and L. W. Gold, Ed., Royal Society of Canada, 1973, p 73.
(41) Reference 16, p 138.
(42) D. A. Draegert, N. W. B. Stone, B. Curnutte, and D. Williams, *J. Opt. Soc. Am.*, **56**, 64 (1966).
(43) G. E. Walrafen, *J. Chem. Phys.*, **40**, 3249 (1964).
(44) K. E. Larsson and U. Dahlborg, *J. Nucl. Energy, Parts A/B*, **16**, 81 (1962).
(45) G. E. Walrafen in "Water, A Comprehensive Treatise", Vol. I, F. Franks, Ed., Plenum Press, New York, N.Y., 1972, Chapter 5.
(46) Reference 16, p 137.
(47) Reference 16, p 139.
(48) The difference in temperature between the initial and final states is not important, because in this theory the relevant energies (e.g., E_{lat}) are taken to be temperature independent.

Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids[1]

F. A. Momany,[2a] R. F. McGuire,[2b] A. W. Burgess,[2c] and H. A. Scheraga*[2d]

Department of Chemistry, Cornell University, Ithaca, New York 14853 *(Received January 31, 1975; Revised Manuscript Received July 10, 1975)*

Empirical interatomic potentials are developed for calculating the energetically most favored conformations of polypeptides and proteins. Geometric parameters, partial atomic charges, nonbonded interaction energies, hydrogen bond energies, and intrinsic torsional potentials are determined for each of the naturally occurring amino acids. The geometric parameters were obtained from a survey of the recent structural literature; the bond lengths and bond angles of a given type of residue appear to be similar from structure to structure. The partial atomic charges (overlap normalized) for every atom of each amino acid were obtained by the CNDO/2 method. Parameters of the nonbonded and hydrogen-bonding potentials were taken from previous calculations on crystals of small molecules; however, the coefficient of the nonbonded repulsive term for two atoms separated by three bonds, the central one being the bond about which rotation can take place (1–4 interactions), was reduced by a factor of 2 to make the repulsive force constants compatible with those computed from Hartree–Fock and Thomas–Fermi–Dirac repulsive potentials. Intrinsic torsional potentials were introduced, where necessary, to reproduce experimental internal rotation barriers. Interaction arrays for interatomic pairwise interactions were defined in order to distinguish hydrogen bonding, and 1–4, from 1–5, 1–6, etc., interactions. Procedures are described for carrying out conformational energy calculations on polypeptides and proteins, using these empirical potentials. Computer programs for performing these calculations are available.

I. Introduction

Recent investigations in this and other laboratories[3-12] have emphasized the need for accurate interatomic potentials to determine polypeptide and protein conformations. The present paper provides the details of the set of potentials presently in use in this laboratory.

In order to obtain a perspective on the present status of potential functions, we begin with a short summary of the evolution of these empirical potentials. The earliest investigations of polypeptide conformation were carried out with hard-sphere repulsive potentials, using the van der Waals contact distance as an overlap criterion; these potentials effectively allowed for interatomic excluded volume effects, and provided rough estimates of the allowed regions of conformational space.[4a]

Next, these potential functions were improved by replacing the hard sphere by a Lennard-Jones 6–12 potential. The attractive (dispersion) term was calculated from atomic polarizabilities, and the position of the minimum of the 6–12 potential was taken to be the van der Waals contact distance.[4a,d] Electrostatic energies were also included by estimating partial atomic charges, first from bond moments, and later by using MO–LCAO σ charges of Del Re[4a] and MO π charges from Hückel theory.[4a] Intrinsic torsional terms were included, with rotational barriers derived from model compounds.[4a] At this stage of development, hydrogen bonding was treated by a modification of the Lippincott–Schroeder potential function.[4a] Using this set of potentials, calculations were carried out to investigate the conformational properties of single amino-acid residues, homopolymers, and cyclic peptides.[4] In a few cases [cyclo-(Gly_3Pro_2) and cyclohexaglycyl],[4d] flexible geometry (i.e., bond stretching and bond angle bending) was allowed by using appropriate force constants for these motions. When applied to proteins (e.g., lysozyme,[13] α-lactalbumin,[14] and rubredoxin[15]), united-atom potentials were used for –CH, $-CH_2$, and $-CH_3$ groups to reduce the number of interactions that had to be computed.

In the present paper, we report the results of our reexamination of the problem of interatomic potentials, whereby we have obtained an improved self-consistent set. First, the coefficients of the repulsive interatomic potential have been refined by treating the packing of molecular crystals.[9] In that work, it was found that the position of the minimum of the interatomic potential was *not* the van der

TABLE I: Selected Backbone Bond Lengths and Bond Angles[a]

Bond[b]	Bond length, Å[c]	Bond[b]	Bond length, Å[c]
N–C^{α}	1.453 ± 0.02	C′=O	1.230 ± 0.01
C^{α}–C′	1.530 ± 0.01	N–H	1.000 ± 0.04
C′–N	1.325 ± 0.005	C^{α}–H	1.000 ± 0.10

Bond angle	Bond angle, deg[c,d]
$C^{\alpha}C'N$	115.0
OC′N	124.5
$C^{\alpha}C'O$	120.5
$C'NC^{\alpha}$	121.0
C′NH	124.0
HNC^{α}	115.0
$NC^{\alpha}C'$	
Gly, Met, Tyr, Phe	111.0
Lys, Ala, Leu, Ile, Asp, Asn, Norleu, Val, Orn	109.3
Arg, Gln, Glu	110.3
Ser	110.0
Cys (both cysteine and cystine)	108.3
Thr	110.4
His	109.6
Trp	108.0

[a] For all amino acid residues except proline and hydroxyproline; see Appendix I for data for these two residues. [b] The C^{α}–C^{β} bond lengths are given with the side-chain dimensions of each residue in Appendix I. [c] The standard deviation was computed directly from the data used to obtain the values cited; the uncertainty in the crystal structure determination of the bond length or bond angle was not included explicitly. [d] The standard deviation in all reported bond angles is ±3°.

Waals contact distance, but somewhat larger.[9] These crystal-refined repulsive terms[9] are used here. Second, the partial atomic charges are determined by all-valence electron molecular orbital methods,[12,16] and the coulombic interaction contribution adjusted to give energies in agreement with both MO calculations and long-range dipole–dipole interactions.[12] Third, the hydrogen-bond energy was calibrated by using the molecular orbital attractive energy[9] and a repulsive energy obtained from the crystal packing studies.[9] The results of these crystal calculations led to an internally self-consistent set of interatomic potential energies (not yet including torsional potentials because intramolecular conformation was not varied in the crystal-packing calculations[9]) for interactions between all types of atoms found in polypeptides. These new potentials (and the torsional potentials and computational procedure developed here) were tested, in paper VI of this series, on the *N*-acetyl-*N*′-methylamides of the naturally occurring amino acids,[17] and all low-energy minimum-energy conformations and their statistical weights (obtainable with these potentials) were found; the predicted conformations and statistical weights[17] were in agreement with the limited amount of experimental data available.[4,18–20] In a subsequent paper of this series,[21] we will present the results of calculations using "united atom" potentials.

This brief summary of previous work is not intended to be a comprehensive one, but rather to provide an indication of the present status of the problem and the applicability of the total potential function presently available. Undoubtedly, this function can be improved further, and, also, other approaches will be used to develop empirical potentials; in fact, some preliminary work in this direction has been started.[22–24] However, considering the reasonableness of the results obtained with the present potential function[4,17–20] (which is the first complete one for polypeptides that is based on both gas phase and X-ray crystal data), and the long time required to develop new ones, the present one will be very useful for calculations on polypeptides and proteins in the foreseeable future.

In section II of this paper, we use recent structural data to obtain "standard" geometries (bond lengths and bond angles) for each of the naturally occurring amino acids. The determination of the overlap normalized (ON) partial atomic charges of both charged and uncharged amino acid residues by the CNDO/2 (complete neglect of differential overlap) method[12,16] is described in section III. In section IV, we consider the nonbonded interactions, and discuss the special nature of 1–4 interactions (interactions between atoms connected through three bonds, the central one being the bond about which rotation can take place). Section VI makes use primarily of experimental data [and some EHT (extended Hückel theory)[25] calculations] to determine intramolecular barriers to rotation about bonds and also to obtain information about 1–4 repulsive interactions. The total intramolecular energy potential function, and the procedure for designation of 1–4 interactions, are summarized in section V. A brief summary of the procedure for carrying out conformational energy calculations on polypeptides is given in section VII.

In all of the results presented here, the IUPAC-IUB nomenclature will be used.[26]

II. Geometry

In conformational energy calculations on polypeptides, the number of variables must be kept to a minimum to conserve computer time. We have chosen to fix the bond lengths and bond angles of the amino acid residues at the values observed in X-ray and neutron diffraction studies of crystals of these molecules, and have taken the dihedral angles for rotation about bonds as the variables (ϕ, ψ, and ω in the backbone,[26] and χ^1, χ^2, etc., in the side chains[26]). It is clear that small variations in bond lengths and bond angles may be expected from crystal to crystal for the same amino acid residue because of differing environments. While such variations from an accepted equilibrium structure can be taken into account by introducing appropriate force constants for stretching and bending into conformational energy calculations on polypeptides,[27,28] it appears from an examination of structural data that variations in bond lengths and bond angles of the polypeptide backbone will depend on short-range interactions, i.e., those between the side chain and backbone of the same residue. Thus, e.g., the $NC^{\alpha}C'$ bond angle of alanine would differ from that of tyrosine, but all alanines (and correspondingly all tyrosines) would have essentially the same bond angle in various (unstrained) polypeptides; this condition seems to prevail in the results of crystal structure studies on amino acids. A similar constancy of bond lengths and bond angles in corresponding monomeric units of carbohydrates has been observed.[29] Therefore, a set of bond lengths and bond angles (different for each type of amino acid) can be selected from crystal data, and then assumed to remain fixed in conformational energy calculations on polypeptides and proteins. The use of fixed geometry for the amino acid residues

TABLE II: Backbone Partial Charges[a]

Atoms[b]	Gly	Pro-(Hypro)	C^{β}-containing residues with charged and uncharged side chains (excluding Asp^-)	Asp^-
N_{13}	−0.344	−0.285	−0.356	−0.356
$H_2(N)$	0.176		0.176	0.176
C_9^{α}	−0.008	0.050	0.064	−0.060
C'_7	0.450	0.455	0.450	0.450
O_{17}	−0.384	−0.385	−0.384	−0.384
H_1^{α}	0.055	0.040	0.020	0.024

[a] In electronic charge units (ecu). [b] The meaning of the subscripts on the atoms[26b] is defined in Table I of ref 9a.

simplifies the calculations on polypeptides by reducing the number of variables, since there is no need to introduce stretching or bending force constants, and a chain of any amino acid sequence can be generated by using the fixed geometry appropriate to each type of amino acid in the sequence.

Several reviews of structural data on amino acids have already appeared.[4a,30-33] We have modified the standard geometric data of ref 4a, using more recent structural information.[32-40] In Table I, we present the revised backbone bond lengths and bond angles. In nearly all of the published structural studies, except the recent ones using neutron diffraction,[33] the positions of the hydrogen atoms have been only moderately well resolved. For this reason, we have positioned the hydrogen atoms according to a method described previously,[9] and these hydrogen positions will be used here. The criteria used for the choice of bond lengths and bond angles are examined separately in Appendices I and II for each amino acid and end group. However, several general rules were observed. First, structures in which amino acids were complexed with heavy metals were not used because of the possibility that the strained rings in these complexes might lead to abnormal bond lengths and bond angles. Second, crystal structures of hydrohalide ionic species were examined carefully for possible deviations in geometry due to ionic effects; those with local deviations arising from ionic bonding were omitted. Third, structures with low rather than with only moderate indices of resolution (R values) were selected. A description of these problems, as well as the resulting backbone and side-chain geometry, is given in Appendix I for each amino acid.

The backbone bond lengths and bond angles for all the naturally occurring amino acids (except proline-type residues) are given in Table I; data for proline and hydroxyproline are given in Appendix I. It was found that, to the precision reported in Table I, these backbone bond lengths and bond angles (excluding the $NC^{\alpha}C'$ bond angle, which varies from one amino acid to another[41]) are remarkably similar, and thus are fixed at the values given in Table I. On the other hand, the addition of various side chains to the backbone seems to affect the geometry around the C^{α} atom, and significant variations in bond angle at this position are observed. Variations in the other bond angles around the C^{α} atom [viz., in $\tau(NC^{\alpha}C^{\beta})$ and $\tau(C^{\beta}C^{\alpha}C')$] indicate that there is a significant influence of the type of side

TABLE III: CNDO/2 (ON) Partial Charges[a] in Saturated Side Chains

Amino acid	
Ala	C^{β}, −0.090; H^{β}, 0.040
Val	C^{β}, 0.008; H^{β}, 0.016; $C^{\gamma 1,2}$, −0.072; H^{γ}, 0.025
Leu	C^{β}, −0.030; H^{β}, 0.022; C^{γ}, −0.011; H^{γ}, 0.025; $C^{\delta 1,2}$, −0.074; H^{δ}, 0.025
Ile	C^{β}, −0.005; H^{β}, 0.025; $C^{\gamma 1}$, −0.020; $H^{\gamma 1}$, 0.015; $C^{\gamma 2}$, −0.075; $H^{\gamma 2}$, 0.025; $C^{\delta 1}$, −0.075; H^{δ}, 0.025
Met	C^{β}, −0.010; H^{β}, 0.030; C^{γ}, −0.120; H^{γ}, 0.045; S^{δ}, 0.035; C^{ϵ}, −0.190; H^{ϵ}, 0.055
CysSH	C^{β}, −0.105; H^{β}, 0.055; S^{γ}, 0.015; H^{γ}, 0.010
Cys	C^{β}, −0.090; H^{β}, 0.055; S^{γ}, 0.010
Thr	C^{β}, 0.160; H^{β}, 0.015; $O^{\gamma 1}$, −0.310; $H^{\gamma 1}$, 0.170; $C^{\gamma 2}$, −0.095; $H^{\gamma 2}$, 0.030
Ser	C^{β}, 0.130; H^{β}, 0.020; O^{γ}, −0.310; H^{γ}, 0.170
Lys	C^{β}, −0.030; H^{β}, 0.015; C^{γ}, −0.025; H^{γ}, 0.020; C^{δ}, −0.030; H^{δ}, 0.020; C^{ϵ}, 0.110; H^{ϵ}, −0.005; N^{ζ}, −0.395; H^{ζ}, 0.150
Lys^+	C^{β}, −0.030; H^{β}, 0.015; C^{γ}, −0.025; H^{γ}, 0.020; C^{δ}, −0.115; H^{δ}, 0.100; C^{ϵ}, 0.050; H^{ϵ}, 0.120; N^{ζ}, −0.320; H^{ζ}, 0.320

[a] In electronic charge units (ecu).

TABLE IV: CNDO/2 (ON) Partial Charges[a] in Unsaturated Nonaromatic Side Chains

Amino acid	
Arg	C^{β}, −0.030; H^{β}, 0.015; C^{γ}, −0.030; H^{γ}, 0.030; C^{δ}, 0.100; H^{δ}, 0.010; N^{ϵ}, −0.350; H^{ϵ}, 0.170; C^{ζ}, 0.480; $N^{\eta 1}$, −0.420; $H^{\eta 1}$, 0.090; $N^{\eta 2}$, −0.410; $H^{\eta 2}$, 0.160
Arg^{+}	C^{β}, −0.030; H^{β}, 0.015; C^{γ}, −0.030; H^{γ}, 0.030; C^{δ}, 0.120; H^{δ}, 0.010; N^{ϵ}, −0.300; H^{ϵ}, 0.230; C^{ζ}, 0.580; $N^{\eta 1}$, −0.390; $H^{\eta 1}$, 0.280; $N^{\eta 2}$, −0.390; $H^{\eta 2}$, 0.280
Asp	C^{β}, −0.100; H^{β}, 0.060; C^{γ}, 0.500; $O^{\delta 1}$, −0.360; $O^{\delta 2}$, −0.350; $H^{\delta 2}$, 0.220
Asp^{-}	C^{β}, −0.170; H^{β}, −0.020; C^{γ}, 0.500; $O^{\delta 1}$, −0.570; $O^{\delta 2}$, −0.570
Glu	C^{β}, −0.030; H^{β}, 0.030; C^{γ}, −0.110; H^{γ}, 0.055; C^{δ}, 0.500; $O^{\epsilon 1}$, −0.360; $O^{\epsilon 2}$, −0.350; $H^{\epsilon 2}$, 0.210
Glu^{-}	C^{β}, −0.120; H^{β}, 0.020; C^{γ}, −0.170; H^{γ}, −0.040; C^{δ}, 0.500; $O^{\epsilon 1}$, −0.570; $O^{\epsilon 2}$, −0.570
Asn	C^{β}, −0.120; H^{β}, 0.055; C^{γ}, 0.460; $O^{\delta 1}$, −0.375; $N^{\delta 2}$, −0.445; $H^{\delta 2}$, 0.200
Gln	C^{β}, −0.030; H^{β}, 0.020; C^{γ}, −0.110; H^{γ}, 0.050; C^{δ}, 0.470; $O^{\delta 1}$, −0.380; $N^{\delta 2}$, −0.430; $H^{\delta 2}$, 0.185

[a] In electronic charge units (ecu).

