Energy Parameters in Polypeptides. I. Charge Distributions and the Hydrogen Bond*

Douglas Poland and Harold A. Scheraga

ABSTRACT: An empirical potential function for the hydrogen bond, based on electrostatic interactions and on an empirical 6–12 potential, is obtained for use in conformational energy calculations. The electrostatic interactions are evaluated in terms of the total charge on each atom, the total charge being the sum of the σ and π charges. The σ charge distribution in homopolymers of all the important amino acids is calculated by the molecular orbital–linear combination of atomic orbital (MO–LCAO) method of Del Re, using the parameters of Berthod and Pullman to deal with the unsaturated groups. Using literature data for the magnitude and direction of dipole moments in small molecules, the π charge density can be calculated empirically for π systems involving three atoms or less; this is done for the carboxyl and amide groups. π charge densities for the aromatic side chains are taken from MO calculations in the literature. The hydrogen-bond potential is constructed by use of two adjustable parameters, determined from the known energy and equilibrium length of the hydrogen bond, and the requirement that the potential go over smoothly to the dipole–dipole potential (calculated from the distribution of charges) at large distances of separation.

In the calculation of the conformation of biological macromolecules, it is necessary to have accurate expressions for the intramolecular energies. In the case of polypeptides, in the absence of solvent, three factors have emerged as being especially important in determining the conformation (Scheraga *et al.*, 1967); these are the hydrogen bond, nonbonded interactions, and electrostatic interactions. All three of these factors are interrelated, as is seen very clearly when one considers our lack of knowledge about the exact nature of the hydrogen bond. If the hydrogen bond is a weak covalent bond, then it is difficult to fit this concept into our simple ideas of valence, because all the atoms involved in the bond have their valence already satisfied. As a result, many authors prefer to think of the hydrogen bond as arising solely from strong electrostatic interactions plus a repulsion from the hard core. However, hydrogen-bond distances are very short and, by analogy with other atom pairs, one would expect enormous repulsive forces at these distances, implying that there must be some softening of the repulsive potential if the only attractive forces are electrostatic. All of these questions are reviewed by Pimentel and McClellan (1960).

Our viewpoint is that quantitative theoretical estimates of the hydrogen-bond energy, *i.e.*, the energy at very short distances, cannot be made at the present time. Thus, to obtain a suitable expression for the hydrogen-bond energy in order to carry out conformational calculations, we must rely on semiempirical equations with adjustable parameters selected to fit the known interaction properties of small molecules. We differ from previous workers (see Pimentel and McClellan, 1960) in that we include, as an important factor in conformational calculations, the long-range attractive tail of the potential function, which arises from the total electrostatic interactions of the molecules involved.

Our approach will be divided into two parts. In the first part, we will calculate the charge distributions in polyamino acids, evaluating the σ charges by the molecular orbital–linear combination of atomic orbital (MO–LCAO) method of Del Re (1958, 1963a,b) and the π charges of the peptide and carboxyl groups by empirically fitting dipole moment data of small molecules; the π charges of the aromatic side chains are taken from the literature (Pullman and Pullman, 1963). These will enable us to evaluate the electrostatic energy. In the second part, we give an empirical potential function for the hydrogen bond, incorporating the nonbonded and electrostatic interactions, and an additional empirical potential function (6–12 and variants thereof) to adjust the total interaction potential to comply with known facts. Since there is no sufficiently detailed knowledge about the *directional* character of the hydrogen bond, we do not introduce explicitly any factor (other than that arising simply from the coulombic interaction between partial charges) which would provide a preference for a given orientation.

Charge Distribution

Calculation of σ Charges. If one accepts the procedure of classifying the bonds of any molecule into σ

* From the Department of Chemistry, Cornell University, Ithaca, New York 14850. *Received July 28, 1967.* This work was supported by a research grant (GB-4766) from the National Science Foundation, and by a research grant (GM-14312) from the National Institute of General Medical Sciences of the National Institutes of Health, U. S. Public Health Service.

TABLE I: Parameters for Calculation of σ Charges.[a]

A	B	δ_A^0	δ_B^0	$(\epsilon_{AB} = \epsilon_{BA})^b$	γ_{AB}	γ_{BA}
Saturated Bonds (Del Re 1958, 1963a,b)						
C	N	0.07	0.24	1.00	0.1	0.1
	O		0.40	0.95	0.1	0.1
	C		0.07	1.00	0.1	0.1
	H		0.00	1.00	0.3	0.4
	S		0.07	0.75	0.2	0.4
H	N	0.00	0.24	0.45	0.4	0.3
	O		0.40	0.45	0.4	0.3
	S		0.07	0.70	0.4	0.3
S	S	0.07	0.07	0.60	0.1	0.1
Unsaturated Bonds[c] (Berthod and Pullman, 1965)						
C, =C< (as in benzene)	C	0.12	0.12	1[d]	0.1	0.1
	H		0.00	1	0.3	0.4
N,  N (as in pyridine)	C	0.38	0.12	0.7	0.1	0.1
N,  NH (as in pyrrole)	C	0.30	0.12	0.7	0.1	0.1
	H		0.00	0.6	0.3	0.4
N,  (as in aniline)	C	0.24	0.12	1	0.1	0.1
	H		0.00	0.45	0.3	0.4
O, O= (as in *N*-methylpyridone)	C	0.28	0.12	0.7	0.1	0.1
O, –OH (as in phenol)	C	0.40	0.12	0.95	0.1	0.1
	H		0.00	0.45	0.3	0.4

[a] The γ's were taken from Del Re (1958, 1963a,b), with the Berthod–Pullman (1965) assumption that they are unaffected by the state of hybridization. [b] This equality follows from the fact that $q_A^{(B)} = -q_B^{(A)}$, and eq 1 and 2. [c] Presumably, the magnitudes of these parameters used to calculate the σ charges are not too seriously affected by the nature of the neighboring unsaturated bonds. [d] We assume this value also holds if carbon B is aliphatic.

and π bonds, one can compute the total charge on any atom by computing the σ and π charges separately and then adding the two together.[1] The total point charge distribution (monopole approximation) should give the direction and magnitude of the experimentally determined dipole moments and also higher moments if these are known. In general, both σ and π charges must be considered in computing the total dipole moment.

Much attention has been paid to the calculation of π charge densities by the MO scheme, and we will discuss the evaluation of π charges in the next section. However, little attention has been given to the σ charge density. Since polypeptides are largely saturated, the σ charges are very important in their influence on the energy of a protein. Sometime ago Del Re (1958) proposed a simple MO–LCAO method of treating saturated molecules that was capable of reproducing dipole moments and known trends of inductive effects in small molecules fairly well. Del Re's method contains many adjustable parameters; some of these are evaluated by adjustment to give the known dipole moments of a small class of reference compounds and others by theoretical and semiempirical guesses. The parameters evaluated by Del Re (1958, 1963a,b) apply only to completely saturated molecules. However, in proteins and nucleic acids, many situations arise where it is necessary to know the σ charge density in a partially unsaturated system; Berthod and Pullman (1965) recently evaluated the necessary parameters for treating such systems. We use the Del Re method

[1] Bradley *et al.* (1963) apply this method to nucleic acids.

TABLE II: σ and Total Charges for Backbone Amide Groups.

	Pro and Hypro			All Others		
	σ	π^a	Total	σ	π^a	Total
H				+0.204		+0.204
N	−0.188	+0.140	−0.048	−0.342	+0.140	−0.202
C′	+0.103	+0.208	+0.311	+0.110	+0.208	+0.318
O	−0.077	−0.348	−0.425	−0.074	−0.348	−0.422
		Sum	−0.162[b]		Sum	−0.102[b]

[a] The π charges come from Table V. [b] The group is not neutral since it is only the *whole molecule* which must be neutral.

here, together with the parameters of Berthod and Pullman, to obtain the σ charges in homopolymers of amino acids. In the calculations, all ionizable groups were considered to be in their uncharged form. The nice feature of the Del Re method is that the σ charge of an atom in a given molecule depends on the complete connexity of *all* the atoms in the molecule; one feels that the situation should be at least this complex in order to approach reality.

According to Del Re (1963b), the σ charge (Q_A) on any atom A may be computed as follows. If atom A is covalently bonded to atom B, then $q_A^{(B)}$ is the contribution to the total charge Q_A due to the bond AB, and is given by

$$q_A^{(B)} = \frac{\tilde{Q}_A^{(B)}}{[1 + (\tilde{Q}_A^{(B)})^2]^{1/2}} \tag{1}$$

where

$$\tilde{Q}_A^{(B)} = \frac{\delta_B - \delta_A}{2\epsilon_{AB}} \tag{2}$$

The ϵ_{AB}'s are constants; the δ's are evaluated from the system of linear equations

$$\delta_A = \delta_A^0 + \sum_{\substack{\text{linked} \\ \text{atoms B}}} \gamma_{AB}\delta_B \tag{3}$$

where δ_A^0 and the γ_{AB}'s are constants. The constants arise from the Coulomb and overlap integrals between the atomic orbitals for atoms A and B, and are treated as adjustable parameters. Even though δ_A is a function of all the δ_B's, in eq 3, this system of equations can be solved, as will be illustrated below. Equation 3 thus make the δ of any atom depend on the connexity of the whole molecule. The total σ charge on A is then given by

$$Q_A = \sum_B q_A^{(B)} \tag{4}$$

Berthod and Pullman (1965) point out that the parameters δ^0 and ϵ_{AB} (but presumably not the γ's) should depend on the type of hybridization involved. The δ^0 and ϵ values for carbon, nitrogen, and oxygen in unsaturated systems thus will differ from the parameters in saturated ones. Table I gives a compilation of parameters taken from Del Re (1958, 1963a,b) (for saturated systems) and from Berthod and Pullman (1965) (for unsaturated systems). Berthod and Pullman give three sets of values for the parameters involving unsaturated nitrogen (Table I), evaluated from aniline, pyrrole, and pyridine. We have used the values given for aniline to compute the σ charge on the amide group, feeling that aniline is a better analog than pyrrole or pyridine.

As an example of the application of eq 1–4, consider the case of polyglycine having a linear, infinite structure with the repeat unit shown in Figure 1. Equation 3 becomes

$$\begin{aligned}
\delta_H &= \delta_H^0 + \gamma_{HN}\delta_N \\
\delta_N &= \delta_N^0 + \gamma_{NH}\delta_H + \gamma_{NC\alpha}\delta_{C\alpha} + \gamma_{NC'}\delta_{C'} \\
\delta_{C'} &= \delta_{C'}^0 + \gamma_{C'N}\delta_N + \gamma_{C'O}\delta_O + \gamma_{C'C\alpha}\delta_{C\alpha} \\
\delta_O &= \delta_O^0 + \gamma_{OC'}\delta_{C'} \\
\delta_{C\alpha} &= \delta_{C\alpha}^0 + \gamma_{C\alpha C'}\delta_{C'} + 2\gamma_{C\alpha H\alpha}\delta_{H\alpha} + \gamma_{C\alpha N}\delta_N \\
\delta_{H\alpha} &= \delta_{H\alpha}^0 + \gamma_{H\alpha C\alpha}\delta_{C\alpha}
\end{aligned} \tag{5}$$

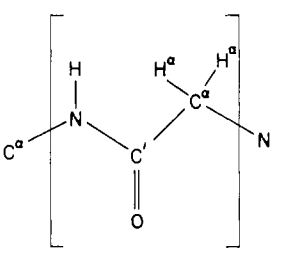

FIGURE 1: The repeat unit of a polyglycine chain.

TABLE III: σ Charges in Saturated Side Chains.[a-d]

Gly,	C^{α}, 0.000														
Ala	C^{α}, 0.046	C^{β}, −0.110	H^{β}, 0.040												
Val	C^{α}, 0.043	C^{β}, −0.023	H^{β}, 0.038	C^{γ}, −0.115	H^{γ}, 0.038										
Leu	C^{α}, 0.044	C^{β}, −0.067	H^{β}, 0.038	C^{γ}, −0.031	H^{γ}, 0.038	C^{δ}, −0.116	H^{δ}, 0.038								
Ile	C^{α}, 0.043	C^{β}, −0.024	H^{β}, 0.038	$C^{\gamma 1}$, −0.075	$H^{\gamma 1}$, 0.038	$C^{\gamma 2}$, −0.115	$H^{\gamma 2}$, 0.038	C^{δ}, −0.115	H^{δ}, 0.038						
Met	C^{α}, 0.046	C^{β}, −0.059	H^{β}, 0.038	C^{γ}, −0.024	H^{γ}, 0.035	S^{δ}, −0.058	C^{ε}, −0.135	H^{ε}, 0.050							
CysSH	C^{α}, 0.056	C^{β}, −0.094	H^{β}, 0.046	S^{γ}, −0.062	H^{γ}, 0.064										
Cys	C^{α}, 0.056	C^{β}, −0.096	H^{β}, 0.046	S^{γ}, 0.004											
Ser	C^{α}, 0.068	C^{β}, +0.040	H^{β}, 0.053	O^{γ}, −0.460	H^{γ}, 0.302										
Thr	C^{α}, 0.063	C^{β}, +0.089	H^{β}, 0.050	$O^{\gamma 1}$, −0.463	$H^{\gamma 1}$, 0.302	$C^{\gamma 2}$, −0.105	$H^{\gamma 2}$, 0.040								
Lys	C^{α}, 0.044	C^{β}, −0.067	H^{β}, 0.038	C^{γ}, −0.072	H^{γ}, 0.038	C^{δ}, −0.065	H^{δ}, 0.038	C^{ε}, −0.013	H^{ε}, 0.046	N^{ζ}, −0.505	H^{ζ}, 0.207				
Pro	C^{α}, 0.037	C^{β}, −0.067	H^{β}, 0.038	C^{γ}, −0.066	H^{γ}, 0.038	C^{δ}, −0.030	H^{δ}, 0.045								
Hypro	C^{α}, 0.037	C^{β}, −0.053	H^{β}, 0.040	C^{γ}, +0.088	H^{γ}, 0.050	O^{δ}, −0.462	H^{δ}, 0.302	C^{δ}, −0.020	H^{δ}, 0.047						

[a] These are also the *total* charges on the side chains. [b] Where a given side-chain carbon or nitrogen atom contains more than one H atom, the charges on all the H atoms are the same; similarly for identical C atoms, as for $C^{\gamma 1}$ and $C^{\gamma 2}$ in valine. [c] The net charge on each side chain (including the C^{α} and H^{α} atoms) is +0.102 to neutralize the charge of −0.102 of the amide group (for all amino acids except proline and hydroxyproline; for the latter two, the net charge on the side chain is +0.162). [d] The σ charge on H^{α} is +0.046 in all cases except glycine, where both H^{α} atoms have a σ charge of +0.051.

This set of six equations in the six unknowns δ_H, δ_N, $\delta_{C'}$, δ_O, $\delta_{C\alpha}$, and $\delta_{H\alpha}$ can be solved, using the δ^0's and γ's from Table I, being careful to take notice of the type of hybridization involved (*e.g.*, C^{α} is different from C′). The calculated δ's are then inserted along with the ϵ's into eq 1 and 2 which, together with eq 4, give the σ charge on each atom.

It should be noted that we have taken cognizance of the fact that we are dealing with an infinite homopolymer since we assume that the charge distribution is the same in each residue; *i.e.*, the factors $\gamma_{NC\alpha}\delta_{C\alpha}$ and $\gamma_{C\alpha N}\delta_N$ actually take one outside the given residue, but we assume that all C^{α}'s and N's have the same charge. In principle, according to this method, the charge distribution of a given amino acid residue should differ depending on the nature of its neighboring residues; thus, for example, ribonuclease would involve about 1000 equations in 1000 unknowns! However, in practice, we find that the backbone charges for the different homopolymers are all about the same. Hence, one can probably use the charges computed here for each amino acid in a *heteropolymer* as an excellent approximation.

In this manner, we have computed the σ charges for homopolymers of all of the important amino acids. These are listed in Tables II and III, where the labeling of the atoms follows the convention recommended at the Workshop on Protein Conformation in Madras (Jan 1967). Before discussing the values presented in Tables II and III, we will compute the π charges, which must be added to the σ charges to obtain the total charge on each atom.

Calculation of π Charges. The π charges of arginine, histidine, phenylalanine, tyrosine, and tryptophan were taken from Pullman and Pullman (1963), and are listed in Table IV. The remaining π systems in polypeptides are the small amide and carboxyl groups. While MO calculations have been made (Pullman and Pullman, 1963) to obtain the π charges of these groups, we prefer to use a completely empirical method, applicable to π systems in such small groups; this procedure avoids the necessity of having to use *two* approximate theoretical calculations. In this method, we require a knowledge of the magnitude and direction of the total dipole moment. We compute the contribution to the dipole moment from the σ charges, and subtract the components of this contribution [$\mu_x(\sigma)$ and $\mu_y(\sigma)$] from the components of the total dipole moment [μ_x(total) and μ_y(total)] to obtain the components [$\mu_x(\pi)$ and $\mu_y(\pi)$] of the contribution of the π charges to the dipole moment. This method is applicable to planar groups of three atoms or less, and is outlined below.

Let the π charges on the three atoms of a planar group be q_1, q_2, and q_3. Then

$$q_1 + q_2 + q_3 = 0 \tag{6}$$

In addition

$$\mu_x(\pi) = \mu_x(q_1, q_2, q_3) \tag{7}$$

TABLE IV: π Charges for Arginine, Histidine, Phenylalanine,[a] Tyrosine,[a] and Tryptophan.

Arg	His[b]	Phe	Tyr	Trp
N^{ϵ}, 0.233	C^{γ}, −0.072	C^{β}, −0.048	C^{γ}, −0.038	C^{γ}, −0.170
C^{ζ}, 0.213	$N^{\delta 1}$, 0.453	H^{β}, 0.027	C^{δ}, 0.003	$C^{\delta 1}$, 0.000
$N^{\eta 1}$, −0.679	$C^{\delta 2}$, −0.104	C^{γ}, 0.040	C^{ϵ}, −0.050	$C^{\delta 2}$, −0.060
$N^{\eta 2}$, 0.233	$C^{\epsilon 1}$, 0.009	C^{δ}, −0.018	C^{ζ}, 0.056	N^{ϵ}, 0.379
	$N^{\epsilon 2}$, −0.286	C^{ϵ}, 0.001	O^{η}, 0.076	$C^{\epsilon 1}$, −0.019
		C^{ζ}, −0.012		$C^{\epsilon 2}$, −0.023
				$C^{\zeta 1}$, −0.037
				$C^{\zeta 2}$, −0.039
				C^{η}, −0.031

[a] Some π charges are assigned to the C^{β} and H^{β} atoms (Pullman and Pullman, 1963) of phenylalanine because of hyperconjugation (to account for the dipole moment). This is not necessary in tyrosine, where the dipole moment can be accounted for in terms of resonance structures involving the OH group; the latter contribute a positive charge to the O^{η} atom. [b] These charges were computed for the tautomeric structure with a hydrogen atom on $N^{\delta 1}$.

$$\mu_y(\pi) = \mu_y(q_1, q_2, q_3) \quad (8)$$

where $\mu_x(\pi)$ and $\mu_y(\pi)$ are obtained as indicated above, and $\mu_x(q_1, q_2, q_3)$ and $\mu_y(q_1, q_2, q_3)$ are computed as $\Sigma q_i r_i$ in the x and y directions, respectively. The three equations (6–8) permit the evaluation of the three unknowns q_1, q_2, and q_3.

Figure 2 shows the direction and magnitude of the dipole moment for methyl formate (Curl, 1959) and formamide (Kurland and Wilson, 1957), which are models for the carboxyl and amide groups, respectively. Table V gives the components of the total experimental dipole moment (computed from the data of Figure 2), the calculated components from the σ charges, and the resultant π charges and their contributions to the dipole moment. It can be seen that the contributions from the π charges account for most of the dipole moment. Resonance diagrams, shown in Figure 3, easily account for the signs of the π charges. We will assume that the π charges are independent of the nature of the R groups (as long as the R groups are not part of the unsaturated system). It should be noted that the π charges for the carboxyl group pertain to the COOH group, and not the COO^- group, since they were obtained from the neutral ester group.

We may now compute the charge distribution for the amino acids containing partially unsaturated side chains. The σ charges were computed by the method of Del Re, and the π charges are those of Tables IV and V. The results are shown in Table VI, where the entries in parentheses are the total charges and the other entries are the σ charges.

TABLE V: Components of the Dipole Moment Contributed by the σ and π Charges.[a]

	Dipole Moment (Debye units) μ_x	μ_y	Atom	$q(\pi)$
	Methyl formate			
σ	+0.08	+0.63	O′	−0.225
π	+1.04	−2.00	C	+0.036
			O	+0.189
Total	+1.12	−1.37		
	Formamide			
σ	−0.80	−0.64	O	−0.348
π	−2.06	−1.72	N	+0.140
			C	+0.208
Total	−2.86	−2.36		

[a] The σ charges on methyl formate are: C, −0.024; H(C), +0.054; O, −0.254; C′ +0.107; O′, −0.058; and H(C′), +0.067; for formamide they are: H(C), +0.063; C, +0.070; O, −0.065; N, −0.516; and H(N), +0.224.

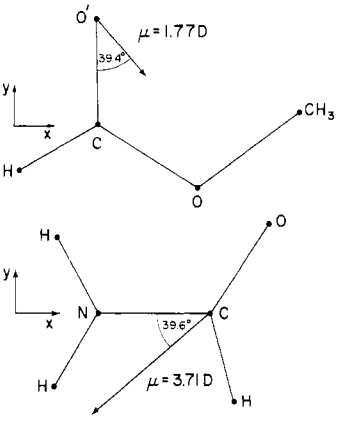

FIGURE 2: Direction and magnitude of the total dipole moment in methyl formate (Curl, 1959) and in formamide (Kurland and Wilson, 1957).

TABLE VI: σ Charges and Total Charges[a,b] for Amino Acids Containing Partially Unsaturated Side Chains.

Arg	His	Phe	Tyr	Trp	Asp	Asn	Glu	Gln
C^{α}, 0.044	C^{α}, 0.049	C^{α}, 0.047	C^{α}, 0.047	C^{α}, 0.045	C^{α}, 0.050	C^{α}, 0.049	C^{α}, 0.045	C^{α}, 0.045
C^{β}, −0.064	C^{β}, −0.049	C^{β}, −0.054 (−0.102)	C^{β}, −0.055	C^{β}, −0.050	C^{β}, −0.043	C^{β}, −0.045	C^{β}, −0.064	C^{β}, −0.065
H^{β}, 0.038	H^{β}, 0.041	H^{β}, 0.040 (0.067)	H^{β}, 0.040	H^{β}, 0.041	H^{β}, 0.042	H^{β}, 0.041	H^{β}, 0.039	H^{β}, 0.039
C^{γ}, −0.064	C^{γ}, 0.120 (0.048)	C^{γ}, −0.009 (0.031)	C^{γ}, −0.006 (−0.044)	C^{γ}, −0.050 (−0.202)	C^{γ}, 0.171 (0.207)	C^{γ}, 0.116 (0.324)	C^{γ}, −0.048	C^{γ}, −0.040
H^{γ}, 0.038	$N^{\delta 1}$, −0.440 (0.013)	C^{δ}, −0.057 (−0.075)	C^{δ}, −0.056 (−0.053)	$C^{\delta 1}$, 0.062 (0.062)	$O^{\delta 1}$, −0.460 (−0.271)	$N^{\delta 1}$, −0.487 (−0.347)	H^{γ}, 0.041	H^{γ}, 0.035
C^{δ}, −0.020	$H^{\delta 1}$, 0.192	H^{δ}, 0.053	H^{δ}, 0.053	$H^{\delta 1}$, 0.058	$H^{\delta 1}$, 0.319	$H^{\delta 1}$, 0.207	C^{δ}, 0.167 (0.203)	C^{δ}, 0.114 (0.322)
H^{δ}, 0.045	$C^{\epsilon 1}$, 0.182 (0.191)	C^{ϵ}, −0.053 (−0.052)	C^{ϵ}, −0.044 (−0.094)	N^{ϵ}, −0.454 (−0.075)	$O^{\delta 2}$, −0.055 (−0.290)	$O^{\delta 2}$, −0.073 (−0.421)	$O^{\epsilon 1}$, −0.459 (−0.270)	$N^{\epsilon 1}$, −0.486 (−0.346)
N^{ϵ}, −0.331 (−0.098)	$H^{\epsilon 1}$, 0.068	H^{ϵ}, 0.053	H^{ϵ}, 0.054	H^{ϵ}, 0.188			$H^{\epsilon 1}$, 0.319	$H^{\epsilon 1}$, 0.207
H^{ϵ}, 0.207	$N^{\epsilon 2}$, −0.293 (−0.579)	C^{ζ}, −0.053 (−0.065)	C^{ζ}, 0.120 (0.176)	$C^{\epsilon 1}$, 0.129 (0.110)			$O^{\epsilon 2}$, −0.064 (−0.289)	$O^{\epsilon 2}$, −0.074 (−0.422)
C^{ζ}, 0.259 (0.472)	$C^{\delta 2}$, 0.083 (−0.021)	H^{ζ}, 0.053	O^{η}, −0.449 (−0.373)	$C^{\delta 2}$, 0.023 (−0.037)				
$N^{\eta 1}$, −0.384 (−1.063)	$H^{\delta 2}$, 0.062		H^{η}, 0.305	$C^{\zeta 1}$, −0.045 (−0.082)				
$H^{\eta 1}$, 0.224				$H^{\zeta 1}$, 0.053				
$N^{\eta 2}$, −0.511 (−0.278)				C^{η}, −0.053 (−0.084)				
$H^{\eta 2}$, 0.227				H^{η}, 0.053				
				$C^{\zeta 2}$, −0.053 (−0.092)				
				$H^{\zeta 2}$, 0.053				
				$C^{\epsilon 2}$, −0.056 (−0.079)				
				$H^{\epsilon 2}$, 0.053				

[a] The numbers in parentheses are the total charges, *i.e.*, where applicable, the π charge has been added to the σ charge; this often changes the sign of the charge. [b] The σ charge on H^{α} is +0.046 in all cases.

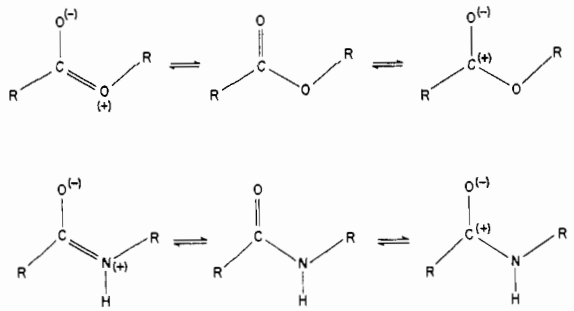

FIGURE 3: Resonance structures of esters and amides, showing π charge distribution in accordance with data of Table V.

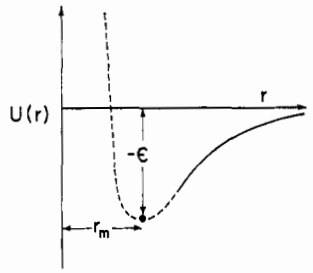

FIGURE 4: Schematic representation of hydrogen-bond potential. The solid line (dipole–dipole interaction potential) and the dot (position of minimum, and energy at the minimum) are the only features that are known. The dashed line is an empirical potential constructed to fit the known data.

Discussion of Charges. Tables II–VI contain the σ, π, and total charges on every atom of the various amino acids which are commonly found in proteins. Examination of these charges reveals that every atom has a net charge, even hydrocarbon groups, which have their symmetry affected (and thus contribute a net dipole moment) when they occur in polar molecules. The amide group (HNCO) has a net negative charge, which is the same for all amino acids except proline and hydroxyproline. This is neutralized by the net positive charge of the side chain (see footnote *c* of Table III). For all practical purposes, the total charges on the amide C′ and O atoms are identical for all amino acids, only the charge on N in proline and hydroxyproline being different. The charge on H^{α} is the same for all amino acids (except for glycine, with a slightly larger value), while the charge on C^{α} varies, being 0 in glycine and +0.040 to +0.070 in all other amino acids.

The charges are given to three significant figures in the tables, even though they are certainly not this accurate, in order to avoid roundoff errors in their computation. If the numbers are subsequently rounded off for conformational calculations, care must be taken to leave the residues with net charges of zero.

Hydrogen-Bond Potential

Form of the Potential. As indicated in the introduction, the theoretical aspects of the hydrogen bond are not yet well understood. For a review of the present status of this problem, the reader is referred to the book by Pimentel and McClellan (1960). Our sole purpose here is to obtain a completely empirical potential function which will be of use for calculations of polypeptide conformation, *i.e.*, we seek a function which is continuous and of simple mathematical form. For this purpose we seek a function U(*r*), where *r* is the distance between the hydrogen and the acceptor atom, which will have three properties: (a) the potential should have a minimum at the experimentally measured distance (r_m); (b) the value of the potential at r_m should be the experimental energy of formation of the hydrogen bond ($-\epsilon$); and (c) the long-range attractive part of the potential should be the known total dipole–dipole interaction potential for the molecules taking part in the association. We differ from previous workers (Pimentel and McClellan, 1960) in that we require condition c as an essential part of the potential (because that is the only part of the potential function known with any certainty, and it is reasonable to expect that the hydrogen-bond potential should go over smoothly and continuously into the dipole–dipole interaction potential), we do not introduce more adjustable parameters than can be determined from experimental data, and we give no *a priori* interpretation to the empirical part of the potential. The electrostatic part of the potential will be given by the monopole approximation, which gives the correct dipole–dipole interaction at large distances; hopefully, it will provide a good estimate of the electrostatic interaction energy as *r* approaches r_m (but not necessarily all the way down to the minimum of the potential function, since we will obtain the shape and location of the minimum completely empirically). In Figure 4, the heavy line and the dot illustrate what we consider to be known with reasonable certainty about the hydrogen-bond potential; the dashed line is the empirical part of the potential.

We shall compute U(*r*) for the following situation

$$R_1AH\cdots BR_2 \tag{9}$$

where *r* is the distance between atoms H and B, all other bonds in the system remaining fixed as *r* varies. The potential U(*r*) is considered to be the sum of two functions

$$U(r) = U_{H\ldots B}(r) + S(r) \tag{10}$$

where $U_{H\ldots B}(r)$ is the interaction between atoms H and B; this function will be obtained empirically. The interaction potential between all of the remaining atoms of the system is S(*r*), which is the sum over all pairwise van der Waals and electrostatic interactions between all atoms *other* than the H···B pairwise interaction. Since then van der Waals' contribution

is very small, S(r) is essentially entirely the electrostatic interaction (calculated as the monopole potential between the partial charges, evaluated in the previous section, with unit dielectric constant). At moderate distances, greater than $r = r_m$, $U_{H\cdots B}(r)$ becomes negligible; hence U(r), *i.e.*, S(r), becomes the known dipole–dipole interaction potential.[2] We will compute S(r) as indicated above, and obtain $U_{H\cdots B}(r)$ empirically so that U(r) will satisfy conditions a–c.

Conditions a and b enable us to use two adjustable parameters for computing $U_{H\cdots B}(r)$. Since there is no unique way of meeting these two conditions, we explore three possible empirical functions (where the parameters to be adjusted are starred).

$$U_{H\cdots B}(r) = a^* e^{-b^* r} - c/r^6 \tag{11a}$$

$$U_{H\cdots B}(r) = f^*/r^{\alpha^*} - c/r^6 \tag{11b}$$

$$U_{H\cdots B}(r) = d^*/r^{12} - c^*/r^6 \tag{11c}$$

In eq 11a,b we assume c to be given (Scott and Scheraga, 1965), and use both adjustable parameters to construct the repulsive part of the potential. In eq 11c, we assume a potential of the "6–12" form, and use one adjustable parameter to describe the strength of the repulsion and the other to describe the strength of the attraction; in all cases, the latter decreases rapidly to zero for $r > r_m$, and hence the potential goes over into the dipole–dipole form. The starred parameters are determined by applying conditions a and b, *i.e.*

$$U(r_m) = -\epsilon$$
$$\left(\frac{\partial U}{\partial r}\right)_{r_m} = 0 \tag{12}$$

giving

$$a^* = [-\epsilon + c/r_m^6 - S(r_m)]e^{b^* r_m}$$
$$b^* = \left[6c/r_m^7 + \left(\frac{\partial S}{\partial r}\right)_{r_m}\right] \Big/ [c/r_m^6 - S(r_m) - \epsilon] \tag{13a}$$

$$\alpha^* = r_m b^*$$
$$f^* = [-\epsilon + c/r_m^6 - S(r_m)]r_m^{\alpha^*} \tag{13b}$$

$$c^* = r_m^6\left[2\epsilon + 2S(r_m) + \frac{r_m}{6}\left(\frac{\partial S}{\partial r}\right)_{r_m}\right]$$
$$d^* = r_m^6 c^* - r_m^{12}[\epsilon + S(r_m)] \tag{13c}$$

$S(r_m)$ and $(\partial S/\partial r)_{r_m}$ must be determined numerically.

In using the above equations to obtain $U_{H\cdots B}(r)$ empirically, it is assumed that the AH···B part of the system remains linear as r varies. Once $U_{H\cdots B}(r)$ is evaluated in this fashion, this same value is assumed to hold for all (nonlinear) orientations of AH···B (for a given type hydrogen bond, *e.g.*, amide hydrogen bond); the departure from linearity reflects itself in altered values of the function S(r), which must be evaluated separately for every conformation every time a conformational calculation involving a hydrogen bond is made. In summary, the calculations reported here provide the values of $U_{H\cdots B}(r)$ for various types of hydrogen bonds.

Values of $U_{H\cdots B}(r)$ for Several Types of Hydrogen Bond. Consider first the gas-phase carboxylic acid dimers, for which experimental data are available. For formic acid (Pimental and McClellan, 1960), ϵ = 14 kcal/mole, $r_{O\cdots O}$ = 2.72 A; assuming that r_{OH} = 1.00 A, we obtain $r_m = r_{H\cdots O}$ = 1.72 A. Using the partial charges computed in this paper, $S(r_m)$ = −8.0 kcal/mole and $(\partial S/\partial r)_{r_m}$ = 9.8 kcal/mole A; $c_{H\cdots O}$ is computed as a geometrical mean from data given by Scott and Scheraga (1965) as 135 (kcal/mole) A^6. Inserting these values in eq 13a, $a^* = 2.89 \times 10^8$ kcal/mole, $b^* = 10.45$ A^{-1}, and (from eq 13b) $\alpha^* = 17.97$. Thus, $U_{H\cdots B}(r)$ for an OH···O hydrogen bond is given by eq 11a with these parameters. If eq 11c is used, the values of the parameters are d^* = 5620 (kcal/mole) A^{12} and c^* = 374 (kcal/mole) A^6.

Consider next the amide hydrogen bond, which is of paramount importance for polypeptide conformational calculations. Unfortunately, the association of the amides has not been studied in the gas phase, presumably because of the low vapor pressure of these compounds. Hence, we must argue by analogy in order to obtain a reasonable value of ϵ. Many studies have been carried out on this association in solution (*e.g.*, in CCl_4 and in benzene) and give a rather consistent value of $\epsilon = 3.5 \pm 0.5$ kcal/mole for a great variety of amides (Pimentel and McClellan, 1960). Now, the effect of solvent on hydrogen bonds is not well understood; the alcohols have the same energy of association in the gas phase as in CCl_4 (about 4 kcal/mole), but the energy of association of carboxylic acids is lowered (Pimentel and McClellan, 1960) by about a factor of two when measured in CCl_4. Since unsubstituted amides form double hydrogen-bonded dimers and have an extensive conjugated structure (Pimentel and McClellan, 1960), just like the carboxylic acids, we feel that the amide hydrogen bond is very much like the carboxylic acid case; this is in contrast to *N*-methylamides, which can form only a single hydrogen-bonded dimer because the amide geometry is invariably *trans*. Sublimation energies (Pimentel and McClellan, 1960) indicate that the amides have a strong hydrogen bond, of approximately the same strength as carboxylic acids. From the data in Table VII, comparing amides with carboxylic acids, we are led to the conclusion that $\epsilon \simeq 7$ kcal/mole is a reasonable guess for the gas-phase energy of the amide hydrogen bond; crystal energy calculations, in progress,

[2] For nonbonded interactions, in situations *not* involving hydrogen bonds, the same form of S(r) is applicable at large internuclear distances (the attractive part). At small internuclear distances, the repulsion arises both from electrostatic repulsion and the repulsive part of the nonbonded 6–12 potential.

TABLE VII: Comparison of Amide and Carboxylic Acid Hydrogen Bonds (Data from Pimentel and McClellan, 1960).

	Energy of Hydrogen Bond (kcal/mole)		
	Gas Phase	CCl_4	Crystal[a]
Carboxylic acid	~7	~3–4	~7
Amide	?	~3.5	~7–8

[a] Based on sublimation energy.

should yield more definitive results than that obtained from a consideration of the sublimation energies. Another reason for assuming that $\epsilon > 3.5$ kcal/mole is that the electrostatic potential (with the charges computed here) already gives −3.5 kcal/mole at a distance equal to the "normal" van der Waals' radii (~2.8 A). If this value also holds at r_m ~1.85 A, then the potential would have to be flat for about 1 A. In fact, if an adjustable repulsive potential of the form ae^{-br} is used, one cannot produce a minimum at r_m for $\epsilon = 3.5$ kcal/mole, but instead one obtains an inflection because the total potential is so flat.

Using *N*-methylacetamide as a model and using charges as calculated in this paper, we obtain $S(r_m) = -5.61$ kcal/mole and $(\partial S/\partial r)_{r_m} = 5.25$ kcal/mole A, taking $r_m = 1.85$ A (from $r_{N\cdots O} = 2.85$ A, and $r_{NH} = 1.00$ A). Since we do not know ϵ, we have computed the adjustable parameters for several values of ϵ (see Table VIII). Figure 5 shows the amide hydrogen-bond potential, using both eq 11a and 11c for $\epsilon = 5.5$ kcal/mole; it is seen that there is a negligible difference between these two potential functions.

It should be noted from Table VIII that, for $\epsilon = 5.5$ kcal/mole, the repulsive parameters are $a^* = 1.90 \times 10^4$ kcal/mole and $b^* = 4.65$ A^{-1}, while crude

TABLE VIII: Parameters for the Amide Hydrogen Bond.[a]

ϵ (kcal/mole)	a^* (kcal/mole)	b^* (A^{-1})
5	7.36×10^3	4.06
5.5	1.90×10^4	4.65
6	6.91×10^4	5.43
7	7.35×10^6	8.17
	c^* (kcal/mole) A^6	d^* (kcal/mole) A^{12}
5.5	56.08	2423

[a] The values of a^* and b^* were computed with the value of $c = 135$ (kcal/mole) A^6 (Scott and Scheraga, 1965).

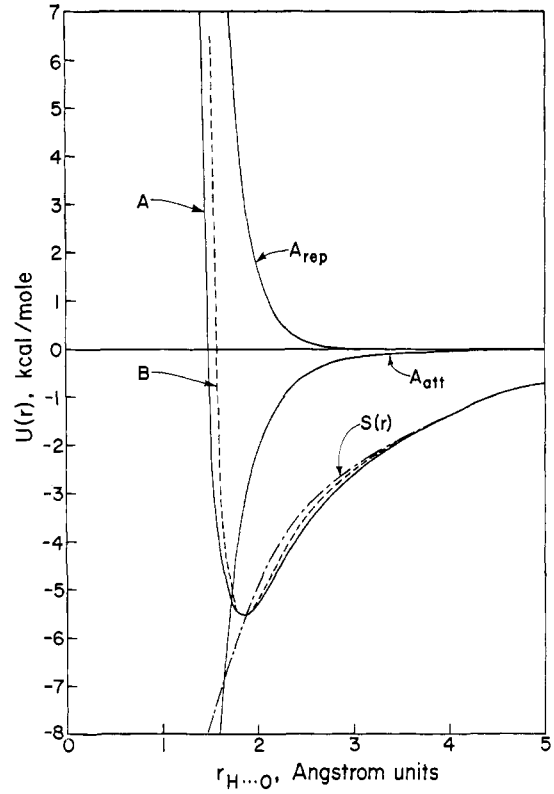

FIGURE 5: Potential function for the amide hydrogen bond, with $\epsilon = 5.5$ kcal/mole. Curve A is the function $U(r)$, being the sum of $S(r)$, $a^*e^{-b^*r}$ (curve A_{rep}), and c/r^6 (curve A_{att}), according to eq 10 and 11a; curve B is the function $U(r)$, being the sum of $S(r)$, d^*/r^{12}, and c^*/r^6, according to eq 10 and 11c.

estimates for NH···O=C nonbonding interactions (obtained as in the paper by Scott and Scheraga, 1965) give $a = 2.37 \times 10^4$ kcal/mole and $b = 4.3$ A^{-1}. Obviously, within the accuracy of our estimates, we could not tell which of these were the more appropriate values, *i.e.*, the "normal" values for the repulsive potential enable a strong short hydrogen bond to form. On the other hand, the values for formic acid ($a^* = 2.89 \times 10^8$ kcal/mole and $b^* = 10.45$ A^{-1}) are quite different from the nonbonding parameters ($a = 2.37 \times 10^4$ kcal/mole and $b = 4.3$ A^{-1}). If the amide hydrogen bond is closely analogous to the carboxylic acid hydrogen bond, as suggested in Table VII, then perhaps ϵ for the amide hydrogen bond should be 7 kcal/mole, corresponding to $a^* \simeq 10^7$ kcal/mole and $b^* \simeq 8$ A^{-1}. For the present, eq 11c (leading to curve B of Figure 5) provides an adequate estimate of $U_{H\cdots B}(r)$ for the amide hydrogen bond; crystal structure calculations, in progress, should provide a refined set of parameters for the specification of $U_{H\cdots B}(r)$.

It is obvious from Figure 5 that the electrostatic potential $S(r)$ is a very important part of $U(r)$. In this connection, there recently has been much discussion

(Sasisekharan, 1959; Ramachandran *et al.*, 1967; Krimm *et al.*, 1967) about the possibility of CH···O hydrogen bonds in polyproline, collagen, and polyglycine. In this case, the electrostatic potential is very small because of the very small charge on the CH group. Hence, the reasonable empirical approach would be to soften the repulsive potential. This would allow a closer approach of these groups than would be expected from a consideration of van der Waals' radii. