, with energies between 3.2 and 5.0 kcal/mol.

All minima of all residues, as listed in the tables in the supplement, were plotted in Figure 1. Generally, they cluster into one or more small areas of each low-energy region, showing that the potential energy surfaces usually do not vary much from one amino acid to the other. The clusters of points in the various low-energy regions are well separated from each other, except near the boundaries between regions D and E, and between regions C and F. However, if the energy minima for each individual amino acid are plotted separately, the clustering of the minima into separate regions is much more marked than in the composite plot of Figure 1. Conformations differing from each other merely by side-chain rotations are always found in the same region as described by the letter code.

No direct comparison can be made between the *exact* locations of minima in Figure 1 and the specific values of φ and ψ observed for amino acid residues in globular proteins. This is because medium- and long-range interactions can shift the φ–ψ values of residues in proteins away from the locations of minima, calculated here for *isolated* single residues, to other locations *within the low-energy region.*

C. Number of Minimum-Energy Conformations. Table V lists the number of starting points from which energy minimization was carried out and the resulting number of minima with $\Delta E < 3$ and $3 \leq \Delta E < 5$ kcal/mol. For comparison, the table also gives the number of low-energy minima obtained by Lewis et al.[5]

IV. Discussion

The purposes of this study have been (a) to locate and describe all low-energy conformations for each of the 20 naturally occurring amino acid residues and (b) to compare these results, obtained using ECEPP,[10,11] with those of Lewis et al.,[5] who used an earlier set of parameters. The basic conclusions of Lewis et al.[5] have been verified in the present study; therefore, we will not discuss in detail all the results but rather indicate important differences.

A. The Effect of Changes in Energy Parameters. The two major improvements in the parameters since the earlier calculations of Lewis et al.[5] are (a) the softening of the repulsive interaction between an amide nitrogen and an amide hydrogen in another peptide group[10,11,13] and (b) the inclusion of an explicit torsional energy for rotations about aliphatic C–C and C–N single bonds.[10,11] The effects of these changes are discussed below.

1. Effects of Softening Amide N···Amide H Repulsions. The change in conformational properties due to softening the repulsion between an amide N and the amide H of another peptide group is shown by the following four examples.

(a) In region B around $(\phi, \psi) = (-90°, 0°)$, the energy is 3–4 kcal/mol rather than the 10 kcal/mol obtained earlier.[5] This is seen clearly by comparing the bridge region of the contour plots of Ala and Gly (Figures 2 and 3) with the same region in Figures 3 and 2 of ref 5.

(b) The minima of Ala and Gly with $|\psi| < 70°$ (i.e., the ones near the bridge region) have shifted toward $\psi = 0°$. For example, the minimum of Ala in the G region was located previously[5] at $(\phi, \psi) = (-160°, -68°)$, but it is now located at $(\phi, \psi) = (-158°, -58°)$ (Table II), a shift in ψ of 10°.

(c) Not only have the locations of certain minima shifted closer to $\psi = 0°$ but also the relative energies of those minima have decreased. Taking again the example of the G minimum

Table IV
Minimum-Energy Conformations of *N*-Acetyl-*N'*-methyl-prolineamide[a]

Code	ΔE[b]	ΔG[c]	$-T\Delta S$	ΔE_{ES}	ΔE_{NB}	ΔE_{TOR}	Dihedral angles, deg[d] ω	ψ
C	0.0[e]	0.0	0.0	0.0	0.0	0.0	178	79
F	0.831	0.416	−0.415	1.182	−0.327	−0.024	180	159
A	1.407	1.536	0.129	1.247	0.185	−0.025	180	−48
A	2.101	2.000	−0.101	1.207	1.117	0.113	−5	−48
F	2.240	2.003	−0.238	1.950	0.538	0.088	−4	162

[a] All minimum-energy conformations are listed. [b] $E_0 = -17.896$ kcal/mol. [c] $G_0 = -0.296$ kcal/mol. The values of ΔG and ΔS are given for $T = 300.0$ K. [d] Only variable dihedral angles are listed; all others have a value of 180°. [e] This conformation has a backbone–backbone hydrogen bond.

of Ala, the value of ΔE was calculated earlier to be 2.49 kcal/mol[5] but is now only 1.58 kcal/mol.