TABLE V: CNDO/2 (ON) Partial Charges[a] in Ring-Bearing Side Chains

Amino acid	
Phe	C^{β}, −0.040; H^{β}, 0.025; C^{γ}, 0.020; C^{δ}, −0.015; H^{δ}, 0.010; C^{ϵ}, −0.015; H^{ϵ}, 0.010; C^{ζ}, 0.015; H^{ζ}, 0.005
Tyr	C^{β}, −0.040; H^{β}, 0.025; C^{γ}, 0.020; C^{δ}, −0.010; H^{δ}, 0.010; C^{ϵ}, −0.060; H^{ϵ}, 0.030; C^{ζ}, 0.225; O^{η}, −0.330; H^{η}, 0.165
Trp	C^{β}, −0.030; H^{β}, 0.015; C^{γ}, −0.025; $C^{\delta 1}$, 0.085; $H^{\delta 1}$, 0.000; $C^{\delta 2}$, 0.005; $N^{\epsilon 1}$, −0.280; $H^{\epsilon 1}$, 0.130; $C^{\epsilon 2}$, 0.150; $C^{\epsilon 3}$, −0.080; $H^{\epsilon 3}$, 0.080; $C^{\zeta 2}$, −0.060; $H^{\zeta 2}$, 0.030; $C^{\zeta 3}$, −0.020; $H^{\zeta 3}$, 0.010; $C^{\eta 2}$, −0.005; $H^{\eta 2}$, 0.010
His ($H^{\delta 1}$)	C^{β}, −0.040; H^{β}, 0.015; C^{γ}, 0.055; $N^{\delta 1}$, −0.250; $H^{\delta 1}$, 0.150; $C^{\delta 2}$, 0.080; $H^{\delta 2}$, 0.035; $C^{\epsilon 1}$, 0.190; $H^{\epsilon 1}$, 0.020; $N^{\epsilon 2}$, −0.240
His ($H^{\epsilon 2}$)	C^{β}, −0.075; H^{β}, 0.040; C^{γ}, 0.100; $N^{\delta 1}$, −0.230; $C^{\delta 2}$, 0.054; $H^{\delta 2}$, 0.020; $C^{\epsilon 1}$, 0.182; $H^{\epsilon 1}$, 0.010; $N^{\epsilon 2}$, −0.255; $H^{\epsilon 2}$, 0.144
His^{+}	C^{β}, −0.050; H^{β}, 0.065; C^{γ}, 0.150; $N^{\delta 1}$, −0.200; $H^{\delta 1}$, 0.275; $C^{\delta 2}$, 0.100; $H^{\delta 2}$, 0.135; $C^{\epsilon 1}$, 0.275; $H^{\epsilon 1}$, 0.150; $N^{\epsilon 2}$, −0.200; $H^{\epsilon 2}$, 0.265
	CNDO/2 (ON) Partial Charges in Proline and Hydroxyproline
Pro {up / down[b]	C^{β}, −0.025; H^{β}, 0.015; C^{γ}, −0.050; H^{γ}, 0.025; C^{δ}, 0.100; H^{δ}, 0.010
HPro {up / down	C^{β}, −0.050; H^{β}, 0.020; C^{γ}, 0.172; H^{γ}, 0.006; $O^{\delta 1}$, −0.323; $H^{\delta 1}$, 0.165; $C^{\delta 2}$, 0.075; $H^{\delta 2}$, 0.020

[a] In electronic charge units (ecu). [b] Up and down refer to ϕ values of −67.6 and −75.0°, respectively (see Table VIII).

chain on these bond angles, which are given in Appendix I for each amino acid. The values of $\tau(NC^{\alpha}C')$ for each amino acid are given in Table I.

III. Partial Atomic Charges

The overlap normalized partial charges for each atom of every amino acid residue were determined by using the complete neglect of differential overlap[12,16,42] (CNDO/2) molecular orbital method. The CNDO/2 theory is an SCF method, which treats all valence electrons, and employs zero differential overlap, but includes electron interactions explicitly. We have found previously[42] that the overlap normalized partial charges reproduce dipole moments of model amide molecules only fairly well, giving dipole moments that are generally somewhat less than the experimental values. The primary reason for using the overlap normalized charges is that they are atom centered, and the charges can be positioned without having to consider, in addition, the locations of lone-pair orbitals, i.e., the charges can be placed on the atomic coordinates, and the empirical formalism with a fixed geometry can be retained. Since the values of the CNDO/2 (ON) charges are low, resulting in low gas phase dipole moments, the effective electrostatic energies are correspondingly low and are equivalent to those computed with a dielectric constant of ~2. We introduce an effective dielectric constant of 2 to further reduce the electrostatic energy so that it corresponds to that in a medium with a dielectric constant of ~4, with charges that reproduce the dipole moment; a value of ~4 is an estimate of the dielectric constant of polypeptide crystals (see section VA). The use of an effective dielectric constant is only an approximation that attempts to duplicate the contributions of atomic polarization to the interaction energy. As pointed out earlier,[9] the nonbonded repulsive parameters determined from crystal packing calculations compensate, to some extent, for other inadequacies of the potential function.

The molecules treated here to obtain the atomic charges were the *N*-acetyl-*N'*-methyl amino acid amides. In some cases, the *N*-formyl molecule was used to reduce the number of orbitals in the CNDO/2 program.[42] In every case, it was necessary to investigate several conformations of the molecule under consideration by changing backbone and side-chain dihedral angles to find conformations in which the partial charges are least dependent on conformation. The imposition of this condition was necessary since any arbitrarily chosen conformation might have some steric interaction which would tend to polarize some particular atoms. The low-energy or the unpolarized state of the molecule (i.e., that conformation which has no close interatomic contacts that would tend to induce a polarization in these atoms) is the one sought here, and we have determined those partial charges which best represent this state. Since we do not allow bond stretching or bond-angle bending, a low-energy conformation (by CNDO/2) is found by varying only the dihedral angles, but this is not necessarily the lowest-energy geometry of the molecule according to this molecular orbital method (i.e., the bond lengths and bond angles are not those which would necessarily be the minimum energy ones for the CNDO/2 method). In many cases, the conformations used in determining the charges are, by intention, not the minimum-energy ones for the CNDO/2 method. For example, by the CNDO/2 method, the equatorial seven-membered ring conformation has a hydrogen bond in which the atoms that participate in the hydrogen bond exhibit considerable polarization,[16] due in part to the rather short H···O distance which is usually obtained when the energy is minimized using the CNDO/2 method;[11,16] thus, conformations such as the equatorial seven-membered ring, which contain internal hydrogen bonds, were *not* used to determine the charges. In order to reduce the influence of conformational variation on the atomic charges as much as possible, we have averaged the charges over several different conformations (viz., β and α_R backbone, and several side-chain conformations) for each residue, and then rounded off insignificant differences. We assume that, by using the average values, the resulting atomic charge distribution will apply for all conformations. In most cases, excluding those atoms involved in hydrogen bonding, the variation of charge with conformation (for low-energy conformations) is very small (i.e., ±0.005 electronic charge units) and will not have any appreciable effect on the electrostatic energy calculated by the monopole approximation.

The partial charges of the backbone atoms are given in Table II, and the side-chain charges are given in Tables III–V. The partial charges of the end-group atoms are given in Appendix II. It appears that the backbone charges are constant for several categories of amino acids, viz., glycine, proline, and hydroxyproline, C^{β}-containing amino acids with charged and uncharged side chains, and the charged side chain of aspartic acid. In every case, the total residue (without the acetyl and methyl end groups) has a net charge of zero, but the sum of the charges on the backbone atoms, N, H, C', and O, is not zero. It can be seen from Table II that the addition of a C^{β} atom to residues other than proline affects mainly the charge on the C^{α} atom, with smaller effects on the N_{13} and $H_1{}^{\alpha}$ atoms;[26b] the charges on the carbonyl group and on $H_2(N)$ remain unaffected. However, in the case of proline, the methylene group on the nitrogen does affect the charge on the N_{13} atom, and that on the C'_7 atom only slightly. When the aspartic acid side chain is charged, there is charge migration into the C^{α} atom, but little effect on the other backbone atoms. It appears that the charges of the backbone of all residues containing a C^{β} atom (charged or uncharged), except proline and Asp^-, are nearly equivalent, with possible deviations being small and arising primarily from changes in conformation.

The use of the CNDO/2 method to obtain partial charges is justified further by an examination of the dependence of ^{13}C chemical shifts[43] on CNDO/2 (ON) partial charges of the carbon atoms of amino acids, shown in Figure 1. The range in C^{α} chemical shifts is only ~10 ppm for the C^{β}-containing residues, while C^{β} chemical shifts cover a range of ~50 ppm depending upon the substituent atoms. The very small range of C' chemical shifts (i.e., ~5 ppm) indicates an even lesser influence of residue type, and probably is more dependent on solvent and hydrogen bonding than on residue side chain.

There appears to be a near-linear correlation between the ^{13}C chemical shifts and the CNDO/2 (ON) partial charges for carbon atoms. Since the experimental chemical shifts reflect the true electronic charge distribution, this correlation indicates that the CNDO/2 (ON) charges are a good measure of the electronic charge distribution. A straight line fit to all points intersects the C' data at q = 0.500 and δ = −130 ppm, and has a slope of ~280 ppm/electronic charge unit (ecu), using the overlap normalized

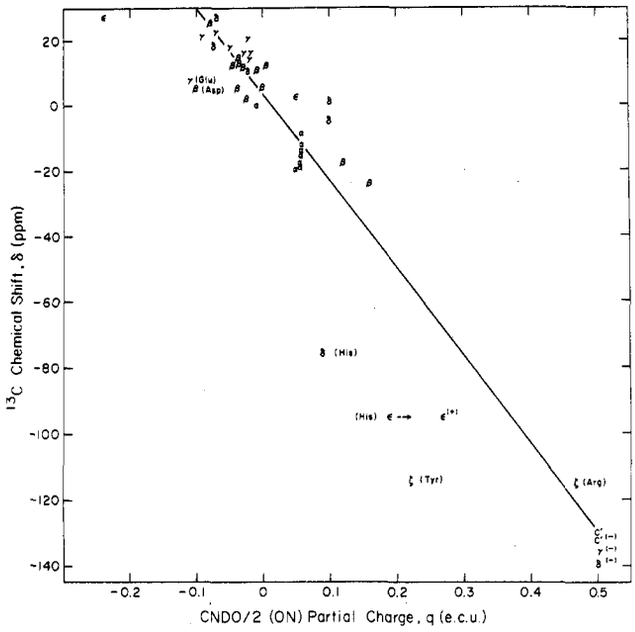

Figure 1. Correlation of ^{13}C chemical shift with CNDO/2 (ON) partial charges for the C^{α}, C^{β}, and other carbon atoms of all the amino acids. Special cases, described in the text, are designated by the name of the amino acid in parentheses. The charged state is indicated by + and − (e.g., His^+, COO^-, etc.).

(ON) charges. The Pople CNDO/2 charges (not shown here) gave a slope of ~370 ppm/electronic charge unit, which is considerably higher than the 150–200 ppm/electronic charge unit estimated theoretically[44] and empirically.[45] The value for the slope of the line in Figure 1 is in close agreement with the slope obtained by Del Re et al.[46] using σ and π charges (i.e., 240 ± 40 ppm/ecu), and the charges obtained here agree fairly closely with those of Del Re et al.[46] and Poland and Scheraga.[47] Some exceptions to the linear correlation are discussed below together with some general remarks that pertain to the atomic charges of the amino acid side chains given in Tables III–V.

(1) The partial charges on the C^{β} atoms in saturated side chains reflect the polarity of the subsequent atoms, being slightly negative when the γ atom is carbon, more negative when the γ atom is sulfur, and becoming positive when the γ atom is oxygen. In the nonaromatic unsaturated side chains, the γ atom also influences the C^{β} charge strongly. As the polar side-chain atoms lie further away from the C^{β} atom, their influence on its charge diminishes rapidly.

(2) The partial charge on the γ atom reflects the covalent character of the bonds linking the polar atoms to it. Two particular cases stand out as deviations from the linear correlation of Figure 1, viz., the C^{β} and C^{γ} atoms of aspartic and glutamic acids, respectively. They are much more negative (by ~0.1 electronic charge unit) than would be expected from the ^{13}C chemical shift data, and become even more negative when the acid side chain is ionized (see Table IV). The negative value is in agreement with that found by Del Re et al.,[46] but not with the chemical shift data. On the other hand, the charges of carboxyl group carbons (viz., C^{δ} of glutamic and C^{γ} of aspartic acid) agree very well with the chemical shift data, indicating a high positive charge ($q \sim 0.5$) in both the uncharged and charged states.

(3) Ring carbon atoms also show departures from the linear relation of Figure 1. For example, the C^{δ} and C^{ϵ} atoms of histidine are shifted to the negative side of the line of best fit. The values for C^{δ} and C^{ϵ} shift to the right (more so for C^{ϵ}) in going from His ($H^{\epsilon 2}$) to His ($H^{\delta 1}$) to His^+ (see Table V). A second example is shown by the C^{ζ} carbon of tyrosine, but we have no explanation for this behavior at the pH (7) of the measurement.

(4) The proton charges also show considerable polarization when attached to the more electronegative atoms. For example, the amine, amide, hydroxyl, and acid protons all have larger positive charges than hydrogens of methyl or methylene groups, showing considerable migration of the s electron of hydrogen toward their respective heavy atoms. Since no charge migration is found from the proton to the sulfur atom in Cys-SH, no large electrostatic contribution to possible S–H···X hydrogen bonds is expected. The methyl and methylene protons are generally found to have small positive charges, the exceptions being the charged side chains, e.g., negative glutamic acid where the H^{γ} atoms become slightly negative.

(5) In the charged side chains, the charge is diffused away from the site where it is formally located. For example, in negative glutamic acid the effect of removing a proton is reflected back to the C^{β} atom, and in Asp^- to the C^{α} atom, thus spreading out the electronic excess over many atoms, resulting in a diffuse charge rather than a charge that is localized on the carboxyl oxygen atoms. This can be seen by summing the carboxyl COO^- charges (of Asp^- or Glu^-) to obtain −0.64 ecu rather than the expected full negative charge on the COO^- group (see also Lys^+ where the sum of the charges on the NH_3^+ group is +0.64). The remaining charge is spread out on the carbon and hydrogen atoms of the side chain. The effect of charged side chains on conformation has been examined in studies of charged polyamino acids.[48]

(6) The charges on the sulfur atoms in a disulfide bond exhibit a small variation with S–S dihedral angle, viz., −0.020, 0.006, and −0.030 ecu at 0, ±90, and 180°, respectively. A value of 0.01 ecu (see Table III) was selected for calculations on proteins since χ^{SS} is usually around ±90°.

The variation in charge with conformation, for changes in dihedral angles around the minimum-energy conformations, is usually found to be small (in the *N*-acetyl-*N'*-methyl amides), and supports our contention that static partial charges may be used in the empirical potentials developed here. When the interatomic contacts are of high energy, the partial charges may change. However, these conformational regions are generally not of interest, and need not be considered further. When hydrogen bonding contacts occur, they may be a significant change in the partial charges of the atoms involved; however, the use of a special hydrogen bonding function simulates both this and other polarization effects.

When there is a large rotation about bonds involved in resonance hybrids, a variation in the partial charges occurs. For example, we have shown previously[16] that a large rotation about the peptide bond (i.e., $\Delta\omega > 40°$) is accompanied by large changes in partial charges of the atoms concerned. The barrier to rotation is very high in this case, and causes the planar conformers (either trans or cis) to be preferred. Small variations of the dihedral angle ω causes little or no variation in charge, and we will use fixed charges for planar and nearly planar peptide groups; the charges on the atoms of the peptide group for both the trans and cis conformations were found here to be the same.

When the CNDO/2 (ON) charges, which pertain to mole-

cules in vacuum, are used for molecules in condensed media, some charge polarization (neglected in empirical computations on polypeptides) may occur. Errors from the omission of polarization are partially reduced because the electrostatic energy decreases when a dielectric constant greater than 1 is used in the computations, and (in the case of hydrogen bonding polar groups) because the general hydrogen bond term (see section VC) includes a contribution from such polarization.[11]

IV. Nonbonded Interactions

In treating the nonbonded intramolecular interactions between pairs of atoms in a polypeptide chain, we have used the Lennard-Jones "6–12" potential function; the parameters for this function were determined and listed in Table IV of ref 9a, and tested on crystal packing of amino acids in ref 9b, and will not be repeated here. These parameters were obtained from crystals in which the molecules were not allowed to undergo bond stretching, bond angle bending, rigid body translational or rotational motion, or internal rotation about single bonds (except for rotations of groups, such as methyl, to position hydrogens). (There is a typographical error in Table II of ref 9a, which appeared only in the typescript and, hence, does *not* affect the computed results; the polarizability of S_{20} should be 0.504×10^{-24} instead of 0.34×10^{-24}.)

A. Special Nature of 1–4 Type Interactions. In order to use the *intermolecular* potentials derived from crystals to compute *intramolecular* energies of polypeptides, when rotations about single bonds are the important degrees of freedom, it must be recognized that, in contrast to 1–5 and higher interactions, the potential energy of interaction of two atoms in a 1–4 interaction (where the atoms are connected by three bonds, the central bond being the one about which rotation may take place), will be dominated by quantum effects such as overlap, exchange, and coulombic contributions. Also, the contributions from molecular translation and rotation to the repulsive potential (inherent in the potentials derived from molecular crystals[9a]) will be smaller for 1–4 interactions than in the case of interactions between atoms separated by more than three bonds about which rotation can take place; i.e., when there are many rotational degrees of freedom between interacting atoms, the atoms will appear to each other to be independent bodies, and should interact similarly to the way they do when they are parts of different molecules in molecular crystals.

There are no data available from which one might compute the energy of a 1–4 interaction, even though this energy is included in experimental barriers to internal rotation. However, such barriers cannot be separated easily into contributions from the interactions between the two atoms in positions 1 and 4 and from interactions between bonding electrons (the so-called intrinsic torsional potential). Therefore, we have chosen to obtain these two separate contributions by using theoretical results to scale the repulsive interaction energy between atoms 1 and 4 and by using intrinsic cosine functions to reproduce the experimental barriers.

The range in which close repulsive terms (i.e., overlap and exchange) contribute significantly is of the order of <5 Å, and is related directly to the quantum mechanical overlap integrals or repulsive potentials for atoms as determined from Hartree–Fock (HF) or Thomas–Fermi–Dirac (TFD) calculations.[49–51] From a theoretical point of view, Bonham[52–54] showed that, to a good approximation, interatomic repulsive potentials (from HF and TFD calculations) between two neutral atoms can be fit to analytical functions, and repulsive force constants (which are in good agreement with those found expgrimentally in Urey–Bradley and other force fields for molecules) can be obtained by differentiation of these analytical functions. We will compare Bonham's results[52–54] with our calculated repulsive force constants, and introduce an approximate empirical factor to scale our empirical repulsive coefficients for 1–4 interactions. In cases where experimental data are available, the resulting potentials are calibrated (also, with the inclusion of intrinsic tortional potentials) using known *experimental* barriers to rotation (see section VI).

Before going into detail about these repulsive terms (in section B), several characteristics of long-range attractive interactions must be discussed. Dispersion energies are usually derived for two isolated and separated atoms or molecules (sufficiently far apart (usually >5 Å) so that the electronic orbitals between the molecules do not interact, or interact only very weakly). In polypeptides and proteins, we must consider interactions between different parts of the *same* molecule. If this molecule were partially conjugated and rigid, it would be difficult to separate out the various contributions to the energy, on the basis of interactions between pairs of atoms, since the π system would extend over the complete molecule. However, because of the separation of peptide groups (and conjugated ring side chains) by saturated single bonds (e.g., N–C^{α} and C^{α}–C′ in the backbones of polypeptides), it is reasonable to apply the Slater–Kirkwood[3] treatment of dispersion forces, and the use of monopole charges for coulombic forces, between atoms sufficiently separated from one another so that overlap is negligible. It seems reasonable to assume that, in the case where no intramolecular hydrogen bonding occurs, e.g., ~4 Å or greater, or even where the neighboring atoms are separated by at least four or more bonds, the electrostatic forces may be calculated in the usual manner. For atoms closer than a typical van der Waals distance (i.e., <4 Å), the repulsive forces should begin to dominate or, at the very least, the electron clouds should be overlapping, and electronic effects of a short-range nature will become important. In calculating the variation in charge with conformation, we found that the charges are perturbed only slightly (~0.01 ecu) at the normal equilibrium contact distances between atoms (excluding hydrogen-bonding interactions). Thus, it would appear that our static monopole charges (which include contributions from both σ and π electrons) will represent the charge distribution adequately at the equilibrium interatomic distances. If the polarizability (assumed to be isotropic) is unchanged at these contact distances, then the perturbation treatment of dispersion forces should also be valid in this region. In the case of hydrogen bonding, a hydrogen-bond function (previously referred to as a "general hydrogen bond" function, GHB,[9,11] but designated here as HB) is used instead of the Lennard-Jones potential, as was described previously.[9,11] It is certainly true that the flexibility of polypeptide molecules, for which this set of potential functions is being developed, allows part of the chain to fold back upon itself, thus bringing remote atoms into contact with one another. However, it seems reasonable to assume that, in these cases, those nonbonded repulsive terms found in the crystal studies,[9] and the associated attractive terms, should describe these

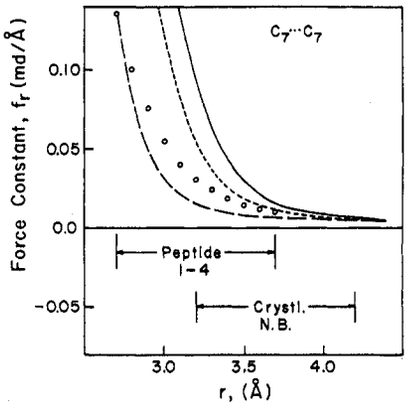

Figure 2. Dependence of the force constant, f_r, on interatomic distance for C···C type interactions. The solid line was obtained from the repulsive coefficient[9] $A_{C_7 \cdots C_7}$ for $C_7 \cdots C_7$ type interactions. The large- and small-dashed lines were obtained by multiplying $A_{C_7 \cdots C_7}$ by 0.2 and 0.6, respectively. The circles were obtained from HF and TFD potentials, both of which give the same force constants for the interaction between two neutral isolated carbon atoms. The range of $C_7 \cdots C_7$ (nonbonded) contact distances found in a variety of molecular crystals[9] is denoted as "Crystl. N.B.". The range of $C_7 \cdots C_7$ distances in a polypeptide chain (when it is a 1–4 interaction, and determined by variation of ϕ, if ω is fixed) is designated as "Peptide 1–4".

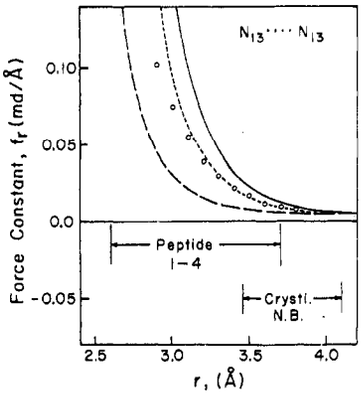

Figure 3. Same as Figure 2, but for $N_{13} \cdots N_{13}$ type interactions (where the distance between successive N_{13} atoms in a polypeptide depends on ψ).

interactions adequately [the assumption being that there are a sufficient number of dihedral angles between the interacting atoms (in 1–5 and higher interactions) so that they behave effectively as independent bodies even though atoms in 1–5 and higher interactions may approach each other as closely as a van der Waals contact].

B. Calibration of 1–4 Interactions. In order to calibrate the empirical potentials obtained from calculations on crystals,[9a] the HF and TFD repulsive force constants, f_r, were calculated, using the analytical functions derived in ref 52 and 53, for all the atoms of interest in this work (viz., H, C, N, O, and S). For example, the HF and TFD force constant curves (which are identical for neutral isolated carbon atoms) are shown in Figure 2 together with that calculated from the empirical repulsive $A_{C_7 \cdots C_7}$ coefficient determined from crystal data.[9a] For over half of the range of $C_7 \cdots C_7$ contact distances found in crystals, the force constants computed from the $A_{C_7 \cdots C_7}$ coefficient agree fairly well with the HF and TFD values. However, at shorter distances (especially in the range where, e.g., the $C_7 \cdots C_7$ 1–4 distances in a polypeptide lie), the HF and TFD repulsive

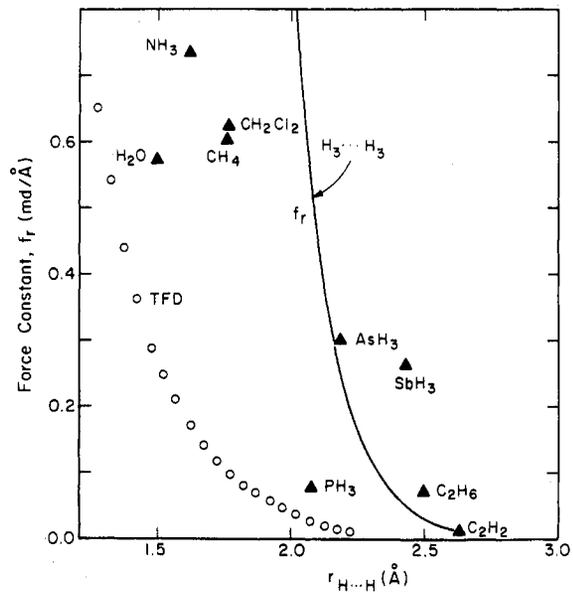

Figure 4. Dependence of the force constant, f_r, on interatomic distance for H···H type interactions. The solid line was obtained from the repulsive coefficient[9] $A_{H_3 \cdots H_3}$. The open circles were obtained from the TFD potential. The filled triangles represent nonbonded force constants for hydrogen atoms, obtained[55,56] from Urey–Bradley type force constant calculations.

force constant curve[52,53] is considerably softer (i.e., is less at any distance) than that calculated from the $A_{C_7 \cdots C_7}$ coefficient. In order to characterize the difference quantitatively, force constant curves, in which $A_{C_7 \cdots C_7}$ is multiplied by a factor of 0.2 and 0.6, respectively, are also shown in Figure 2. It can be seen that a factor of about 0.5 will produce the best fit to the HF and TFD force constant curves in the range of $C_7 \cdots C_7$ 1–4 contacts found in polypeptides. Force constant curves are also shown in Figure 3 for $N_{13} \cdots N_{13}$ type interactions. In this case, the theoretical (HF) curve lies just below that for which $A_{N_{13} \cdots N_{13}}$ is multiplied by 0.6, over the range of 1–4 distances between $N_{13} \cdots N_{13}$ type atoms in a polypeptide chain. In a similar manner, we investigated the force constant data for all pair combinations of heavy atoms (i.e., C, N, O, and S), and found that a factor of 0.5 gives reasonable agreement with (HF) theory in almost all cases.

The TFD theory is not applicable for interactions involving hydrogen atoms (it being a good representation of the electron density only for heavy atoms[52]), and the HF data were not well fit[53] by the analytical functions used.[53] Hence, we have obtained some reasonable estimates of H···H (and H···C[9a]) short-range repulsive force constants from an analysis of Urey–Bradley force fields of various molecules. Force-constant data of Bartell[55] and Mizushima[56] for gas-phase H···H intramolecular contact distances observed in hydrogen-containing molecules are presented in Figure 4. It is apparent that the TFD curve for H···H repulsive force constants lies to the left of most of the Urey–Bradley force constants, and the curve obtained from the second derivative of the $H_3 \cdots H_3$ nonbonded potential[9a] (and similarly for $H_1 \cdots H_1$, not shown here) is close to the data for those molecules where the hydrogens are ~2.2 Å or greater apart, but rises too steeply at shorter contact distances. While a scaling factor of 0.5 is assigned to the repulsive coefficient for 1–4 type H···H interactions, the characteristics of 1–4 type H···H nonbonded interactions are gen-

erally accounted for by the empirical intrinsic torsional term based on experimental barriers for internal rotation. These results are discussed in section VI.

It should be noted that a softer repulsive potential function (e.g., an exponential or a lower power, such as the $1/r^9$ function used by other authors[27]) would tend to fit the theoretical (HF and TFD) force constant data at the shorter contact distances and also, most probably, in the range of intermolecular contact distances found in crystals. We have resorted to the use of a scaling factor, F, for the *repulsive* coefficient for 1–4 interactions in order to compensate for the vibrational contributions to the $1/r^{12}$ term (inherent in the results of the crystal studies[9]), and we have shown[17] that this approach reproduces observed experimental conformational data.

V. Total Intramolecular Energy Terms

In these calculations, the total interaction potential energy between atoms has been partitioned into electrostatic, nonbonded repulsion, dispersion attraction, and hydrogen-bond terms. In section VI, we will show that those *conformational* properties (viz., barriers to rotation) which are not fully accounted for by the foregoing contributions to the intramolecular energy can be reproduced correctly by the addition of a further, simple one-term intrinsic torsional potential.

A. Electrostatic Energy. Atom-centered monopole charges (which are chosen to represent the continuous charge distribution) are used for all electrostatic interactions. All 1–4, 1–5, and higher type interactions are included, and the potential energy for any pair of atoms is calculated from

$$U_{el}(r_{ij}) = 332.0 q_i q_j / D r_{ij} \quad (1)$$

where q_i and q_j are the partial charges (see section III and Appendix II). The distance between interacting atoms, r_{ij}, is specified by the atomic coordinates; 332.0 is the conversion factor to give $U_{el}(r_{ij})$ in units of kcal/mole when r_{ij} is in Ångström units and q is in units of electronic charge. The "effective dielectric constant", D, is taken[11] as 2 in all calculations used here. The value of 2 for D was obtained by analysis of CNDO/2 calculations[11] on hydrogen-bonded dimers. Since the CNDO/2 (ON) partial charges lead to dipole moments which are smaller than the experimental values, the use of an "effective dielectric constant" $D = 2$ is operationally equivalent to a macroscopic value of $D \sim 4$ to 8, and is probably close to the experimental dielectric constant for polypeptides in polar media.[57,58] This "dielectric constant" is only marginally applicable for the 1–4 electrostatic terms since the use of point charges at the short distances over which 1–4 interactions occur is not rigorously correct. Also, while one might expect a value of $D = 1$ to apply to 1–4 cis interactions, a larger value should apply to 1–4 trans interactions (i.e., those "through" a bond). Since we have no way of evaluating this effect, the value of $D = 2$ as the "effective dielectric constant" was retained for all interactions.

B. Nonbonded Energy. The form of the nonbonded repulsion and dispersion attraction energies is taken to be

$$U_{NB}(r_{ij}) = F A^{kl}/r_{ij}^{12} - C^{kl}/r_{ij}^{6} \quad (2)$$

for any two atoms, where A^{kl} is the repulsive coefficient determined from the crystal calculations,[9a] and C^{kl} was calculated[9a] using the Slater–Kirkwood formalism[3] (see Table IV of ref 9a). F is a factor whose magnitude was determined by consideration of the repulsive force constants and taken to be 0.5 for 1–4 type interactions, and 1.0 for all other type interactions as described in section IVB). The atom types are designated as k and l (see Table I of ref 9a),[26b] and specify which nonbonded coefficients are to be used in eq 2.

C. Hydrogen-Bond Energy. Interactions between those atoms taking part in a hydrogen bond [viz., atoms $H_2(N)$, $H_4(O)$ to atoms O_{17}, O_{18}, N_{13}, or N_{14}] are calculated using a hydrogen bond function[11] of the type

$$U_{HB}(r_{H\cdots X}) = A'_{H\cdots X}/r_{H\cdots X}^{12} - B_{H\cdots X}/r_{H\cdots X}^{10} \quad (3)$$

where $A'_{H\cdots X}$ and $B_{H\cdots X}$ are specific coefficients for different combinations of hydrogen-bonding atoms (see Table V of ref 9a and Table I of ref 9b), and are taken without modification for all types of interactions (i.e., 1–4, 1–5, and greater). A recent study[59] of hydrogen bonding, using the CNDO/2 method and model compounds, similar to our studies of ref 11, has led to an alternative functional form for the H···O hydrogen-bond term than the one used here. However, the 10–12 form, which was calibrated using crystal data,[9a] and which was shown to reproduce experimental crystal packing,[9] has been retained here.

D. Torsional Energy. In selected cases, the energy associated with the barriers to rotation is not completely accounted for by the above energy contributions. In these cases (described in section VI), a torsional contribution of the type

$$U_{TOR}(\theta) = (U_0/2)(1 \pm \cos n\theta) \quad (4)$$

is introduced, where U_0 is the height of the n-fold barrier. In order to obtain the correct value for U_0, all other interactions which contribute to the torsional energy must be calculated first, and the difference between the total interaction energy and the experimentally observed (or theoretical) rotational energy is assigned to U_0 (see section VI). In all cases examined here (section VI), a single cosine term was found to be sufficient to obtain a good fit to experimental (or theoretical) data.

E. Total Energy. The total interaction energy is the sum of the terms described above over all atom pairs of the 1–4, 1–5, or higher types, and the torsional terms (see section VI). We repeat here our statement in ref 9a that it is only the *total* potential that has any meaning for conformational-energy calculations on polypeptides, and thus one should not mix individual parts of our total potential with the complementary parts of the total potential of other authors.

F. Designation of Type of Interaction. The calculation of interactions between atoms is carried out in the following manner.

(a) Interactions between atoms whose relative positions are not dependent on rotations either around single bonds or about the peptide bond are not calculated.

(b) Interactions between atoms separated by at least three bonds, and for which there is only one bond about which rotation can take place, are designated as 1–4 interactions. For example, in Figure 5, the 1–4 interactions are those between the following pairs of atoms: 1, 4; 1, 5; 1, 6; 2, 7; 2, 8; 3, 9; 3, 10; 4, 7; 4, 8; 5, 7; 5, 8; 7, 9; and 7, 10. In each of the interactions listed (except the HB term for 7, 9) the A^{kl} coefficient is multiplied by 0.5. Pairs such as 3, 9 and 3, 10 involve rotation about the peptide bond; in the general scheme presented here, such rotation is allowed.

(c) Interactions between atoms separated by four or more bonds, and where rotation can occur around at least

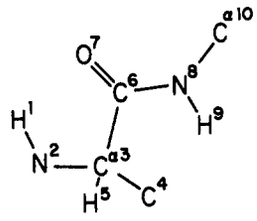

Figure 5. Schematic numbering scheme for the atoms of the backbone of a polypeptide chain.

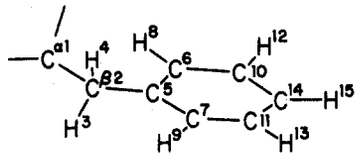

Figure 6. Schematic numbering scheme for the atoms of the phenylalanine side chain.

two bonds, are calculated using the nonbonded terms obtained from the crystal calculations,[9a] with $F = 1.0$. In the example shown in Figure 5, $F = 1.0$ for atoms pairs 1, 9; 1, 10; 2, 10; 4, 9; 4, 10; 5, 9; and 5, 10, while 1, 7; 1, 8; and 2, 9 are HB type interactions (i.e., eq 3 is used to calculate the energy). Some particular cases are described below to illustrate these definitions. The phenyl ring of phenylalanine and tyrosine is treated as rigid; thus all interactions, as shown in Figure 6, between atoms 1, (6 through 15) and similarly 3, (6 through 15) and 4, (6 through 15) are treated as 1–4 type interactions (i.e., there is only one degree of rotational freedom between these atoms). No interactions between atoms in the phenyl ring are included since there is no movement of one atom with respect to another. It is also apparent that atoms, such as C^{10} and H^{12}, are far enough from H^3 or C^1 so that the nonbonded repulsive term is negligible. Similar interactions occur in the proline residue. For example, in Figure 7, the atom pairs 1, [(5 through 8) and (10 through 17)] are all treated as 1–4 type; similarly with atom 3, whereas $F = 1.0$ is used for interactions between 1, 9 and 3, 9. Atom 2 is special since the bond angles defined by atoms 2, 4, 5 and 2, 4, 6 differ for the trans and cis peptide (as described in Appendix I), and its interaction with the proline ring atoms is included in the internal energy of the proline ring (whose energy also changes when the C^γ is up or down, or when the ring is planar); however, this energy is not calculated explicitly each time, but is included as a constant depending on the proline conformation used (see section P of Appendix I). A hydrogen bond (HB) term is included in place of the nonbonded term for interactions between atoms $H_2(N)$, $H_4(O)$, and O_{17}, O_{18}, N_{13}, and N_{14}. For example, in Figure 8, the following interactions are taken as 10–12 HB terms, rather than as 6–12 nonbonded: atoms $1(H_2)\cdots7(O_{17})$, $1(H_2)\cdots8(N_{13})$, $1(H_2)\cdots13(O_{17})$, and $1(H_2)\cdots14(O_{18})$, $2(N_{13})\cdots9(H_2)$, $2(N_{13})\cdots15(H_4)$, $7(O_{17})\cdots9(H_2)$, $7(O_{17})\cdots15(H_4)$, $9(H_2)\cdots13(O_{17})$, $9(H_2)\cdots14(O_{18})$, and $13(O_{17})\cdots15(H_4)$; i.e., these interactions contribute to hydrogen bond formations, as described in previous papers.[9,11] The $1(H_2)\cdots8(N_{13})$ interaction (Figure 8) was not discussed previously in ref 9, but is discussed in ref 19 and 20, and is taken to be the same as the $H_2(N)$ to N_{14} type HB.

The atom numbers introduced in Figures 5–8 are used here only for illustrative purposes, and are not those which are reported in the discussion of the computer programs

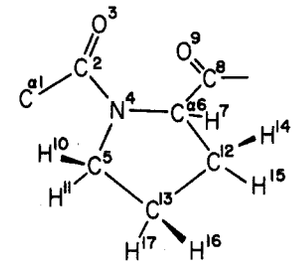

Figure 7. Schematic numbering scheme for the atoms in a section containing a proline residue.

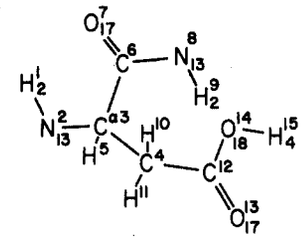

Figure 8. Schematic numbering scheme and types for the atoms in a section containing an aspartic acid residue. The atom types 2, 13, 17, and 18 are indicated by subscripts.

used to carry out these energy calculations.[60] We have not discussed here all the interactions for each amino acid. However, by using the examples shown above, the interaction array for any other amino acid may be developed easily. A listing of coordinates, from which bond lengths and bond angles for the amino acids may be derived, parameters of the potential functions, types of arrays, and a computer program written in Fortran IV are available;[60] the numerical values are an updated version of those reported in ref 17.

VI. Torsional Potentials

Despite the fact that several theoretical studies[61-72] have been carried out in an attempt to identify the principal energy components responsible for the barrier to internal rotation about single bonds, the origin of this barrier is not well understood. Therefore, in cases where these are not adequately represented by the empirical electrostatic, nonbonded, and hydrogen-bonding contributions discussed in sections VA–VC, we have used experimental data[73] to calibrate intrinsic torsional barriers. In those cases, where experimental data on barrier heights or periodicities of the intrinsic torsional potentials are not available, we have used the results of EHT or CNDO/2 calculations to estimate these quantities. The results for the backbone and side chains of various amino acids follow.

A. Internal Rotation about the Backbone Bonds. The most stable conformation of the peptide group in proteins and model compounds (excluding cyclic peptides and *N*-methylated amino acids) has been established to be the planar trans form, by many X-ray and neutron diffraction studies[32] and more recently by gas-phase electron diffraction.[36,37] The barrier to rotation about the peptide bond has also been investigated both theoretically[42] and experimentally (see ref 42 for references to experimental barriers), and the potential function was found[42] to have twofold symmetry with a barrier height between 14[74] and 25 kcal/mol.[42] We use an intrinsic torsional term of the form

$$U(\omega) = (U_\omega/2)(1 - \cos 2\omega) \qquad (5)$$

TABLE VI: Comparison of Calculated and Experimental Rotational Barriers for Model Compounds

Molecule	U_0[a]	Barrier height, kcal/mol		Rel. conformational energy, kcal/mol		Dihedral angle[c] of stable conformers, deg	
		Expt[b]	Calcd	Expt	Calcd	Expt	Calcd
Ethane (C–C)[d]	2.7	2.9	2.9			60	60
Methanol (C–O)[d]	0.6	1.07	0.8			60	60
Ethanol (C–O)[d]	0.6	0.8	0.8			60	60
Methylamine (C–N)[d]	1.8	1.98	2.0			60	60
Propane (C–C)[d]	2.7	3.3	3.0			60	60
n-Butane (C–C)[d]	2.7	~3.8 (T→G)	3.8 (T→G)	0 (T)	0 (T)	180 (T)	180 (T)
		~7[e] (G→G)	31 (G→G)	0.96[f] (G)	1.50 (G)	67.5[g] (G)	68 (G)
1-Propanol (C–C)[d]	2.7		3.6 (G→T)	0 (G)	0 (G)	60 (G)	60 (G)
			4.3 (G→G)	0.3[h] (T)	0.3 (T)	180 (T)	180 (T)
n-Propylamine (C–C)[d]	2.7		3.3 (G→T)	0 (G)	0 (G)	60 (G)	62 (G)
			5.6 (G→G)		0.6	180 (T)	180 (T)
2-Aminoethanol (C–C)[d]	2.7	3.4[i] (G→T)	3.2 (G→T)	0 (G)	0 (G)	55[i] (G)	60 (G)
			3.0 (G→G)		0.1 (T)		180 (T)
Ethylene glycol (C–C)[d]	2.7		3.3 (G→T)	0 (G)	0 (G)	74[j] (G)	50 (G)
			2.3 (G→G)	>0.7[k] (T)	0.3 (T)		180 (T)
Ethylenediamine (C–C)[d]	2.7		3.5 (G→T)	0 (G)	0 (G)	64[l] (G)	53 (G)
			2.6 (G→G)		0.8 (T)	180 (T)	180 (T)

[a] U_0 is the height at the maximum of the intrinsic torsional potential for rotation around the C–C, C–N, and C–O bonds (see eq 6). [b] Reference 73, unless indicated otherwise. [c] For the disubstituted ethanes, the dihedral angle around *only* the central C–C bond is reported. The other dihedral angles were optimized until they were in their lowest energy conformation. [d] U_0 refers to this bond, in all cases. For those molecules containing not only C–C but also a C–O or C–N bond or both, the value of U_0 and the conformational data are given only for rotation about the C–C bond. The values of U_0 for the C–O and C–N bonds are taken as those listed for methanol and methylamine, respectively. [e] Reference 76. [f] Reference 77. [g] Reference 78. [h] Reference 79. [i] Reference 80. [j] Reference 81. [k] Reference 82. [l] Reference 83.

for variation of the peptide bond dihedral angle ω, where U_ω is taken to be 20 kcal/mol, in agreement with our earlier studies.[42,75] The value of 20 kcal/mol is in agreement with most of the experimental data cited in ref 42. The use of eq 5, together with the other contributions to the total energy, makes the trans peptide favored over the cis by several kcal/mole (the value of the energy difference depends somewhat on the values of ϕ and ψ on each side of the peptide bond); in this calculation, it is assumed that the six atoms which constitute the peptide group all lie on one plane when $\omega = 0$ or 180°, i.e., there is no puckering around the C′ or N atoms.

As far as the torsional energy contributions to rotation about the C′N–C$^\alpha$C′ (ϕ) and NC$^\alpha$–C′N (ψ) bonds are concerned, experimental data on the lowest vibrational states of molecules containing no hydrogen bonds would be of considerable help; unfortunately, such data are not available at present. Attempts have been made[27] to gain some insight into these torsional terms (from vibrational analysis), but at present resort must be had to such internal rotation in model compounds[42] to obtain information about them.

We have previously described the results of molecular orbital calculations on model compounds such as *N*-methylformamide, acetamide, and others,[42] in which the rotations of methyl groups (representing the equivalents of variations of ϕ and ψ) were investigated. In these cases, we found very small torsional barriers, and this result was in agreement with the conclusions of Scott and Scheraga,[3] who described various model compounds and the available experimental rotational barriers. In the absence of good experimental data or reliable molecular orbital calculations, we assume, in contrast to the procedure recommended earlier in ref 4a, that there is no intrinsic torsional contribution to rotation about the N–C$^\alpha$ and C$^\alpha$–C′ bonds of the backbone; i.e., such rotations are represented adequately by the interatomic interactions across these bonds. The results presented in paper VI[17] of this series (based on the potentials reported here, *without* intrinsic torsional terms for ϕ and ψ) gave reasonable agreement with the few available experimental data cited there.[17]

B. Internal Rotation about Side-Chain Bonds. In contrast to the treatment of ϕ and ψ, where no intrinsic torsional potentials were used, side-chain rotations (involving variations of χ's) do require torsional potentials to match experimental rotational barriers. It is assumed here that the total barrier to rotation in each side chain consists of all the 1–4 and higher nonbonded, HB, and electrostatic interatomic contributions across the bond about which rotation takes place, as well as a contribution characteristic of the bond itself, i.e., an intrinsic torsional term. The use of this assumption is not intended to imply any knowledge about the physical origin of the barrier to rotation, but only to reconcile the total energy with experimental observations that depend on this barrier. For example, in paragraphs 1–12 of section VIB, we will see that it will be necessary to introduce a contribution that is characteristic of most aliphatic (sp^3 hybridized) C–C bonds (saturated hydrocarbons being similar to the saturated amino acid side chains). The magnitudes of U_0 and n of eq 4 are evaluated by first calculating all other contributions to the particular barrier under consideration, and then adding an intrinsic torsional term to bring the total energy into agreement with experimental (or, in some cases where no experimental data are available, other theoretical) data.

The compounds used to obtain the torsional terms are ethane, methanol, ethanol, and methylamine. The torsional terms were then tested on propane, *n*-butane, 1-propanol,

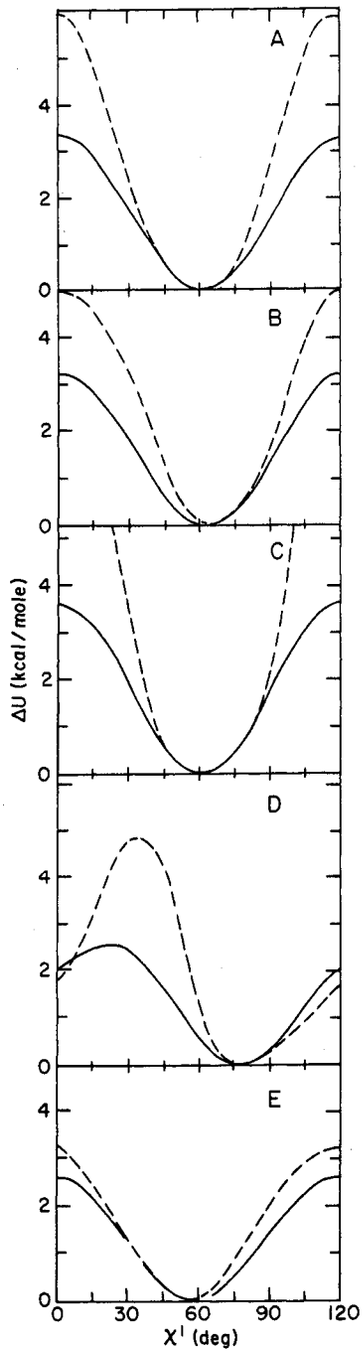

Figure 9. Energy for variation of χ^1 in *N*-acetyl-L-alanine-*N'*-methylamide, for various backbone conformations: (A) α_R ($\phi = -50°$; $\psi = -40°$); (B) α_L ($\phi = 50°$; $\psi = 40°$); (C) equatorial seven-membered ring ($\phi = -70°$; $\psi = 70°$); the value of ΔU at $\chi = 0°$ is 7.7 kcal/mol; (D) axial seven-membered ring ($\phi = 70°$; $\psi = -70°$); and (E) extended ($\phi = -170°$; $\psi = 170°$). The solid and dashed curves are the EHT and total empirical (including torsional term) results, respectively. Each curve was normalized to zero at its own minimum.

n-propylamine, 2-aminoethanol, ethylene-1,2-diol, and ethylenediamine. The partial charges (overlap normalized) for these compounds were obtained by CNDO/2 calculations. For the first set of molecules, the energy difference (including 1–4 and 1–5 nonbonded, hydrogen bonding, and electrostatic energies) between the eclipsed and staggered forms was computed. In order to reduce the difference between this calculated energy and the corresponding experimental value as much as possible, a torsional term was introduced; the values used for U_0 are shown in Table VI. These torsional terms were then tested in a similar manner, on the disubstituted ethanes, except that the total energy was minimized with respect to rotations of the end groups at each value of the dihedral angle, θ, of the central C–C bond to obtain the dependence of the total energy on θ. In all cases the value of n was taken to be 3 for torsions about C–C, C–N, and C–O single bonds. The results of the application of these parameters to some disubstituted ethanes are also presented in Table VI. The rotational energies about all C–C bonds are well reproduced when eq 6 is included

$$U(\chi^i) = [U_0(\chi^i)/2][1 + \cos n\chi^i] \quad (6)$$

where $U_0 = 2.7$ kcal/mol and $n = 3$ (these values are used for all side-chain saturated C–C bonds). Similarly, the C–O (hydroxyl) and C–N (amine) torsional terms (with U_0 from Table VI, and $n = 3$) are calculated, using eq 6 for those amino acid side chains in which these groups occur.

The bond lengths and bond angles used for these model compounds were the same as those described for similar groups in amino acids (see Appendices I and II). Slight differences between this "standard" geometry and the experimental values for each of the compounds of Table VI do not affect the conclusions drawn from the calculations.

1. Alanine. Alanine is the simplest amino acid with a rotatable side chain. In computing empirical energies for alanine, and all other amino acids, the scaling factor $F = 0.5$ was applied in all 1–4 type interactions (including those involving hydrogens), and the torsional potential for the side-chain rotation (described in eq 6) was included. The empirical and EHT energies as a function of χ^1 for five different backbone conformations are shown in Figure 9; i.e., when χ^1 is varied, only the hydrogen atoms of the β carbon of *N*-acetyl-L-alanine-*N'*-methylamide move. Each curve was normalized to zero at its own minimum; since our interest is only in the positions of the minima and of the low-energy maxima, the absolute energies are not relevant here. Since no experimental data on the barrier to rotation or the low-energy conformation of the methyl side chain of alanine are available, we have used the EHT results for comparison with the empirical results (which contain the torsional term of eq 6, based on ethane-type molecules). The axial seven-membered ring is a high-energy conformation,[17] except for glycine, and the large deviations between the empirical and EHT curves (Figure 9D) have no consequences, as far as their applicability to the location of most probable conformations of polypeptides is concerned. In the remaining four backbone conformations (Figure 9A, B, C, E), which are low-energy but not necessarily minimum-energy ones,[17] the empirical EHT curves agree very well in the neighborhood of the minima, but deviate (principally in barrier height) in the regions of the maxima; i.e., the empirical energies give higher barrier heights. (This result is generally true for variation of χ^1 in all amino acids, but has no bearing in calculations on polypeptides.) To understand the reason for the disagreement in the barrier heights, it may be recalled that the EHT method is a molecular orbital method that expresses overlap in an exponential form which tends to be softer than the $1/r^{12}$ form used here. This same behavior was observed (section IVA) in the differences in the force constants obtained with TFD and HF potentials, on the one hand, and with empirical potentials on the other. The torsional term (eq 6) helps to bring the minima of the empirical and EHT curves into good agreement,

as it did with empirical and observed barriers of the model compounds of Table VI. The resulting discrepancies at the maxima are of no consequence since the maxima occur in high-energy regions of conformational space.

2. Valine. The empirical and EHT energies as a function of χ^1 (with $\chi^{2,1} = \chi^{2,2} = 60°$) for various values of ϕ and ψ of *N*-acetyl-L-valine-*N'*-methylamide are shown in Figure 10. The torsional potential of eq 6 was included in the empirical energy. In both calculations, the positions of the minima are nearly identical, although at higher energy (i.e., ~5 kcal above the lowest-energy minimum) the empirical energy rises faster than the EHT energy, as in the case of alanine.

The dependence of the energy on $\chi^{2,1}$ and $\chi^{2,2}$ (not shown here) is similar for both empirical energies (including eq 6) and EHT energies. The variations in $\chi^{2,1}$ and $\chi^{2,2}$ are strongly coupled through the hydrogen atom interactions (these interactions, rather than the intrinsic torsional potential, dominate the formation of low energy conformations), so that both must be varied simultaneously to obtain a low-energy side-chain conformation for any given backbone conformation.

3. Leucine and Isoleucine. The energy for rotation about the single bonds of the nonpolar side chains of leucine and isoleucine are similar to those for methylene–methylene interactions in butane and propane (see Table VI), and the dependence of the energy on χ^1, χ^2, etc., is not shown here. The barrier for variation of χ^1 in leucine (for low-energy backbone conformations) is ~2.0 kcal/mol (and threefold) without a torsional term, similar to alanine and χ^1 of valine. However, as was found for valine, the empirical energy minima (for χ^1) agree more closely with the EHT results when the torsional term of eq 6 is added. The torsional term is also included for χ^2, $\chi^{3,1}$, and $\chi^{3,2}$ of leucine (as for the branch in valine), and for χ^1, $\chi^{2,1}$, $\chi^{2,2}$, and $\chi^{3,1}$ of isoleucine.

4. Methionine. Variations of the side-chain dihedral angles χ^1 and χ^2 of methionine are similar to those in leucine and isoleucine (i.e., eq 6, with $U_0 = 2.7$ kcal/mol and $n = 3$, is used for these dihedral angles). Variations of the dihedral angles χ^3 and χ^4 are different from any model ethane-type rotations. Some relevant experimental data are available. In rotation around a C–S bond, dimethyl sulfide has a threefold barrier of ~2.1–2.5 kcal/mol (trans to gauche),[73] and diethyl sulfide has a threefold barrier of ~1.75 kcal/mol.[73] Based on these experimental data, we have included an intrinsic torsional term for χ^3 and χ^4 of the form

$$U(\chi^{3,4}) = (2.0/2)(1 + \cos 3\chi^{3,4}) \quad (7)$$

With the inclusion of eq 7 for χ^3, and a choice of $\phi = \psi = -60°$, $\chi^1 = \chi^2 = \chi^4 = 180°$ in *N*-acetyl-*N'*-methyl-L-methioninamide, minima are obtained at 180 and ±70°, the latter being ~0.4 kcal/mol higher than the former, with an intervening barrier of ~2 kcal/mol. The barrier between the −70 and +70° conformations is ~4.5 kcal/mol.

5. Cysteine and Cystine. The side chain of cysteine differs from that of methionine in that the $-CH_2-S-CH_3$ group of methionine is S–H. Equation 6 (with $U_0 = 2.7$ kcal/mol and $n = 3$) is used for χ^1. The following experimental[73] data are available for rotation of the H(S) atom, with the experimental barrier being threefold in all cases: $U_0 = 1.27$ kcal/mol for methanethiol, 1.64 kcal/mol for ethanethiol, and 1.36 kcal/mol for 2-methyl-2-propanethiol. From these data, we select an intrinsic torsional potential of the form

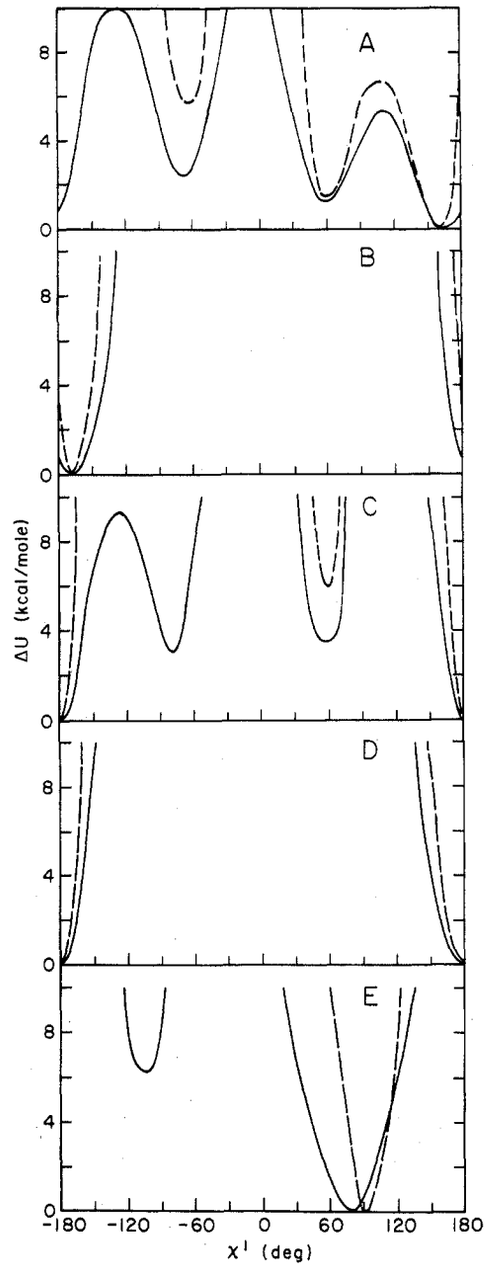

Figure 10. Energy for variation of χ^1 in *N*-acetyl-L-valine-*N'*-methylamide, for various backbone conformations: (A) α_R ($\phi = -50°$; $\psi = -40°$); (B) α_L ($\phi = 50°$; $\psi = 40°$), (C) equatorial seven-membered ring ($\phi = -70°$; $\psi = 70°$), (D) axial seven-membered ring ($\phi = 70°$; $\psi = -70°$), and (E) extended ($\phi = -170°$; $\psi = 170°$). $\chi^{2,1}$ and $\chi^{2,2}$ were held fixed at 60°. The solid and dashed curves are the EHT and total empirical (including torsional term) results, respectively. Each curve was normalized to zero at its own minimum.

$$U(\chi^2) = (1.5/2)(1 + \cos 3\chi^2) \quad (8)$$

The empirical energy of *N*-acetyl-*N'*-methylcysteinamide at $\phi = \psi = -60°$, $\chi^1 = 180°$, without the torsional term of eq 8, is 0.4 kcal/mol at $\chi^2 = 180°$ and 0.0 kcal/mol at ±80°. When the torsional term of eq 8 is included, the total energy has minima at $\chi^2 = 180°$ and ±70° (0.2 kcal/mol higher at ±70°), with a barrier between 180 and −70° of ~1.50 kcal/mol, and between −70 and +70° of ~3.8 kcal/mol. The lower barrier is in good agreement with the experimental values cited above.[73] No experimental data are available for the gauche–gauche barrier or for the difference in energy between the minima.

Cystine has a disulfide bond and is considered to be similar to the model compound dimethyl disulfide. The experimental data[73] for both dimethyl disulfide and hydrogen persulfide indicate a barrier height of U_0 = 7 kcal/mol. Structural studies[84,85] give a value of 84–85° for the CS–SC dihedral angle. The observed[73] barrier is most probably that for a gauche to trans rotation since theoretical calculations[86] indicate that the gauche–cis–gauche pathway has a barrier nearly twice as high as the gauche–trans pathway. In using the potential function described in Appendix III, the disulfide bond is formed only as the calculation proceeds, i.e., an algorithm is used to direct the formation of this bond. This algorithm leads to the correct bond lengths and bond angles for the disulfide bridge, and also includes an intrinsic torsional potential which has minima at ±90° and barriers to rotation around the S–S bond of approximately 7 kcal/mol. This algorithm is described in Appendix III. The torsional barrier for the CC–SS bond is the same as that used for cysteine (eq 8).

6. Serine and Threonine. In order to estimate the rotational barrier for the hydroxyl groups of serine and threonine, the results for model compounds have been used (see footnote *a* of Table VI). The rotation of the hydroxyl group in threonine (*N*-formyl-L-threonine-*N'*-methylamide) is not the same as in serine since any movement of the hydroxyl hydrogen is influenced by the neighboring (γ) methyl group. For $\phi = -50°$, $\psi = -40°$, $\chi^1 = -60°$, $\chi^{2,2} = 180°$, the minima were found to occur at $\chi^{2,1} = -180$ and $-60°$, with an intervening maximum at $-120°$. The minimum at $-60°$ is ~1.4 kcal/mol higher in energy than that at $-180°$. An intrinsic torsional term is included both for $\chi^{2,1}$ (U_0 = 0.6 kcal/mol) and $\chi^{2,2}$ (U_0 = 2.7 kcal/mol).

7. Lysine and Ornithine. The torsional potentials for χ^1, χ^2, χ^3, and χ^4 of lysine and χ^1, χ^2, and χ^3 of ornithine are computed from eq 6 with U_0 = 2.7 kcal/mol and n = 3. The torsional term for χ^5 of lysine and χ^4 of ornithine was taken as that for methylamine (Table VI), when the ϵ (and δ) amino group was either charged or uncharged. For the charged $-NH_3^+$ and uncharged $-NH_2$ amino groups, rotational barriers of ~0.5 and ~0.2 kcal/mol are obtained if the intrinsic torsional term is not included.

8. Arginine. The side-chain torsional terms for χ^1, χ^2, and χ^3 of arginine are taken to be that of eq 6 with U_0 = 2.7 kcal/mol and n = 3. For rotation about the C^δ–N^ϵ bond, we anticipate a very low rotational barrier (less than 1.0 kcal/mol) because barriers for rotation of methyl groups attached to planar conjugated groups (e.g., C'O–NH–CH_3) are generally in the range of 0 to 1.0 kcal/mol.[42] Since the empirical calculations, *without* an intrinsic torsional term, give a barrier of ~0.3 kcal/mol for rotation about the C^δ–N^ϵ bond, no intrinsic torsional term is used for χ^4.

The case of the N^ϵ–C^ζ bond (involving χ^5) is obviously different. The guanidyl group is nearly planar, indicating a resonance structure between various charged forms. Thus, the guanidyl group will have high barriers to rotation about each bond (i.e., $\chi^{6,1}$ and $\chi^{6,2}$ as well as χ^5). We have carried out CNDO/2[12] molecular orbital studies on these three torsional terms for charged arginine, and found barriers to rotation of ~18–20 kcal/mol for χ^5, $\chi^{6,1}$, and $\chi^{6,2}$. Without an intrinsic torsional term, the empirical energy for variation of χ^5 was found to be ~7 kcal/mol at $\chi^5 = 0°$ and ~0 kcal/mol at 90°. With the inclusion of a torsional term of the form

$$U(\chi^5) = (18.0/2)(1 - \cos 2\chi^5) \quad (9)$$

minima are obtained at $\sim\chi^5 = \pm 15°$, with a total barrier to rotation of ~14 kcal/mol at 90°. A similar result is obtained for $\chi^{6,1}$ and $\chi^{6,2}$ when a torsional term of the form

$$U(\chi^{6,1} \text{ or } \chi^{6,2}) = (20.0/2)(1 - \cos 2\chi^{6,1 \text{ or } 6,2}) \quad (10)$$

is used. Equations 9 and 10 are also used for the uncharged guanidyl group. These results are in excellent agreement with X-ray[87] and neutron diffraction data[88] which also show small deviations from planarity in the guanidyl group. There appear to be no experimental data for barrier heights represented by these torsional terms against which to check our values.

9. Aspartic Acid and Glutamic Acid. The aspartic and glutamic acid side chains differ only in the number of $-CH_2$ groups (1 in Asp and 2 in Glu). As in the previous sections, χ^1 of aspartic acid and χ^1 and χ^2 of glutamic acid are taken to have a torsional term of eq 6 (with U_0 = 2.7 kcal/mol and n = 3). The torsional energies for χ^2 of aspartic acid and χ^3 of glutamic acid are likely to be small (i.e., less than 1.0 kcal/mol) since experimental data on model compounds such as acetic acid[73] give torsional barriers of ~0.5 kcal/mol. No intrinsic torsional term is used here for χ^2 of aspartic and χ^3 of glutamic acid. For variation of $\chi^{3,2}$ (CC–OH) in Asp, the experimental evidence[73] indicates that the trans form (i.e., $\chi^{3,2} = 180°$) is favored, with cis–trans barrier to rotation of between 10–17 kcal/mol.[73] The energy difference between the cis and trans minima is found experimentally[73] to be ~4 kcal/mol. The empirical calculations, without any intrinsic torsional component, give a minimum at $\chi^{3,2} = \pm 180°$ and a maximum of ~5 kcal/mol at $\chi^{3,2} = 0°$. In order to match the observed data described above, we have added an intrinsic torsional term of the form

$$U(\chi^{3,2}) = (8.0/2)(1 - \cos 2\chi^{3,2}) \quad (11)$$

which results in a total empirical energy upon rotation about the CC–OH bond with the lowest-energy minimum at $\chi^{3,2} = \pm 180°$, a maximum of ~10 kcal/mol at ±90°, and a second minimum at $\chi^{3,2} = 0°$, which is 5 kcal/mol above the lowest energy minimum.

The torsional term developed here for the $\chi^{3,2}$ term of aspartic acid is used for the $\chi^{4,2}$ term of glutamic acid. The same torsional term is used for the α-carboxyl end group of a polypeptide chain. In the case of the carboxylate (or charged state), no torsional term is used for χ^2 of aspartic acid and χ^3 of glutamic acid.

10. Asparagine and Glutamine. As with aspartic and glutamic acids, the torsional term for the C^α–C^β bond is taken to be that of eq 6 (with U_0 = 2.7 kcal/mol and n = 3). However, asparagine and glutamine have amide side-chain groups which involve rotation about the $C^\alpha C^\beta$–$C^\gamma N$ bond of asparagine (i.e., variation of χ^2 of asparagine, and χ^3 of glutamine). Similarly to the variation of ψ of the backbone, no intrinsic torsional term is included for variation of χ^2 in asparagine and χ^3 in glutamine. The amide group is taken to be planar, and requires an intrinsic torsional term of the type

$$U(\chi^2) = (15/2)(1 - \cos 2\chi^2) \quad (12)$$

to provide agreement with experimental[73] and theoretical[16] results. The same torsional potential is used for $U(\chi^3)$ of Gln.

11. Histidine, Phenylalanine, Tyrosine, and Tryptophan. The ring-bearing amino acids are similar in that

each has C^{α}–C^{β} and C^{β}–C^{γ} bonds about which rotation can occur, and each amino acid has a planar ring (involving the C^{γ} atom) whose position depends upon rotation about these bonds. Equation 6 (with U_0 = 2.7 kcal/mol and n = 3) is used to describe the torsion about the C^{α}–C^{β} bond for all ring-bearing side chains.

Because the imidazole ring can hydrogen bond to the backbone through the N^{δ} or N^{ϵ} group, histidine exhibits side-chain torsional energies which depend strongly on the backbone conformation. However, because of the similarity of the rings at the C^{γ} atom, we expect the *intrinsic* torsional potential about the C^{β}–C^{γ} bond to be similar to that in phenylalanine, in the sense that planar rings involve low barriers, so that only the phenylalanine side chain is examined in detail here. The results for phenylalanine are assumed to be applicable to histidine, and also to tyrosine and tryptophan. The variation of χ^2 of phenylalanine leads to a very small barrier to rotation (~0.5 kcal/mol), in agreement with available experimental data;[73] thus, no torsional term is used for rotation about the C^{β}–C^{γ} bond of any of the ring-bearing side chains.

In tyrosine, the hydroxyl group on the phenyl ring can also be rotated. In this case, resonance appears to constrain the proton of the hydroxyl group to the plane of the ring (i.e., $\chi^6 \approx 0$ or 180°). An intrinsic torsional term of the following form is used for this rotational energy:

$$U(\chi^6) = (3.5/2)(1 - \cos 2\chi^6) \qquad (13)$$

The remaining contributions to the empirical energy are small. Thus, the resulting total empirical energy (including eq 13) is in good agreement with available experimental data[73] for the hydroxyl group attached to aromatic rings; i.e., the experimental torsional barrier to rotation of the hydroxyl group in phenol is found[73] to be 3.1 and 3.5 kcal/mol.

12. Proline and Hydroxyproline. Proline has no side-chain bonds which rotation can take place for any one given form of the pyrrolidine ring (i.e., with the C^{γ} atom up, down, or planar; see section P of Appendix I). However, the conversion from one form to another involves changes in various dihedral angles, causing the internal energy (i.e., that energy associated with the interactions within the ring and between the atoms in the backbone and those of the ring) to change. The torsional potential for the backbone peptide bond of proline and hydroxyproline is taken to be the same as that described in section VIA for all other peptide bonds. The internal energies[89] for up, down, and planar proline, for cis and trans peptide bonds, are given in section P of Appendix I.

The rotation of the hydroxyl group in hydroxyproline was investigated empirically, and an intrinsic torsional potential (the same as that for the hydroxyl torsion in serine and threonine) was required (see Table VI).

VII. Discussion

The total empirical energy includes all of the terms described in section V and VI. The conformational energy of a polypeptide in a given conformation may be computed by the following procedure. First, a conformation (including end groups) is generated[60] using the geometry described here and any desired set of dihedral angles. The position of every atom is thus obtained in a cartesian coordinate system. Each atom is then assigned a type[26b] (i.e., H_1, H_2, ..., C_6, ..., N_{13}, etc.) and an array number (i.e., 1 through the total number of atoms), starting with the N-terminal end group and ending with the last atom of the C-terminal end group. Second, a decision (predefined for each amino acid residue and end group) is made as to which type of interactions are involved for each atom pair, i.e., 1–4, hydrogen bonding, or nonbonding with F = 1.0. The partial charges, nonbonded coefficients, and torsional terms are assigned to the appropriate atoms or bonds of each residue. Third, the total energy (for the given conformation) is calculated for the molecule. If desired (as, for example, in energy minimization) a new conformation is generated, and the energy is recalculated. Programs which use the parameters reported here are available.[60] For an examination of several studies using these techniques and parameters, ref 17, 19, 21, and 89 may be consulted. The ability of these computational methods to predict physically observable data may be evaluated by comparing the conformations calculated for the *N*-acetyl-*N'*-methylamides of the 20 naturally occurring amino acids[17] with the numerous experimental data cited in ref 17.

Further support for the reduction of the repulsive coefficient for 1–4 interactions comes from a similar conclusion reached by Nelson and Hermans.[90] They found, as we did, that parameters obtained from intermolecular interactions in molecular crystals cannot be taken over directly to polypeptide calculations without some consideration of the vibrational contributions to the potentials. Our approach to the calculations of intramolecular terms differs from that of Nelson and Hermans[90] in the treatment of 1–4 interactions.

It will be many years before rigorous molecular orbital methods (that are reliable for conformational studies) will be able to treat large polypeptides. Thus, we have developed empirical potential energies which fit both experimental conformational data and barrier heights. It appears at present that the computational ease and speed in applying the empirical techniques developed here to the study of *macromolecules* outweigh the advantages of the various molecular orbital methods. These empirical methods should be of value for many different aspects of the study of molecular conformation.

Acknowledgment. We are indebted to Dr. S. Tanaka and Dr. D. H. Wertz for helpful discussions on the geometry, partial charges, and energies of proline residues, and to Marcia Pottle, Shirley Rumsey, and Dr. G. F. Endres for checking the parameters and the computer program.

Appendix I. Geometry of Individual Amino Acid Residues

The bond lengths and bond angles for the backbone atoms of each amino acid residue (–NH–CHR–CO) are given in Table I. In this Appendix, we give the remaining geometry for all side-chain atoms, as well as proline and hydroxyproline. The positioning of hydrogen atoms, whose coordinates are usually not well defined by X-ray diffraction methods, was described earlier (see Table III of ref 10a). The same procedure has been used here; the positions deduced for the hydrogen atoms are in good agreement with recent neutron diffraction and gas-phase electron diffraction data, some of which became available (and are referenced here) subsequent to our computations.

The following geometries apply to *neutral* (i.e., uncharged) residues, and (except where indicated otherwise) also to charged residues. Only neutral geometries are shown in Figure 11–15.

A. Glycine

The geometry of the glycine residue, deduced from several recent,[91-94] as well as older,[4a,31] studies, is shown in Fig. 11A. The main differences from earlier data[4a] are the N-C^{α} bond length of 1.453 Å and the C'=O bond length of 1.230 Å, both slightly shorter than reported previously.[4a] The $NC^{\alpha}C'$ bond angle of 111°, which is an average of those reported in refs. 91-94, is greater than that reported previously,[4a] while the other angles are very similar.

B. Alanine

The geometry of the alanine residue[95-97] is shown in Fig. 11B. The $NC^{\alpha}C'$ bond angle is smaller than that in glycine, and the other atoms around the C^{α} atom depart from a tetrahedral arrangement. The $C'C^{\alpha}C^{\beta}$ bond angle is found to be greater than the $NC^{\alpha}C'$ bond angle, a result which is observed for many of the naturally occurring amino acids.

C. Valine

The geometry chosen for the valine residue[98-100] is shown in Fig. 11C. The C^{γ} atoms of valine appear to tip (by several degrees) around the C^{β} atom in a recent crystal structure,[98] probably because of crystal packing interactions. We have chosen to make the $C^{\alpha}C^{\beta}C^{\gamma 1}$ and $C^{\alpha}C^{\beta}C^{\gamma 2}$ bond angles equivalent and nontetrahedral, as can be seen in Fig. 11C, since we believe the observed difference in these angles arises because the backbone is in the β structure in the crystal. By making these angles equivalent, we hope to avoid biasing the backbone of valine toward the β structure[116-119] by 0.06 Å. Neutron diffraction studies[118,119] yield a value of 1.48 Å for this bond length in agreement with one x-ray value.[117] An earlier x-ray study,[116] using the anomalous dispersion method, gave this bond length as 1.54 Å. We have used 1.48 Å for both the charged and uncharged residues since no geometry is available for the uncharged side chain. Similar geometry for ornithine (with a C^{δ}-N^{ϵ} distance of 1.486 Å) has also been reported.[120] In the charged Lys^{+} residue, the only difference is the hydrogen geometry on N^{ϵ} [i.e., τ(CNH) = 109.47° and τ(HNH) = 109.47°].

K. Arginine

The geometry of the uncharged arginine residue[87] is shown in Fig. 12E. The two C^{ζ}-N^{η} bonds were observed to have different bond lengths[87] in the uncharged state. In the charged state, the two C^{ζ}-N and N^{ϵ}-C^{ζ} bonds are all taken as equivalent in length (1.34 Å), in agreement with recent neutron diffraction results.[88] The geometry of the NH_2 groups in both the charged and uncharged state is selected so that the hydrogens lie in the plane defined by the N^{ϵ}, C^{ζ} and N^{η} atoms with HNH angles of 120°, in good agreement with the hydrogen positions found experimentally.[88] One hydrogen is on $N^{\eta 1}$ in the uncharged state and two in the charged state.

L. Aspartic and Glutamic Acids

The geometry of the aspartic acid residue[121,122] is shown in Fig. 13A. The uncharged side-chain carboxyl group has two unequal C-O bond lengths. The charged carboxylate ion group is made to have equivalent oxygens with a $C^{\beta}C^{\gamma}O^{\delta 1,2}$ bond angle of 117° and C=O bond lengths of 1.25 Å, in agreement with previous work,[4a]

The geometry of the uncharged glutamic acid residue is shown in Fig. 13B. The geometry of the charged state of the carboxylate ion group is the same for both aspartic and glutamic acids.

M. Asparagine and Glutamine

The geometries of the asparagine[123-125] and glutamine[126] residues are shown in Figs. 13C and 13D, respectively. The side-chain amide groups have typical backbone amide geometries[30-33] with the HNH bond angle taken as 120°, in good agreement with the neutron diffraction results.[126]

N. Histidine

The geometry of one form of the neutral histidine residue[127-128] is shown in Fig. 14A. The second form of the residue is obtained by removing the hydrogen from the N^{δ} atom and placing it on the N^{ϵ} atom. Since it was not possible to locate these hydrogens from the x-ray data, we use the same ring geometry for both neutral forms. The charged histidine residue is obtained when hydrogens are placed on both ring nitrogens, and no difference in ring geometry is introduced in going from the neutral to the charged species. The differences in partial atomic charge for the different charged species were given in Table V.

O. Phenylalanine, Tyrosine, and Tryptophan

The geometry of the phenylalanine residue[115,129-132] is shown in Fig. 14B (all bond lengths of the ring are not given because of symmetry). The structural data for this residue are unusual in that the $C'C^{\alpha}C^{\beta}$ angles range from 104° to 117°, while the $C^{\alpha}C^{\beta}C^{\gamma}$ angle ranges from 103° to 120° for different structural determinations. These and other very large ranges of geometry make it difficult to choose a standard geometry for this residue. The values for these bond angles for phenylalanine (shown in Fig. 14B) are in close agreement with the most recent structural data[130,132] and with those for the observed bond angles of tyrosine[130,133,134] whose geometry is shown in Fig. 14C. The geometry of tryptophan[135,136] is shown in Fig. 14D.

P. Proline and Hydroxyproline

The geometry selected for proline (with the pyrrolidine ring in the up, down, and planar conformations, where these terms refer to the relative positions of the C^{γ} and C^{β} atoms; see Fig. 15A, B, C) are given in Table VII. Because of this possible puckering of the pyrrolidine ring, three different geometries are described in Table VII, and the corresponding dihedral angles are given in Table VIII. The data of Table VII conform as closely as possible to reliable x-ray and neutron diffraction data.[137-148] The dihedral angles of Table VIII were obtained by varying φ, χ^1, χ^2, χ^3, and χ^4 (keeping bond lengths and bond angles fixed) to obtain exact ring closure. It should be noted that, in the two resulting puckered conformations of the pyrrolidine ring, the position of the C^{γ} atom is determined by φ and χ^1 (positive values of χ^1 pertain to a clockwise rotation of the C^{γ} atom when looking from C^{α} to C^{β}). The two puckered conformations are designated up (U) and down (D), with φ = -67.6°, χ^1 = -6.11° (for U) and φ = -75.0°, χ^1 = +18.67° (for D); these values of φ and χ^1 pertain to both the cis and trans conformations about the X-Pro bond. It can be seen in Fig. 15A and B that, in both the U and D forms, both the C^{β} and C^{γ} atoms lie out of the plane defined by the C^{α}, N and C^{δ} atoms.

Besides these differences in the puckering of the pyrrolidine ring, there is also a different bond length for the X-Pro peptide bond,[141,145] viz., 1.36 Å instead of 1.325 Å. Further, in trans X-Pro (ω_{i-1} = 180°), the bond angles[141,145] are $\tau(C'_{i-1}N_iC^{\alpha}_i)$ = 121° and $\tau(C'_{i-1}N_iC^{\delta}_i)$ = 126° while, in cis X-Pro (ω_{i-1} = 0°), they are reversed,[19,149] viz., $\tau(C'_{i-1}N_iC^{\alpha}_i)$ = 126° and $\tau(C'_{i-1}N_iC^{\delta}_i)$ = 121°. These data agree with those from other N-disubstituted structures.[150-152] X-HyPro is treated in the same way as X-Pro.

As a result of both the different types of puckering and cis-trans isomerism about the X-Pro peptide bond, the internal energy of the "fixed" part of the proline residue differs; i.e., the internal energy is different for down-trans, down-cis, up-trans, up-cis, planar-trans, and planar-cis. These internal conformational energies (which include the C'_{i-1} and C'_i atoms; see Fig. 1 of ref. 89) are given in Table IX.[89]

The geometry of hydroxyproline[137,138,144] is shown in Fig. 15D, E, F. The hydroxyl group is positioned as shown in Fig. 15D and E, in agreement with the experimental results.[137,138,144]

The planar structures of proline and hydroxyproline (φ = -57.6°), shown in Figs. 15C and 15F, represent a simplification of the two puckered forms. However, there is no experimental evidence for its existence, and therefore caution must be exercised in its use in conformational energy calculations.

G. Cysteine and Cystine

The geometry of the cysteine residue[109,110] is shown in Fig. 12A. The $NC^{\alpha}C'$ bond angle is observed to be smaller than tetrahedral, and the C-S bond length is 1.83 Å as compared to 1.86 Å reported earlier.[4a]

The geometry of cystine[111-113] is not shown. The most important difference between cysteine and cystine is the increase in the $C^{\alpha}C^{\beta}S$ bond angle from 110° in cysteine to 114.8° in cystine; otherwise the geometries of these two amino acids are the same. Upon formation of the disulfide, a CSS bond angle of 104.0°, and an S-S bond length of 2.04 Å are used, and an intrinsic barrier of ~7 kcal/mole is introduced for variation of the CS-SC dihedral angle (see Appendix III).

H. Serine

The geometry of the serine residue[114] is shown in Fig. 12B. The $C^{\alpha}C^{\beta}O$ bond angle is 112°, compared to 110° reported previously.[4a]

I. Threonine

The geometry of the threonine residue[115] is shown in Fig. 12C. The $C^{\alpha}C^{\beta}O$ bond angle (104°) is smaller than that of serine (112°); this may be a result of the steric influence of the C^{γ} methyl group, or of different electrostatic or hydrogen-bonding interactions.

J. Lysine and Ornithine

The geometry of the uncharged lysine residue is shown in Fig. 12D. The C^{ϵ}-N^{ζ} bond length differs from structure to

Appendix II

Treatment of End Groups

The termination of the polypeptide chain may be carried out in several ways, by using different N- and C-terminal end groups. The bond lengths and bond angles of the end groups are given in Tables X and XI, and their partial charges are given in Tables XII and XIII.

Any N-terminal end group may be used with any C-terminal end group. If it is desired to maintain electroneutrality, then end groups of equal and opposite charge must be used. However, the use of the program does not demand the imposition of electroneutrality. Since no provision is made in the computer program for contributions from solvent or counterions, there is no compensation of charge on a molecule whose net charge is not zero.

In general, the end groups are small fragments such as $-NH_2$ and -COOH. However, the whole N-terminal proline residue is treated as an end group when the nitrogen of the pyrrolidine ring is sp^3 hybridized, because the geometry of such a pyrrolidine ring differs from that of an internal proline residue whose nitrogen is sp^2 hybridized. Two forms are required for a neutral N-terminal proline group, cis-H-proline and trans-H-proline, depending on whether the H on the ring nitrogen is cis or trans, respectively, to the C' atom of the same residue. The backbone of pyroglutamic acid also is involved in a ring, and requires special treatment (see Tables X and XII).

All of the end groups have zero net charge except $-NH_2$, $-NH_3^+$, cis-H-proline, trans-H-proline, ^+H_2N-proline (the charged form of N-terminal proline), -COOH, and COO^-. As with the charges on each residue, the charges on the end groups were determined from CNDO/2 calculations of appropriate model compounds, e.g., NH_2CH_2COOH. Since the charges on the backbone -NH and CO- of each residue are included in the residue charges, which net to zero, some slight alteration of these charges is necessary, when end groups are added, to match the CNDO/2 charges, within ~0.003 e.c.u. For example, when an H atom is added to the N-terminus to complete the $-NH_2$ end group, the extra H-atom is given a charge of +0.176, the same as that of the H of the -NH group (see Table II). The N-terminal residue (including the end group) thus has a net charge of +0.176. This net charge must be balanced by a C-terminal group with a net charge of -0.176 if it is desired to keep the molecule electrically neutral. For an $-NH_3^+$ end group, an additional charge of +0.109 e.c.u. is added to the +0.176 of the H of the -NH group, and the two new H's are each assigned a charge of +0.285 e.c.u. Thus, all three hydrogens have a charge of +0.285 each, and the net charge of 0.109 + 0.285 + 0.285, or +0.680, matches the CNDO/2 result for the end group. In the case of cis-H-proline and trans-H-proline, the net charge of +0.176 e.c.u. has been distributed over the atoms of the pyrrolidine ring, the net charge of +0.680 for the ^+H_2N-proline has been treated similarly. This re-distribution was done in such a way as to preserve the relative charges obtained from the CNDO/2 calculations. In a similar way, charges of -0.380 and +0.204 on the O and H of the carboxyl -OH, give a net charge of -0.176 to the C-terminal -COOH group (the same magnitude as the N-terminal $-NH_2$ group). Likewise, augmentation of the charge on the carbonyl O by -0.148 e.c.u. from -0.384 (Table II) to -0.532, and the assignment of -0.532 to the end group O, yields two equivalently charged oxygens in the $-COO^-$ end group, with a net charge of (-0.148) + (-0.532), or -0.680 e.c.u., which is the CNDO/2 net charge on the C-terminal residue (including the end group).

Table VII

Geometry of Trans[a] L-Proline and L-Hydroxyproline

Bond	Proline Bond Length, Å Up and Down[b]	Proline Bond Length, Å Planar[b]	Hydroxyproline Bond Length, Å Up and Down[c]	Hydroxyproline Bond Length, Å Planar[d]
C'=O[e]	1.230			
C'-N (X-Pro)	1.360			
N-C^{α}	1.453			
C^{α}-C^{β}	1.520	1.500		1.500
C^{β}-C^{γ}	1.520	1.470		1.470
C^{γ}-C^{δ}	1.520	1.452		1.452
C^{δ}-N	1.460			
C^{α}-C'	1.530			
C'=O (Pro-X)	1.230			
C'-N (Pro-X)	1.325			
C^{α}-H	1.000			
C^{β}-H	1.090			
C^{γ}-H	1.090		1.000	
C^{δ}-H	1.090			
C^{γ}-O	-		1.490	1.460
O-H	-		1.000	

Bond Angle	Proline Bond Angle, Deg Down	Proline Bond Angle, Deg Up[b]	Proline Bond Angle, Deg Planar[b]	Hydroxyproline Bond Angle, Deg Down[c]	Hydroxyproline Bond Angle, Deg Up[d]	Hydroxyproline Bond Angle, Deg Planar[d]
$C^{\alpha}NC'$ (X-Pro)	115.0					
$C^{\alpha}C'O$[e]	120.5					
OC'N (X-Pro)	124.5					
$C'NC^{\alpha}$ (X-Pro)	121.0[f]					
$C'NC^{\delta}$ (X-Pro)	126.0[f]					

Table VII (continued)

Bond Angle	Proline Bond Angle, Deg. Down	Proline Bond Angle, Deg. Up[b]	Proline Bond Angle, Deg. Planar[b]	Hydroxyproline Bond Angle, Deg. Down[c]	Hydroxyproline Bond Angle, Deg. Up[d]	Hydroxyproline Bond Angle, Deg. Planar[d]
$C^{\alpha}NC^{\delta}$	113.0					
$NC^{\alpha}H$	110.3		110.1			110.1
$NC^{\alpha}C^{\beta}$	105.0					
$NC^{\alpha}C'$	111.0					
$HC^{\alpha}C^{\beta}$	111.8		111.3			111.3
$HC^{\alpha}C'$	105.9		106.5			106.5
$C^{\beta}C^{\alpha}C'$	113.0					
$C^{\alpha}C^{\beta}H$	111.0	110.8	110.8		110.8	110.8
$C^{\alpha}C^{\beta}C^{\gamma}$	105.9	106.6	106.6		106.6	106.6
$HC^{\beta}H$	107.0					
$HC^{\beta}C^{\gamma}$	111.0	110.8	110.8		110.8	110.8
$C^{\beta}C^{\gamma}H$	110.5	110.9	109.6	112.1	113.0	109.6
$C^{\beta}C^{\gamma}C^{\delta}$	107.9	106.3	111.4		106.3	111.4
$HC^{\gamma}C^{\delta}$	110.5	110.9	109.6	112.1	113.0	111.8
$HC^{\gamma}H$	107.0			-	-	-
$NC^{\delta}H$	111.3	111.7	111.4		111.7	111.4
$NC^{\delta}C^{\gamma}$	104.7	103.2	104.0		103.2	104.0
$HC^{\delta}H$	107.0		107.2			107.2
$HC^{\delta}C^{\gamma}$	111.3	111.7	111.4		111.7	111.4
$C^{\alpha}C'N$ (Pro-X)	115.0					
$C^{\alpha}C'O$	120.5					
OC'N (Pro-X)	124.5					
$C^{\beta}C^{\gamma}O$	-			106.0		109.0
$OC^{\gamma}H$	-			112.4	111.9	109.0
$C^{\delta}C^{\gamma}O$	-			106.0		
$C^{\gamma}OH$	-			110.0		

Footnotes to Table VII

(a) Changes in bond angles for cis L-proline and cis L-hydroxyproline are described in this section.

(b) The values for the up and planar pyrrolidine rings are the same as those for down, except where indicated otherwise.

(c) Same as Down Proline except where indicated otherwise.

(d) Same as Down Hydroxyproline except where indicated otherwise.

(e) For the residue preceding proline.

(f) These bond angles are reversed[a] for cis L-proline and cis L-hydroxyproline.

Table VIII

Dihedral Angles of the Puckered Pyrrolidine Ring of Proline

Dihedral Angle[26]	Angle (Deg.) Up	Angle (Deg.) Down	Angle (Deg.) Planar
$\phi(C'N\text{-}C^{\alpha}C')$	-67.6	-75.0	-57.6
$\chi^1(NC^{\alpha}\text{-}C^{\beta}C^{\gamma})$	-6.11	18.67	0.0
$\chi^2(C^{\alpha}C^{\beta}\text{-}C^{\gamma}C^{\delta})$	19.07	-14.02	0.0
$\chi^3(C^{\beta}C^{\gamma}\text{-}C^{\delta}N)$	-24.20	3.83	0.0
$\chi^4(C^{\gamma}C^{\delta}\text{-}NC^{\alpha})$	21.64	8.65	0.0

Table IX

Internal Conformational Energy of the Fixed Part of a Proline Residue[89]

Puckering Conformation[a]	Peptide Bond Conformation	Internal Conformational Energy[b] kcal/mole
D	Cis	0.000[c]
U	Cis	0.202
P	Cis	0.920
D	Trans	0.336
U	Trans	0.550
P	Trans	1.366

(a) The designations, down (D), up (U), and planar (P) pertain to the (ϕ = -75.0°, χ^1 = +18.67°), (ϕ = -67.6°, χ^1 = -6.11°), and (ϕ = -57.6°, χ^1 = 0.0°) positions, respectively, of the C^{γ} atom.

(b) The energy pertains to the unit shown in Fig. 1 of ref. 89, and comprises the nonbonded and electrostatic contributions.

(c) All values are expressed as relative values with respect to the down-cis conformation by subtracting 12.355 kcal/mole from each of the computed energies.

Table X

Geometry for N-Terminal End Groups[a]

End Group	Bond Length[b] (Å)		Bond Angle (Deg.)	
$-NH_2$[c]	$N_{14}\text{-}H_2^{1,2}$	1.014	HNH	106.8
			HNC^{α}	112.0
$-NH_3^+$[c]	$N_{13}\text{-}H_2^{1,2,3}$	1.014	HNH	109.47
			HNC^{α}	109.47
$-NHCH_3$[c]	$C_6\text{-}H_1^{1,2,3}$	1.090	HCH	109.47
	$C_6\text{-}N_{14}$	1.490	HCN	109.47
			CNH(N)	124.0
			CNC^{α}	121.0
$-NHCOCH_3$	$C_7\text{=}O_{17}$	1.230	CCO	120.5
	$C_6\text{-}C_7$	1.490	HCC	109.47
	$C_6\text{-}H_1^{1,2,3}$	1.090	HCH	109.47
	$C_7\text{-}N_{13}$	1.325	CCN	115.0
			OCN	124.5
			CNH(N)	124.0
			CNC^{α}	121.0
-NHCHO	$C_7\text{=}O_{17}$	1.240	HCO	132.5
	$C_7\text{-}H_3$	1.090	HCN	103.9
	$C_7\text{-}N_{13}$	1.340	OCN	123.6
			CNH(N)	124.0
			CNC^{α}	121.0

(a) Data for N-terminal cis-H-proline, trans-H-proline, ^+H_2N-proline, and pyroglutamic acid are also available in the computer program.[60] The geometry of the proline-containing end groups is a slight modification of that found by T. F. Koetzle (private communication) for L-proline.

(b) The subscripts refer to atom types.[26b]

(c) These end groups are not used with proline-type residues; instead N-terminal proline is treated as a special end group when its nitrogen is tetrahedral.[60]

Table XI

Geometry of C-Terminal End Groups

End Group	Bond Length[a] (Å)		Bond Angle (Deg)	
-COOH	$C_7\text{=}O_{17}$	1.230	OCO	124.5
	$C_7\text{-}O_{18}$	1.290	$C^{\alpha}C'O_{17}$	120.5
	$O_{18}\text{-}H_4$	1.000	$C^{\alpha}C'O_{18}$	115.0
			COH	110.0
$-COO^-$	$C_7\text{=}O_{17}^{1,2}$	1.250	OCO	126.0
			$C^{\alpha}C'O^{1,2}$	117.0
$-COCH_3$	$C_7\text{=}O_{17}$	1.230	OC'C	124.5
	$C_7\text{-}C_6$(Me)	1.330	$C^{\alpha}C'O$	120.5
	$C_6\text{-}H_1^{1,2,3}$	1.090	$C^{\alpha}C'C$	115.0
			C'CH	109.47
			HCH	109.47
$-CONH_2$	$C_7\text{=}O_{17}$	1.230	OC'N	124.5
	$C_7\text{-}N_{13}$	1.325	$C^{\alpha}C'O$	120.5
	$N_{13}\text{-}H_2^{1,2}$	1.000	$C^{\alpha}C'N$	115.0
			C'NH	120.0
			HNH	120.0
$-CONHCH_3$	$C_7\text{=}O_{17}$	1.230	OC'N	124.5
	$C_7\text{-}N_{13}$	1.325	$C^{\alpha}C'O$	120.5
	$N_{13}\text{-}H_2$	1.000	$C^{\alpha}C'N$	115.0
	$N_{13}\text{-}C_6$(Me)	1.453	C'NH	124.0
	$C_6\text{-}H_1^{1,2,3}$	1.090	C'NC	121.0
			HNC	115.0
			NCH	109.47
			HCH	109.47

Table XI (continued)

End Group	Bond Length (Å)		Bond Angle (Deg)	
$-CON(CH_3)_2$	$C_7\text{=}O_{17}$	1.230	OC'N	124.5
	$C_7\text{-}N_{13}$	1.360[b]	$C^{\alpha}C'O$	120.5
	$N_{13}\text{-}C_6^{1,2}$(Me)	1.473	$C^{\alpha}C'N$	115.0
			$C'NC^1$	126.0[c]
			$C'NC^2$	121.0[c]
	$C_6\text{-}H_1$	1.090	C^1NC^2	113.0
			NCH	109.47
			HCH	109.47
$-COOCH_3$	$C_7\text{=}O_{17}$	1.240	OC'O	123.0
	$C_7\text{-}O_{19}$	1.360	$C^{\alpha}C'O_{17}$	123.0
	$O_{19}\text{-}C_6$(Me)	1.450	$C^{\alpha}C'O_{19}$	114.0
	$C_6\text{-}H_1$	1.090	C'OC	112.0
			OCH	109.47
			HCH	109.47
$-COOC_2H_5$	$C_7\text{=}O_{17}$	1.240	OC'O	123.0
	$C_7\text{-}O_{19}$	1.360	$C^{\alpha}C'O_{17}$	123.0
	$O_{19}\text{-}C_6$(Et)	1.450	$C^{\alpha}C'O_{19}$	114.0
	$C_6\text{-}H_1$	1.090	C'OC	112.0
	$C_6\text{-}C_6$	1.530	OCH	109.69
			OCC	111.0
			HCH(sec)	107.0
			HCC(Me)	109.69
			$CC_{Me}H$	109.47
			HCH(Me)	109.47

(a) The subscripts refer to atom types.[26b]

(b) See refs. 150-152.

(c) C^1 is always trans to the carbonyl oxygen, and C^2 is always cis to the carbonyl oxygen.

Table XII

Partial Charges of N-Terminal End Groups[a]

End Group	Charge on Atoms[b] (e.c.u.)		Net Charge on Group (e.c.u.)
$-NH_2$[c]	$H_2^{1,2}$	(+0.176)	+0.176
$-NH_3^+$[c]	$H_2^{1,2,3}$	(+0.285)	+0.680
$-NHCH_3$[c]	$H_1^{1,2,3}$(Me)	(+0.020)	0.0
	C_6	(-0.060)	
$-NHCOCH_3$	$H_1^{1,2,3}$(Me)	(+0.020)	0.0
	C_6	(-0.129)	
	C_7	(+0.455)	
	O_{17}	(-0.386)	
-NHCHO	H_3	(-0.050)	0.0
	O_{17}	(-0.380)	
	C_7	(+0.430)	

(a) Charges for N-terminal cis-H-proline, trans-H-proline, ^+H_2N-proline, and pyroglutamic acid are also available in the computer program.[60]

(b) The subscripts refer to atom types.[26b]

(c) These end groups are not used with proline-type residues; instead N-terminal proline is treated as a special end group when its nitrogen is tetrahedral.[60]

Table XIII

Partial Charges of C-Terminal End Groups

End Group	Charge on Atoms[a] (e.c.u.)		Net Charge on Group (e.c.u.)
-COOH	O_{18}	(-0.380)	-0.176
	H_4	(+0.204)	
$-COO^-$	$O_{17}^{1,2}$	(-0.532)	-0.680
$-COCH_3$	C_6	(-0.180)	0.0
	$H_1^{1,2,3}$	(+0.060)	
$-CONH_2$	N_{13}	(-0.390)	0.0
	$H_2^{1,2}$	(+0.195)	
$-CONHCH_3$	N_{13}	(-0.345)	0.0
	H_2	(+0.163)	
	C_6	(+0.050)	
	$H_1^{1,2,3}$	(+0.044)	
$-CON(CH_3)_2$	N_{13}	(-0.280)	0.0
	$C_6^{1,2}$	(+0.050)	
	H_1^{1-6}	(+0.030)	
$-COOCH_3$	O_{19}	(-0.300)	0.0
	C_6	(+0.135)	
	$H_1^{1,2,3}$	(+0.055)	
$-COOC_2H_5$	O_{19}	(-0.300)	0.0
	C_6(Sec)	(+0.155)	
	C_6(Me)	(-0.100)	
	$H_1^{1,2}$	(+0.040)	
	$H_1^{1,2,3}$	(+0.055)	

(a) The subscripts refer to atom types.[26b]

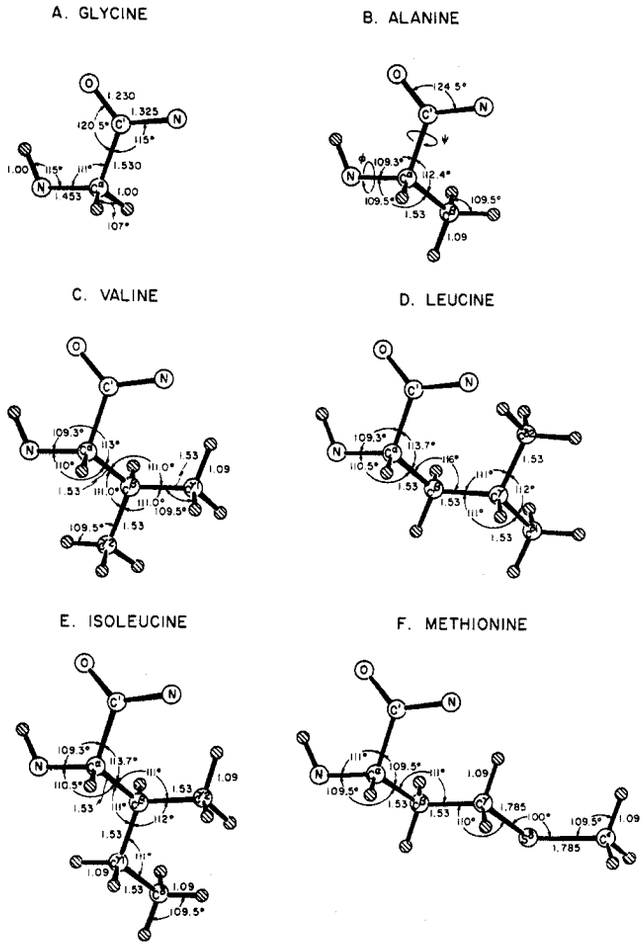

Figure 11. Geometry of several L amino acid residues.

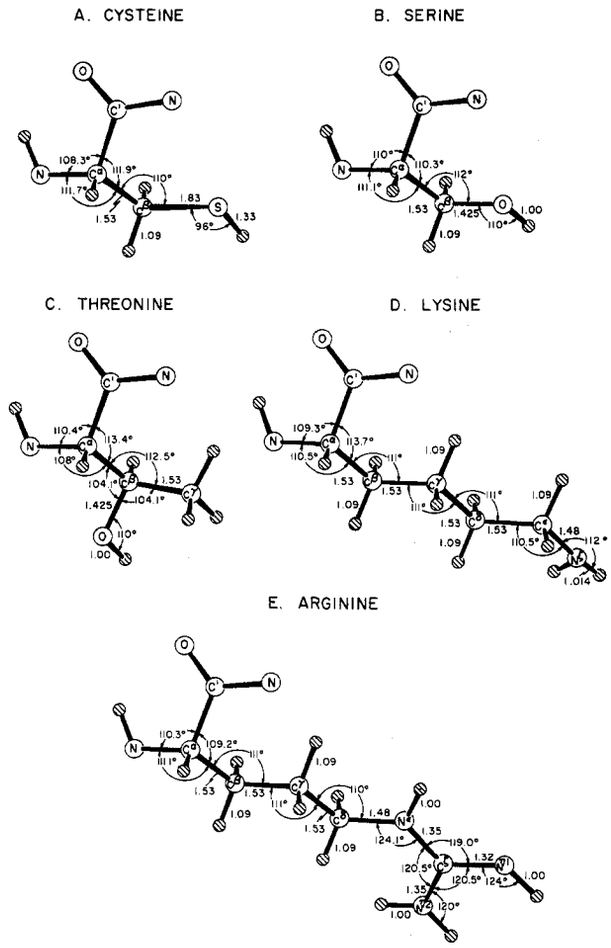

Figure 12. Geometry of several L amino acid residues.

Appendix III. Algorithm for Directing the Formation of Disulfide Bonds

When disulfide bonds occur in a protein, it is necessary to direct the approach of the two half-cystine residues (referred to as i and j) so that the S–S bond length, the bond angles $\tau(C_i^\beta S_i S_j)$ and $\tau(S_i S_j C_j^\beta)$, and the dihedral angle $C_i^\beta S_i S_j C_j^\beta (\chi^{S_i S_j})$ are close to the corresponding experimental values. Since bond lengths and bond angles are generally fixed in these calculations,[60] a functional form for the loop-closing energy has been chosen so that its value goes to zero when the disulfide bond has been formed properly, i.e., when the calculated bond lengths and bond angles are the same as the values chosen from experiment [$S_i S_j$ bond length[84,153] = 2.04 Å, $\tau(C_i^\beta S_i S_j)$ and $\tau(S_i S_j C_j^\beta)$ bond angles[84,153] = 104°]. Another term is included in this algorithm to simulate the twofold potential for rotation about the SS bond (variation of $\chi^{S_i S_j}$). Experimentally, this barrier has been reported[73] to be approximately 7 kcal/mol with the maxima occurring at $\chi^{S_i S_j} = 0$ and 180°, and minima at ±90°.

The disulfide-bond-forming algorithm is defined analytically by

$$U(\chi^{SS}) = A(r_4 - r_{4,0})^2 \tag{14}$$

and

$$U_{loop} = B \sum_{i=1}^{3} (r_i - r_{i,0})^2 \tag{15}$$

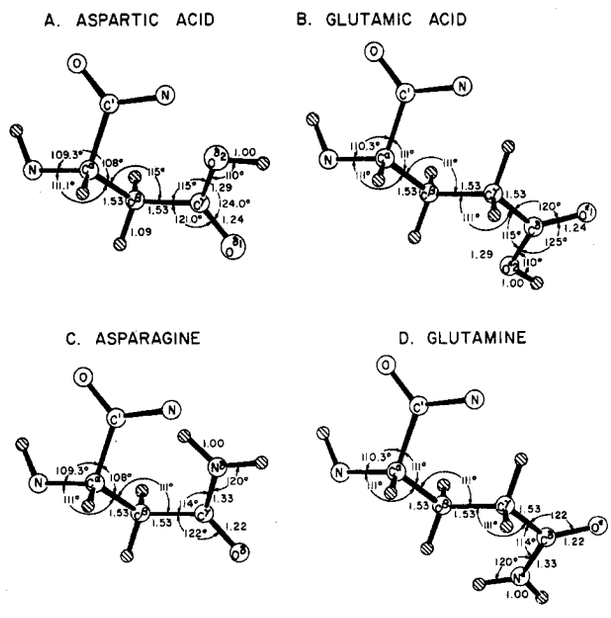

Figure 13. Geometry of several L amino acid residues. While the geometry of the COOH group would be expected to be the same for aspartic and glutamic acids, the small differences shown here appeared in the X-ray results.

where A is a constant, chosen as 10 kcal/mol Å^2 (to make the average value of the barrier to rotation at $\chi^{S_i S_j} = 0$ and 180° equal to ~7 kcal/mol[73]), B is arbitrarily set equal to

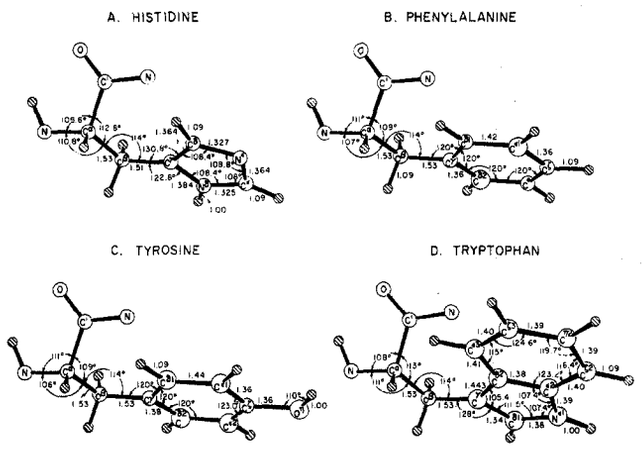

Figure 14. Geometry of several L amino acid residues. The following additional bond angles pertain to tryptophan: $\tau(C^{\beta}C^{\gamma}C^{\delta 2}) = 126.6°$; $\tau(C^{\gamma}C^{\delta 2}C^{\epsilon 3}) = 130.3°$; $\tau(C^{\epsilon 2}C^{\delta 2}C^{\epsilon 3}) = 121.4°$; $\tau(C^{\gamma}C^{\delta 2}C^{\epsilon 2}) = 108.3°$; $\tau(N^{\epsilon 1}C^{\epsilon 2}C^{\zeta 2}) = 129.4°$; and the following to tyrosine: $\tau(C^{\delta 1}C^{\epsilon 1}C^{\zeta}) = \tau(C^{\delta 2}C^{\epsilon 2}C^{\zeta}) = \tau(C^{\epsilon 1}C^{\zeta}O^{\eta}) = \tau(C^{\epsilon 2}C^{\zeta}O^{\eta}) = 118.5°$.

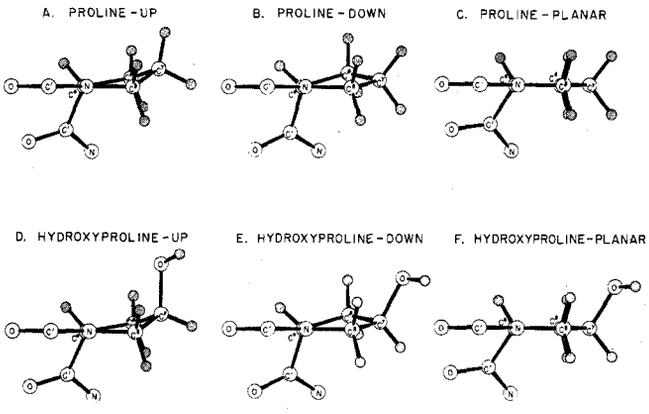

Figure 15. Geometry of L-proline and L-hydroxyproline. See Tables VII and VIII for data.

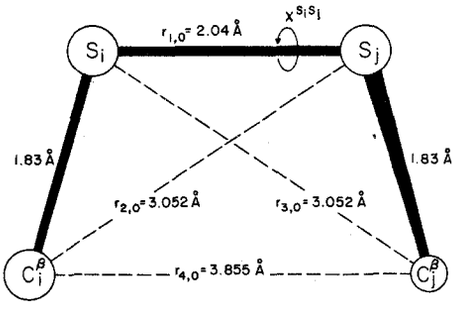

Figure 16. Relative positions of atoms involved in formation of a disulfide bond. The zero subscripts indicate the experimental values, for a conformation in which $\chi^{S_iS_j} = \pm 90°$.

100 kcal/mol Å^2 (arbitrarily, because $U_{loop} = 0$, independent of the value of B, at the experimental bond lengths and bond angles), r_1, r_2, r_3, and r_4 are the distances between atom pairs S_iS_j, $C_i^{\beta}S_j$, $S_iC_j^{\beta}$, and $C_i^{\beta}C_j^{\beta}$, and $r_{1,0}$, $r_{2,0}$, $r_{3,0}$, and $r_{4,0}$ are the respective experimental equilibrium distances when $\chi^{S_iS_j} = \pm 90°$ (see Figure 16). The distances r_1, r_2, and r_3, together with the $C_i^{\beta}S_i$ and $C_j^{\beta}S_j$ bond lengths (1.83 Å), completely define the bond angles and bond lengths around the two sulfur atoms S_i and S_j, while the distance r_4 is used to estimate the torsional energy due to variation of the dihedral angle $\chi^{S_iS_j}$ away from the two minima (±90°).

The simple form of this algorithm avoids the calculation of bond angles, and directs the juxtaposition of the two half-cystines so that the disulfide bond forms with the experimental bond length, bond angles, and SS torsion angle.

In using this algorithm, the energies of eq 14 and 15 are added to the total energy of section VE, and this new total is minimized.

Addendum

When these potential functions were applied to the cis and trans peptide conformations of several model compounds, and the librational entropy included in the free energy, the trans conformation was found to be much more stable than the cis for both nonproline and proline residues, but less so for proline.[154] For example, the probabilities of formation of the cis conformation were 8.5×10^{-7} and 4.1×10^{-2} for the *N*-acetyl-*N'*-methylamides of Gly–Gly and Gly–Pro, respectively.[154]

Miniprint Material Available. Full-sized photocopies of the miniprinted material from this paper only or microfiche (105 × 148 mm, 24× reduction, negatives) containing all of the miniprinted and supplementary material for the papers in this issue may be obtained from the Journals Department, American Chemical Society, 1155 16th St., N.W., Washington, D.C. 20036. Remit check or money order for $4.50 for photocopy or $2.50 for microfiche, referring to code number JPC-75-2361.

References and Notes

(1) This work was supported by research grants from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312), from the National Science Foundation (BMS71-00872 A04), and from Walter and George Todd.
(2) (a) Special Fellow of the National Institute of General Medical Sciences, National Institutes of Health, 1968–1969. (b) NIH Postdoctoral Trainee, 1968–1969; Postdoctoral Fellow of the National Institute of General Medical Sciences, National Institutes of Health, 1969–1971. (c) Recipient of an Australian–American Travel Grant. (d) To whom requests for reprints should be addressed.
(3) R. A. Scott and H. A. Scheraga, *J. Chem. Phys.*, **45**, 2091 (1966).
(4) (a) H. A. Scheraga, *Adv. Phys. Org. Chem.*, **6**, 103 (1968); (b) Harvey Lectures, **63**, 99 (1968); (c) Nobel Symposium 11 on "Symmetry and Function of Biological Systems at the Macromolecular Level", A. Engstrom and B. Strandberg, Ed., Almqvist and Wiksell, Stockholm, p 43, 1969; (d) *Chem. Rev.*, **71**, 195 (1971); (e) *Jerusalem Symp. Quantum Chem. Biochem.*, **5**, 51 (1973); (f) "Current Topics in Biochemistry, 1973" C. B. Anfinsen and A. N. Schechter, Ed., Academic Press, New York, N.Y., 1974, p 1; (g) "Peptides, Polypeptides, and Proteins", E. R. Blout, F. A. Bovey, M. Goodman, and N. Lotan, Ed., Wiley, New York, N.Y., 1974, p 49.
(5) G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.*, **23**, 284 (1968).
(6) A. J. Hopfinger, *Biopolymers*, **10**, 1299 (1971).
(7) P. K. Ponnuswamy and V. Sasisekharan, *Int. J. Protein Res.*, **2**, 37, 47 (1970); **3**, 9 (1971).
(8) D. R. Ferro and J. Hermans, Jr., *Biopolymers*, **11**, 105 (1972).
(9) (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).
(10) (a) A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974); (b) A. T. Hagler and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5327 (1974).
(11) R. F. McGuire, F. A. Momany, and H. A. Scheraga, *J. Phys. Chem.*, **76**, 375 (1972).
(12) J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **44**, 3289 (1966).
(13) P. K. Warme and H. A. Scheraga, *Biochemistry*, **13**, 757 (1974).
(14) P. K. Warme, F. A. Momany, S. V. Rumball, R. W. Tuttle, and H. A. Scheraga, *Biochemistry*, **13**, 768 (1974).
(15) D. Rasse, P. K. Warme, and H. A. Scheraga, *Proc. Natl. Acad. Sci., U.S.*, **71**, 3736 (1974).
(16) F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, *J. Phys. Chem.*, **75**, 2286 (1971).
(17) P. N. Lewis, F. A. Momany, and H. A. Scheraga, *Isr. J. Chem.*, **11**, 121 (1973).
(18) The results reported earlier[17] for *N*-acetyl-*N*-methylproline amide have since been modified[19] because of the use of more recent structural data on proline. The newer geometric parameters for proline are included in the present paper. In addition, some of the computed con-

formations[17] have been reconsidered[20] in light of a reinterpretation of some of the experimental data.
(19) A. W. Burgess, F. A. Momany, and H. A. Scheraga, *Proc. Natl. Acad. Sci., U.S.*, **70,** 1456 (1973); **71,** 4640 (1974).
(20) A. W. Burgess and H. A. Scheraga, *Biopolymers,* **12,** 2177 (1973).
(21) A. W. Burgess and H. A. Scheraga, *J. Phys. Chem.,* to be submitted for publication.
(22) L. L. Shipman, A. W. Burgess, and H. A. Scheraga, *Proc. Natl. Acad. Sci., U.S.*, **72,** 543 (1975).
(23) A. W. Burgess, L. L. Shipman, and H. A. Scheraga, *Proc. Natl. Acad. Sci., U.S.*, **72,** 854 (1975).
(24) A. W. Burgess, L. L. Shipman, R. A. Nemenoff, and H. A. Scheraga, *J. Am. Chem. Soc.*, in press.
(25) R. Hoffmann, *J. Chem. Phys.*, **39,** 1397 (1963); **40,** 2474, 2480, 2745 (1964).
(26) (a) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry,* **9,** 3471 (1970). (b) Atom types (designated by subscripts) are described in Table I of ref 9a.
(27) A. Warshel, M. Levitt, and S. Lifson, *J. Mol. Spectrosc.*, **33,** 84 (1970).
(28) G. C. C. Niu, N. Gō, and H. A. Scheraga, *Macromolecules,* **6,** 91 (1973).
(29) S. Arnott and W. E. Scott, *J. Chem. Soc., Perkin Trans. 2,* 324 (1972).
(30) R. E. Marsh and J. Donohue, *Adv. Protein Chem.*, **22,** 235 (1967).
(31) G. V. Gurskaya, "The Molecular Structure of Amino Acids", Consultants Bureau, New York, N.Y., 1968.
(32) O. Kennard and D. G. Watson, Ed., *Mol. Struct. Dimensions,* **1,** 343 (1970).
(33) W. C. Hamilton, M. N. Frey, L. Golic, T. F. Koetzle, M. S. Lehmann, and J. J. Verbist, *Mater. Res. Bull.*, **7,** 1225 (1972).
(34) One problem in adopting geometric parameters obtained from amino acids, dipeptides, etc. is that most of these molecules exist in a β structure in the crystals. If the geometry depends greatly on the conformation, this dependence would escape us since there are essentially no small to medium-sized molecules, for which precise crystal data exist, that are not β structures.
(35) An inconsistency in geometry has recently been brought to our attention, viz., the gas-phase structures of *N*-methylacetamide[36] and acetamide[37] both have C′–N bond lengths (~1.38 Å) which are considerably larger than those (1.29 Å in NMA and 1.334 Å in acetamide) observed in crystals of these molecules.[38–40] Since the other bond lengths and bond angles appear to be similar in the gas phase and in the crystal, this discrepancy in the peptide bond length will require further study. A possible resolution of this dilemma may lie in the recent work of F. K. Winkler and J. D. Dunitz (paper presented at the Penn State Meeting on Intermolecular Forces, Aug 1974). These workers showed that protonation of the C′=O oxygen of a peptide lengthens the C′–O bond and shortens the C′–N peptide bond, and that a peptide bond length of 1.32 Å corresponds to about 15–20% protonation, which could arise (in the crystal) from intermolecular hydrogen bond formation. Since the C′=O groups of polypeptides and proteins are very often hydrogen bonded, it appears best to use the crystal rather than the gas-phase value for the peptide bond length.
(36) M. Kitano, T. Fukuyama, and K. Kuchitsu, *Bull. Chem. Soc. Jpn.*, **46,** 384 (1973).
(37) M. Kitano and K. Kuchitsu, *Bull. Chem. Soc. Jpn.*, **46,** 3048 (1973).
(38) J. L. Katz and B. Post, *Acta Crystallogr.*, **13,** 624 (1960).
(39) F. Senti and D. Harker, *J. Am. Chem. Soc.*, **62,** 2008 (1940).
(40) W. C. Hamilton, *Acta Crystallogr.*, **18,** 866 (1965).
(41) C. Ramakrishnan and R. Balasubramanian, *Int. J. Peptide Protein Res.*, **4,** 79 (1972).
(42) J. F. Yan, F. A. Momany, R. Hoffmann, and H. A. Scheraga, *J. Phys. Chem.*, **74,** 420 (1970).
(43) W. Horsley, H. Sternlicht, and J. S. Cohen, *J. Am. Chem. Soc.*, **92,** 680 (1970).
(44) J. A. Pople, *Mol. Phys.*, **7,** 301 (1963).
(45) H. Spiesecke and W. G. Schneider, *Tetrahedron Lett.*, **No. 14,** 468 (1961).
(46) G. Del Re, B. Pullman, and T. Yonezawa, *Biochim. Biophys. Acta,* **75,** 153 (1963).
(47) D. Poland and H. A. Scheraga, *Biochemistry,* **6,** 3791 (1967).
(48) F. T. Hesselink, T. Ooi, and H. A. Scheraga, *Macromolecules,* **6,** 541 (1973).
(49) R. S. Mulliken, *J. Am. Chem. Soc.*, **72,** 4493 (1950).
(50) R. S. Mulliken, *Rec. Chem. Prog.*, **13,** 67 (1952).
(51) H. E. Simmons and J. K. Williams, *J. Am. Chem. Soc.*, **86,** 3222 (1964).
(52) R. A. Bonham and T. G. Strand, *J. Chem. Phys.*, **39,** 2200 (1963).
(53) T. G. Strand and R. A. Bonham, *J. Chem. Phys.*, **40,** 1686 (1964).
(54) R. A. Bonham, *J. Mol. Spectrosc.*, **14,** 404 (1964).
(55) L. S. Bartell and K. Kuchitsu, *J. Chem. Phys.*, **37,** 691 (1962).
(56) S. Mizushima, "Study Week on Molecular Forces", Pontificia Academia Scientiarum, Elsevier, New York, N.Y., 1967, p 415.
(57) The dielectric constants of crystalline amides and amide polymers are approximately[58] 4, and it is assumed here that a similar value holds for polypeptides.
(58) "Handbook of Chemistry and Physics", 41st ed, Chemical Rubber Publishing Co., Cleveland, Ohio, 1959, p 2523.
(59) R. D. Singh and D. R. Ferro, *J. Phys. Chem.*, **78,** 970 (1974).
(60) These computer programs and their 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.
(61) R. M. Pitzer and W. N. Lipscomb, *J. Chem. Phys.*, **39,** 1995 (1963).
(62) E. Clementi and D. R. Davis, *J. Chem. Phys.*, **45,** 2593 (1966).
(63) J. P. Lowe, *J. Chem. Phys.*, **45,** 3059 (1966).
(64) A. Veillard, *Theor. Chim. Acta (Berlin),* **5,** 413 (1966).
(65) W. H. Fink and L. C. Allen, *J. Chem. Phys.*, **46,** 2261 (1967).
(66) L. C. Allen, *Chem. Phys. Lett.*, **2,** 597 (1968).
(67) H. J. Monkhorst, *Chem. Phys. Lett.*, **3,** 289 (1969).
(68) J. P. Lowe, *J. Chem. Phys.*, **51,** 832 (1969).
(69) J. P. Lowe, *J. Am. Chem. Soc.*, **92,** 3799 (1970).
(70) L. Radom, W. J. Hehre, and J. A. Pople, *J. Am. Chem. Soc.*, **94,** 2371 (1972).
(71) L. Radom, W. A. Lathan, W. J. Hehre, and J. A. Pople, *Aust. J. Chem.*, **25,** 1601 (1972).
(72) L. C. Allen and J. Arents, *J. Chem. Phys.*, **57,** 1818 (1972).
(73) J. P. Lowe, *Prog. Phys. Org. Chem.*, **6,** 1 (1968).
(74) T. Miyazawa, *Bull. Chem. Soc. Jpn.*, **34,** 691 (1961).
(75) F. A. Momany, R. F. McGuire, J. F. Yan, and H. A. Scheraga, *J. Phys. Chem.*, **74,** 2424 (1970).
(76) J. E. Piercy and M. G. S. Rao, *J. Chem. Phys.*, **46,** 3951 (1967).
(77) A. L. Verma, W. F. Murphy, and H. J. Bernstein, *J. Chem. Phys.*, **60,** 1540 (1974).
(78) K. Kuchitsu, *J. Chem. Phys.*, **49,** 4456 (1968).
(79) A. A. Abdurahmanov, R. A. Rahimova, and L. M. Imanov, *Phys. Lett. A,* **32,** 123 (1970).
(80) R. E. Penn and R. F. Curl, Jr., *J. Chem. Phys.*, **55,** 651 (1971).
(81) O. Bastiansen, *Acta Chem. Scand.*, **3,** 415 (1949).
(82) H. Matsuura, M. Hiraishi, and T. Miyazawa, *Spectrochim. Acta, Part A,* **28,** 2299 (1972).
(83) A. Yokozeki and K. Kuchitsu, *Bull. Chem. Soc. Jpn.*, **44,** 2926 (1971).
(84) D. Sutter, H. Dreizler, and H. D. Rudolph, *Z. Naturforsch. A,* **20,** 1676 (1965).
(85) B. Beagley and K. T. McAloon, *Trans. Faraday Soc.*, **67,** 3216 (1971).
(86) H. E. Van Wart, L. L. Shipman, and H. A. Scheraga, *J. Phys. Chem.*, **78,** 1848 (1974).
(87) K. Aoki, K. Nagano, and Y. Iitaka, *Acta Crystallogr., Sect. B,* **27,** 11 (1971). See this paper for other references.
(88) M. S. Lehmann, J. J. Verbist, W. C. Hamilton, and T. F. Koetzle, *J. Chem. Soc., Perkin Trans. 2,* 133 (1973).
(89) S. Tanaka and H. A. Scheraga, *Macromolecules,* **7,** 698 (1974).
(90) D. J. Nelson and J. Hermans, Jr., *Biopolymers,* **12,** 1269 (1973).
(91) A. Sugihara, T. Ashida, Y. Sasada, and M. Kakudo, *Acta Crystallogr., Sect. B,* **24,** 203 (1968).
(92) R. Parthasarathy, *Acta Crystallogr., Sect. B,* **25,** 509 (1969).
(93) D. van der Helm and H. B. Nicholas, Jr., *Acta Crystallogr., Sect. B,* **26,** 1858 (1970).
(94) T. F. Koetzle and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **28,** 2083 (1972).
(95) Y. Tokuma, T. Ashida, and M. Kakudo, *Acta Crystallogr., Sect. B,* **25,** 1367 (1969).
(96) F. E. Cole, *Acta Crystallogr., Sect. B,* **26,** 622 (1970).
(97) M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *J. Am. Chem. Soc.*, **94,** 2657 (1972).
(98) O. Ando, T. Ashida, Y. Sasada, and M. Kakudo, *Acta Crystallogr.*, **23,** 172 (1967).
(99) K. Torii and Y. Iitaka, *Acta Crystallogr., Sect. B,* **26,** 1317 (1970).
(100) T. F. Koetzle, L. Golic, M. S. Lehmann, J. J. Verbist, and W. C. Hamilton, *J. Chem. Phys.*, **60,** 4690 (1974).
(101) E. Subramanian, *Acta Crystallogr.*, **22,** 910 (1967).
(102) R. Chandrasekaran and E. Subramanian, *Acta Crystallogr., Sect. B,* **25,** 2599 (1969).
(103) M. O. Chaney, O. Seely, and L. K. Steinrauf, *Acta Crystallogr., Sect. B,* **27,** 544 (1971).
(104) L. Golic and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **28,** 1265 (1972).
(105) C. M. Weeks, A. Cooper, and D. A. Norton, *Acta Crystallogr., Sect. B,* **25,** 443 (1969).
(106) K. Torii and Y. Iitaka, *Acta Crystallogr., Sect. B,* **27,** 2237 (1971).
(107) A. McL. Mathieson, *Acta Crystallogr.*, **5,** 332 (1952); **6,** 399 (1953).
(108) K. Torii and Y. Iitaka, *Acta Crystallogr., Sect. B,* **29,** 2799 (1973).
(109) M. M. Harding and H. A. Long, *Acta Crystallogr., Sect. B,* **24,** 1096 (1968).
(110) R. R. Ayyar, *Z. Kristallogr.*, **126,** 227 (1968).
(111) S. C. Gupta, A. Sequeira, and R. Chidambaram, *Acta Crystallogr., Sect. B,* **30,** 562 (1974).
(112) M. O. Chaney and L. K. Steinrauf, *Acta Crystallogr., Sect. B,* **30,** 711 (1974).
(113) D. D. Jones, I. Bernal, M. N. Frey, and T. F. Koetzle, *Acta Crystallogr., Sect. B,* **30,** 1220 (1974).
(114) M. N. Frey, M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **29,** 876 (1973).
(115) M. Mallikarjunan, S. T. Rao, K. Venkatesan, and V. R. Sarma, *Acta Crystallogr., Sect. B,* **25,** 220 (1969).
(116) S. Raman, *Z. Kristallogr.*, **111,** S. 301 (1959).
(117) D. A. Wright and R. E. Marsh, *Acta Crystallogr.*, **15,** 54 (1962).
(118) R. R. Bugayong, A. Sequeira, and R. Chidambaram, *Acta Crystallogr., Sect. B,* **28,** 3214 (1972).
(119) T. F. Koetzle, M. S. Lehmann, J. J. Verbist, and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **28,** 3207 (1972).

(120) A. Chiba, T. Ueki, T. Ashida, Y. Sasada, and M. Kakudo, *Acta Crystallogr., Sect. B*, **22**, 863 (1967).
(121) S. T. Rao, R. Srinivasan, and V. Valambal, *Ind. J. Pure Appl. Phys.*, **6**, 523 (1968).
(122) M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *J. Cryst. Mol. Struct.*, **2**, 225 (1972).
(123) M. Ramanadham, S. K. Sikka, and R. Chidambaram, *Acta Crystallogr., Sect. B*, **28**, 3000 (1972).
(124) J. J. Verbist, M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **28**, 3006 (1972).
(125) G. Kartha and A. de Vries, *Nature (London)*, **192**, 862 (1961).
(126) T. F. Koetzle, M. N. Frey, M. S. Lehmann, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **29**, 2571 (1973).
(127) K. Oda and H. Koyama, *Acta Crystallogr., Sect. B*, **28**, 639 (1972). See this paper for other references.
(128) M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *Int. J. Peptide Protein Res.*, **4**, 229 (1972).
(129) C. H. Wei, D. G. Doherty, and J. R. Einstein, *Acta Crystallogr., Sect. B*, **28**, 907 (1972).
(130) R. E. Stenkamp and L. H. Jensen, *Acta Crystallogr., Sect. B*, **29**, 2872 (1973).
(131) P. M. Cotrait and Y. Barrans, *Acta Crystallogr., Sect. B*, **30**, 1018 (1974).
(132) Y. Harada and Y. Iitaka, *Acta Crystallogr., Sect. B*, **30**, 726 (1974).
(133) P. M. Cotrait and J. P. Bideau, *Acta Crystallogr., Sect. B*, **30**, 1024 (1974).
(134) D. W. Smits and E. H. Wiebenga, *Acta Crystallogr.*, **6**, 531 (1953).
(135) J. A. Hamilton and L. K. Steinrauf, *Acta Crystallogr.*, **23**, 817 (1967).
(136) T. Takigawa, T. Ashida, Y. Sasada, and M. Kakudo, *Bull. Chem. Soc. Jpn.*, **39**, 2369 (1966).
(137) J. Zussman, *Acta Crystallogr.*, **4**, 493 (1951).
(138) J. Donohue and K. N. Trueblood, *Acta Crystallogr.*, **5**, 419 (1952).
(139) J. Fridrichsons and A. McL. Mathieson, *Acta Crystallogr.*, **15**, 569 (1962).
(140) R. L. Kayushina and B. K. Vainshtein, *Kristallografiya*, **10**, 833 (1965).
(141) T. Ueki, T. Ashida, M. Kakudo, Y. Sasada, Y. Katsube, *Acta Crystallogr., Sect. B*, **25**, 1840 (1969).
(142) T. Matsuzaki and Y. Iitaka, *Acta Crystallogr., Sect. B*, **27**, 507 (1971).
(143) J. J. Verbist, M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *Nature (London)*, **235**, 328 (1972).
(144) T. F. Koetzle, M. S. Lehmann, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **29**, 231 (1973).
(145) G. Kartha, T. Ashida, and M. Kakudo, *Acta Crystallogr., Sect. B*, **30**, 1861 (1974).
(146) E. Benedetti, M. R. Ciajolo, and A. Maisto, *Acta Crystallogr., Sect. B*, **30**, 1783 (1974).
(147) I. L. Karle, *J. Am. Chem. Soc.*, **94**, 81 (1972).
(148) G. Kartha, G. Ambady, and P. V. Shankar, *Nature (London)*, **247**, 204 (1974).
(149) G. N. Ramachandran and V. Sasisekharan, *Adv. Protein Chem.*, **23**, 283 (1968).
(150) W. R. Krigbaum, R. J. Roe, and J. D. Woods, *Acta Crystallogr., Sect. B*, **24**, 1304 (1968).
(151) P. Ganis, G. Avitabile, E. Benedetti, C. Pedone, and M. Goodman, *Proc. Natl. Acad. Sci., U.S.*, **67**, 426 (1970).
(152) J. J. Stezowski and R. E. Hughes, private communication.
(153) D. D. Jones, I. Bernal, M. N. Frey, and T. F. Koetzle, *Acta Crystallogr., Sect. B*, **30**, 1220 (1974).
(154) S. S. Zimmerman and H. A. Scheraga, *J. Am. Chem. Soc.*, to be submitted for publication.

Mean Activity Coefficients for the Simple Electrolyte in Aqueous Mixtures of Polyelectrolyte and Simple Electrolyte. The Systems Potassium Chloride–Potassium Poly(styrenesulfonate), Magnesium Chloride–Magnesium Poly(styrenesulfonate), and Calcium Chloride–Calcium Poly(styrenesulfonate)

Jan C. T. Kwak,* Mary C. O'Brien, and David A. MacLean

Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, Canada *(Received May 27, 1975)*

Publication costs assisted by the National Research Council of Canada

Mean molal activity coefficients of the simple electrolyte are reported in aqueous solutions of the K, Mg, or Ca salts of poly(styrenesulfonic acid) (PSA) with added KCl, $MgCl_2$, or $CaCl_2$. PSA concentrations range from 0.005 to 0.07 *m* in KCl–KPSA and from 0.006 to 0.05 *m* in $MgCl_2$–$Mg(PSA)_2$ and $CaCl_2$–$Ca(PSA)_2$; PSA:Cl^- ratios are varied from excess PSA to excess Cl^-. An electrochemical cell with a cation-exchange membrane as cation-selective electrode and an Ag–AgCl electrode is used in all systems. In the divalent ion systems, liquid ion-exchange membrane electrodes for Mg^{2+} and Ca^{2+} are used as well. The results for log $\gamma_\pm$ are compared to a "limiting law" derived by Manning, applying a correction for Debye–Hückel type interactions between the small ions. In the $MgCl_2$–$Mg(PSA)_2$ and $CaCl_2$–$Ca(PSA)_2$ systems good agreement of the experimental data with the corrected form of the limiting law is found over the complete concentration range studied. In the KCl–KPSA systems, experimental results for $\gamma_\pm$ in mixtures with a large excess of polyelectrolyte are significantly higher than the corrected limiting law values. This was also observed in the NaCl–NaPSA system studied earlier.

The measurement or prediction of the activities of the simple ions in mixtures of polyelectrolytes and simple electrolytes is of importance for the calculation of electrical potentials across cell membranes and for the description of ionic transport across these membranes. It is well known that very considerable deviations from ideality occur even in fairly dilute polyelectrolyte solutions. These deviations could appreciably alter for instance the permeability ratio of various small ions calculated from resting potentials across membranes in living cells. This paper describes the results of an experimental program designed to measure the mean activity coefficient of the added electrolyte in various mixtures of polyelectrolyte and simple electrolyte, using thermodynamically well-defined electrochemical methods and polyions of which the structural parameters are known. Comparison of our results for the mean activity