(d) An important effect of the decreased interaction energy between an amide N and an amide H is the appearance of a new low-energy minimum of Ala at $(\phi, \psi) = (-150°, 72°)$. This minimum has an energy only 0.72 kcal/mol above the global minimum and is the conformation of lowest free energy (at $T = 300$ K) (see Table II). An analogous minimum exists in this region (D) of ϕ–ψ space for many of the single residues (see tables in supplementary material).

2. Effects of Adding Explicit Torsional Potentials. The second major change since the work of Lewis et al.[5] has been the addition of explicit torsional potentials[10,11] for rotations about aliphatic C–C and C–N single bonds in the side chains of the amino acid residues. This change does not, of course, affect Gly and Pro but does influence all other residues. Changes in Ala, for example, are seen by comparing Table II in this paper with Table I in ref 5. In the earlier work,[5] the value of χ^1 deviated from its presumed minimum-energy position of 60° by more than 10° in four of the seven minima; in the present study, such a deviation has been found in only two cases. At those minima, with dihedral angles $(\phi, \psi, \chi^1) = (64°, -178°, 80°)$ and $(\phi, \psi, \chi^1) = (78°, -64°, 87°)$ (Table II), the torsional energies ΔE_{TOR} are 0.70 and 1.16 kcal/mol, and the total energies are high, viz., 5.03 and 8.78 kcal/mol, respectively. By comparison, the analogous minima at $(\phi, \psi, \chi^1) = (65°, -177°, 90°)$ and $(\phi, \psi, \chi^1) = (78°, -65°, 93°)$ given by Lewis et al.[5] have, of course, no explicit torsional energies and have total energies of 3.95 and 7.55 kcal/mol, respectively.

In residues with large side chains, the effects are even greater. For essentially every residue with more than one side-chain single bond about which rotation can take place (see tables in supplementary material), it is observed that (a) some conformations which were minima without the introduction of explicit torsional potentials[5] are not minima after the torsional potentials are included, (b) some new minima appear, (c) the locations of certain minima are altered, and (d) the relative energies of minima are changed. In the cases of Ser and Asp, for example, where many of the minimum-energy conformations have backbone–side chain hydrogen bonds which cause the side-chain dihedral angles to deviate substantially from their intrinsic torsional minima, the confor-

Table V
Numbers of Starting Points and of Local Minima for the 20 Naturally Occurring Amino Acid Residues

Residue	No. of starting points	No. of minima with $\Delta E < 3$ kcal/mol: This study	No. of minima with $\Delta E < 3$ kcal/mol: Lewis et al.[5]	No. of minima with $3 \leq \Delta E < 5$ kcal/mol: This study	No. of minima with $3 \leq \Delta E < 5$ kcal/mol: Lewis et al.[5]
Ala	9	6	6	0	0
Arg	1600	144	72	*a*	24
Asn	162	28	12	10	8
Asp	324	32	8	10	7
Cys	81	44	31	3	11
Gln	486	69	30	31	2
Glu	972	60	25	34	6
Gly	9	7	7	0	0
His (N^δ)	162	35	11	3	3
His (N^ϵ)	162	36	*b*	2	*b*
Ile	81	11	8	6	3
Leu	81	13	13	15	5
Lys	1540	176	71	*a*	4
Met	243	56	49	25	4
Phe	81	12	6	3	4
Pro	6	5	2	0	0
Ser	81	9	8	30	16
Thr	81	16	10	9	7
Trp	162	29	11	4	3
Tyr	162	24	7	6	5
Val	27	8	6	2	0

[a] Tables in the supplementary material[26] list only minima with $\Delta E < 3$ kcal/mol for Arg and Lys. [b] Lewis et al.[5] did not determine minima for His with a proton on the N^ϵ atom.

mation of lowest energy is not the same as it was without the use of explicit torsional potentials.[5]

B. Effect of Changes in Geometry and Energy Parameters in Proline. Two recent improvements in proline parameters have been made. First, the geometry of the proline residue was modified by Burgess et al.[27] because of the publication of more recent x-ray crystal data; and second, the analysis by Tanaka and Scheraga[18] showed that the internal conformational energy in the fixed part of the proline residue depends on the conformation (cis or trans) of the peptide bond preceding the pyrrolidine ring. The effects of incorporating these changes into the calculations are seen in Table IV and in Figure 4. The locations and relative energies of the minima are somewhat different from those given by Lewis et al.[5] The energy difference between the cis ($\omega \simeq 0°$) and trans ($\omega \simeq 180°$) forms is about 2 kcal/mol rather than about 7 kcal/mol as reported previously.[5] This lower value is closer to the experimental value[28] of about 1.8 kcal/mol (i.e., ~5% cis) for the X-Pro peptide bond in various linear peptides.

C. Effects of Differences in Methods of Searching Conformation Space. The results suggest that the present method of searching conformation space to locate energy minima is an improvement over the previous one,[5] for the following two reasons. First, we succeeded in locating all the minima found by Lewis et al.[5] (when proper consideration is made for differences in parameters). And second, for all residues (except Gly and Pro), we found more local energy minima (see Table V) than were found previously.[5] (This is due in part also to differences in parameters.)

D. Conformations with Hydrogen Bonds. The tables in the supplementary material[26] indicate the minimum-energy conformations which contain backbone–side chain and/or backbone–backbone hydrogen bonds. In many cases, the assignment of hydrogen bonds differs from that of the earlier work[5] because (a) the location in conformational space of the minima obtained with ECEPP is often shifted from that obtained[5] with the earlier parameters, and (b) in the present analysis, a precise definition of a hydrogen bond (H···A distance not greater than 2.3 Å) has been employed. For example, the global minimum of Asn was reported previously[5] to have two backbone–side chain hydrogen bonds, but the analogous conformation in the present analysis has no hydrogen bond.

Another important difference between the present analysis of hydrogen bonds and the earlier one[5] is that, in this study, we have checked for backbone–side chain hydrogen bonds not only in Asp, Asn, Glu, Gln, Ser, and Thr but also in Arg, His, Lys, and Tyr. In fact, a hydrogen bond is found to exist in one conformation of Arg and in four conformations of His (N^{δ}), but not in Lys or Tyr.

Backbone–side chain hydrogen bonds not considered previously[5] also occur when the carboxyl O–H bond is trans to the C=O bond in the side chain of Asp or Glu. Although the trans is intrinsically less favorable than the cis form by about 5 kcal/mol,[10] the OH group in the trans form is more accessible for hydrogen bonding with the backbone. Figure 5 shows an example of such a hydrogen bond in Asp. The value of ΔE for this structure is only 0.24 kcal/mol (although ΔG = 2.2 kcal/mol). Five low-energy ($\Delta E < 3$ kcal/mol) conformations of Asp and two low-energy conformations of Glu contain this type of backbone–side chain hydrogen bond.

E. Librational Entropy and Free Energy. In Tables II–IV in the text and in the tables of the supplementary material,[26] the effect of librational entropy on the relative stabilities of low-energy conformations is expressed in terms of $-T\Delta S$, where T = 300 K. We should stress that eq 1 and 2, which are used to calculate the entropy and free energy, are approximations.[22] We have shown that these approximations are generally good,[29] but when the energy barrier between two minima is low, the approximation of eq 1 may not be quantitatively valid.[22] Moreover, the hydrogen-bond potential used in ECEPP may be too "stiff".[4] As a result, the calculated entropy of structures with hydrogen bonds may be too low and the free energy too high. Keeping these precautions in mind, we list below some general observations from the calculations of librational entropies.

1. In ten of the twenty residues, the global energy minimum is not the conformation of lowest free energy (for example, see minima of Ala and Gly with $\Delta G < 0$ in Tables II and III), and in general the entropy causes a change in the relative stabilities of minima.

2. Although the values of $-T\Delta S$ range from −2.0 to 2.2 kcal/mol, the majority (>90%) of low-energy minima have values of $-T\Delta S$ which fall in the range of −1.0 to 0.5 kcal/mol.

3. Conformations which lie in regions G and A* tend to be destabilized by entropy. For example, the G and A* minima of Ala have $-T\Delta S$ values which are 0.3 to 1.1 kcal/mol higher than the $-T\Delta S$ of any one of the other minima. Similarly, for other residues, the values of $-T\Delta S$ are always positive for the G and A* conformations, except in Glu and Gln for which the global minimum ($-T\Delta S \equiv 0$) is also very low in entropy.

The reason for the low entropy of G and A* minima is seen qualitatively in Figure 2. The potential energy well in each of those regions is relatively narrow, as compared with the wells around the other minima.

4. In general, conformations which contain backbone–side chain hydrogen bonds are low in entropy. For example, in Asp, eight of the nine low-energy ($\Delta E < 3$ kcal/mol) conformations with backbone–side chain hydrogen bonds have positive values of $-T\Delta S$. (Of the six conformations without such a hydrogen bond but with a positive value of $-T\Delta S$, five are in regions G or A*.)

5. On the other hand, backbone–backbone hydrogen bonds generally do not increase the values of $-T\Delta S$. This is because such hydrogen bonds are weak (the H···O or H···N distance is

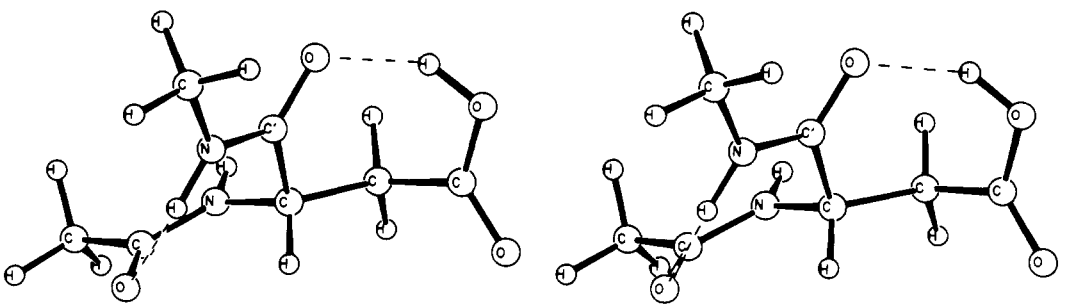

Figure 5. Stereo diagram of *N*-acetyl-*N'*-methyl-aspartic acid amide, showing a backbone–backbone (C_7^{eq}) and a backbone–side chain hydrogen bond (dashed lines). The dihedral angles of Asp in this conformation are (ϕ, ψ, χ^1, χ^2, $\chi^{3,2}$) = (−83°, 76°, −179°, 90°, 17°).

Table VI
The Ratio of Molecular Populations in the C_5 and the (C_7^{eq} + γ) Conformations[a] in *N*-Acetyl-*N'*-methylamides of Amino Acids

Residue[b]	Temp,[c] °C	Ratio of populations $(C_5)/(C_7^{eq} + \gamma)$ Calcd: This work	Calcd: Ref 5	Obsd	Ref
Gly	50	0.6	0.7	0.7	6
Ala	50	1.0	1.6	0.4	6
	90	1.1	1.6	1.7	7
Val	50	0.2	0.2	0.3	6
Cys	25	0.6	0.9	$\ll 1$	9
Ser	90	0.4	1.3	0.3	7
Leu	25	0.5	1.8	0.3	9
Met	25	0.3	0.3	>0.4	9
Asn	25	0.6	0.9	$\ll 1$	9
Gln	25	0.5	0.8	$\ll 1$	9
Phe	25	1.6	2.2	1.2	9

[a] The populations were computed as the sums of the relative populations of all conformations in regions E, C, and F for the states C_5, C_7^{eq}, and γ, respectively, based on the ratios of statistical weights for the individual conformations. [b] Those amino acid residues are listed for which experimental data are available. [c] The calculations were done using the temperatures reported in the experimental studies.

relatively large and the N–H···O or N–H···N angle deviates widely from linearity) and because most of them occur in conformations found in the C and E regions where the potential energy surface is flat.

F. Comparison with Experiment. It is clear from experimental data[6-9,30,31] that the C_7^{eq} and C_5 conformations (occurring in regions C and E, respectively) are stable structures in *N*-acetyl-*N'*-methylamides of single residues and that other stable conformations, such as the γ conformation (occurring in region F), also exist.[32] Our calculations (see tables in the supplementary material) show that the populations of conformations in region F are significant, as compared with region C, for Phe and the other amino acids with aromatic side chains, as well as for Ile, in agreement with interpretations of infrared spectra obtained in inert solvents.[32] For all other amino acid residues, conformations in region F make a very small contribution to the conformer population. Our computed ratios of populations in the C_5 and the (C_7^{eq} + γ) conformations are compared with the earlier calculations[5] and with experimental data in Table VI. In view of the uncertainties in the interpretation of the experimental data,[32] the results shown here are in good agreement with the experiments. Some theoretical treatments failed to show a minimum in the region of the C_7^{eq} conformation[33] or wrongly concluded that the C_7^{ax} conformation is more stable than the C_7^{eq} in Ala.[34]

The calculated conformations can also be compared with those observed for amino acid derivatives in crystals. The validity of such comparisons is very limited, however, because intermolecular interactions, due to crystal packing, can alter the dihedral angles from those occurring in the isolated molecule. Nevertheless, it is worth noting the results of x-ray analyses on two amino acid derivatives: (a) The observed[16] conformation of Tyr (ϕ, ψ, χ^1, χ^2 = −155°, 174°, 57°, 90°) in glycyl-tyrosine·2H_2O is very close to one of the low-energy minima (ΔE = 0.016 kcal/mol) computed here (viz., ϕ, ψ, χ^1, χ^2 = −157°, 158°, 60°, 89°). (b) The side-chain conformation in the crystal[35] of tryptophan *O*-ethyl ester is near (χ^1, χ^2) = (60°, −93°). In one of the low-energy extended conformations (region E), these dihedral angles were calculated to be (60°, −88°). For these two amino acid derivatives, no comparable minima were found in the earlier study.[5]

V. Conclusions

Employing a set of empirical energy parameters which includes not only the parameters obtained from intermolecular interactions in crystals[4] but also parameters obtained by considering intramolecular interactions,[10] we have located the minimum-energy conformations (within 5 kcal/mol of the global minimum) of the *N*-acetyl-*N'*-methyl amides of the 20 naturally occurring amino acids. These minima have been characterized not only by their conformational energies but also by their conformational entropies and free energies. In addition, all minima have been analyzed systematically for backbone–backbone and backbone–side chain hydrogen bonds. We have found that, although these results agree in general with the earlier analysis of Lewis et al.,[5] there are significant differences.

These results are essential for the frequent use of single-residue minima in the analysis of larger peptides and proteins.

Supplementary Material Available: Tables 1-21 containing data for the low-energy minima of the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids (28 pages). Ordering information can be found on any current masthead page.

References and Notes

(1) This work was supported by research grants from the National Institute of General Medical Sciences, the National Institutes of Health, U.S. Public Health Service (GM-14312), and the National Science Foundation (BMS75-08691).
(2) NIH Postdoctoral Trainee, 1973–1974; NIH Postdoctoral Fellow, 1974–1976.
(3) For a recent review, see H. A. Scheraga, "Peptides, Polypeptides, and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Ed., Wiley, New York, N.Y., 1974, p 49.
(4) (a) F. A. Momany, L. M. Carruthers, R. F. McGuire, and H. A. Scheraga, *J. Phys. Chem.*, **78**, 1595 (1974); (b) F. A. Momany, L. M. Carruthers, and H. A. Scheraga, *ibid.*, **78**, 1621 (1974).
(5) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Isr. J. Chem.*, **11**, 121 (1973).
(6) M. Avignon and P. V. Huong, *Biopolymers*, **9**, 427 (1970).
(7) J. Smolikova, A. Vitek, and K. Blaha, *Collect. Czech. Chem. Commun.*, **36**, 2474 (1971).
(8) M. Marraud, J. Néel, M. Avignon, and P. V. Huong, *J. Chim. Phys. Phys.-Chim. Biol.*, **67**, 959 (1970).
(9) M. T. Cung, M. Marraud, and J. Néel, *Jerusalem Symp. Quantum Chem. Biochem.*, **5**, 69 (1973).
(10) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
(11) The FORTRAN computer program for ECEPP, its description, and all associated geometric and energy parameters are available on magnetic tape from the Quantum Chemistry Program Exchange. Write to Quantum Chemistry Program Exchange, Chemistry Department, Room 204, Indiana University, Bloomington, Ind. 47401 for standard program request sheets, and then order No. QCPE 286.
(12) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, **9**, 3471 (1970).
(13) K. Nishikawa, F. A. Momany, and H. A. Scheraga, *Macromolecules*, **7**, 797 (1974). See also the Note Added in Proof in ref 5.
(14) F. M. Pohl, *Nature (London), New Biol.*, 277 (1971).
(15) I. L. Karle, *J. Am. Chem. Soc.*, **96**, 4000 (1974).
(16) M. Cotrait and J. Bideau, *Acta Crystallogr., Sect. B*, **30**, 1024 (1974).
(17) T. Matsuzaki and Y. Iitaka, *Acta Crystallogr., Sect. B*, **27**, 507 (1971).
(18) S. Tanaka and H. A. Scheraga, *Macromolecules*, **7**, 698 (1974).
(19) H. A. Scheraga, *Adv. Phys. Org. Chem.*, **6**, 103 (1968).
(20) P. J. Flory, "Statistical Mechanics of Chain Molecules", Interscience, New York, N.Y., 1969, p 56.
(21) M. J. D. Powell, *Comput. J.*, **7**, 155 (1964).
(22) (a) N. Gō and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969); (b) *Macromolecules*, **9**, 535 (1976).
(23) See, for example, G. Dahlquist and A. Bjorck, "Numerical Methods", Prentice-Hall, Englewood Cliffs, N.J., 1974, p 249.
(24) N. Gō, M. Gō, and H. A. Scheraga, *Macromolecules*, **7**, 137 (1974).
(25) M. H. Miller and H. A. Scheraga, *Polym. Symp.*, in press.

(26) See paragraph at end of paper regarding supplementary material.
(27) A. W. Burgess, F. A. Momany, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 1456 (1973).
(28) K. Wüthrich, Ch. Grathwohl, and R. Schwyzer, "Peptides, Polypeptides, and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Ed., Wiley, New York, N.Y., 1974, p 300.
(29) S. S. Zimmerman and H. A. Scheraga, "Peptides: Chemistry, Structure, and Biology", R. Walter and J. Meienhofer, Ed., Ann Arbor Science Publishers, Ann Arbor, Mich., 1975, p 263.
(30) Y. Grenie, M. Avignon, and C. Garrigou-Lagrange, *J. Mol. Struct.*, **24**, 293 (1975).
(31) G. M. Crippen and J. T. Yang, *J. Phys. Chem.*, **78**, 1127 (1974).
(32) A. W. Burgess and H. A. Scheraga, *Biopolymers*, **12**, 2177 (1973).
(33) See, for example, G. N. Ramachandran, ref 28, p 14.
(34) See, for example, B. Pullman and B. Maigret, *Jerusalem Symp. Quantum Chem. Biochem.*, **5**, 13 (1973).
(35) B. K. Vijayalakshmi and R. Srinivasan, *Acta Crystallogr., Sect. B*, **31**, 999 (1975).

Statistical Mechanical Treatment of Protein Conformation. 5. A Multistate Model for Specific-Sequence Copolymers of Amino Acids[1]

Seiji Tanaka[2a] and Harold A. Scheraga*[2b]

Department of Chemistry, Cornell University, Ithaca, New York 14853.
Received March 19, 1976

ABSTRACT: One-dimensional short-range interaction models for specific-sequence copolymers of amino acids have been developed in this series of papers. In the present paper, a multistate model [involving right-handed helical (h_R), extended (ϵ), chain-reversal (R and S), left-handed helical (h_L), right-handed bridge-region (ζ_R), left-handed bridge-region (ζ_L), and coil (or other) (c) states] is developed for the prediction of protein backbone conformation. This model involves ten parameters (w_{hR}, v_{hR}, v_ϵ, v_R, v_S, w_{hL}, v_{hL}, $u_{\zeta R}$, $u_{\zeta L}$, and u_c) and requires a 10×10 statistical weight matrix. Assuming that the left-handed helical *sequence* cannot occur in proteins, this 10×10 matrix can be reduced to a 9×9 matrix with nine parameters (w_{hR}, v_{hR}, v_ϵ, v_R, v_S, v_{hL}, $u_{\zeta R}$, $u_{\zeta L}$, and u_c). A nearest neighbor approximation of this multistate model is also formulated; with the omission of left-handed helical *sequences*, and the inclusion of the *left*-handed bridge region in the c state, this approximate model requires a 7×7 matrix with statistical weights w_{hR}^*, v_{hR}^*, v_R^*, v_S^*, v_{hL}^*, $u_{\zeta R}^*$, and u_c^*, expressed as values relative to the statistical weight of the ϵ state. The statistical weights for the multistate model are evaluated from the atomic coordinates of the x-ray structures of 26 native proteins. These statistical weights and the multistate model are applied in the prediction of the backbone conformations of proteins. The conformational probabilities of finding a residue in h_R, ϵ, R, S, h_L, ζ_R, or c states, defined as relative values with respect to their average values over the whole molecule, are calculated for bovine pancreatic trypsin inhibitor and clostridial flavodoxin, in order to select the most probable conformation for each residue of these proteins. The predicted results are compared to experimental observations and are discussed together with the reliability of the statistical weights. In the Appendix, the property of asymmetric nucleation of helical sequences is introduced into the (nearest neighbor) multistate model.

In this series of papers,[3-6] we have developed a statistical mechanical treatment of protein conformation within the context of one-dimensional short-range interaction models. These will be referred to here as papers I,[3] II,[4] III,[5] and IV,[6] with equations designated as I-1, II-1, III-1, etc.

In paper I, we presented a method for evaluating empirical statistical weights for various conformational states of amino acid residues on the basis of the reported (x-ray) conformations of native proteins.[3] In paper II,[4] we divided the conformational space of a residue into helical (h), extended (ϵ), and coil (other) (c) regions and formulated a three-state model for specific-sequence polypeptides, to predict protein conformation in paper III.[5] In paper IV,[6] we extended this treatment to a four-state model that included chain-reversal (R and S) states, as well as h, ϵ, and c states; also, in paper IV (as well as in the present paper), we used the X-ray coordinates (instead of the crystallographers' description of the conformational states) to evaluate the statistical weights. Since many (conformationally undefined) amino acid residues remain in the c state, even in the four-state model (see the last column of Table II of paper IV[6]), we can specify these conformations more precisely by further dividing the c region, in the present paper, and thereby develop a multistate model for treating protein conformation.

Even though the multistate model developed here specifies the conformational state of a residue more precisely, this does not render the four-state model obsolete for two reasons: (a) the four-state model is more convenient to use (because of its smaller size statistical weight matrix) and gives valid information about helical, extended, and chain-reversal conformations, and (b) the multistate model will be much improved when more x-ray data become available to provide more accurate statistical weights for the many states involved.

In section I of this paper, we formulate a multistate model for treating polypeptide conformation. In section II, a nearest neighbor treatment of the multistate model is presented. X-ray data on native proteins are analyzed in section III and are used in section IV to evaluate the statistical weights. The method of calculating conformational probabilities (for detecting backbone conformations in proteins) is presented in section V and applied and discussed in section VI. Finally, in section VII, we summarize the results obtained in papers I–V in relation to their applicability to the computation of the three-dimensional structures of proteins. In the Appendix, the (nearest neighbor) multistate model, developed in section II, is incorporated into our earlier model[7] of the helix–coil transition, in which asymmetric nucleation of helical sequences is taken into account.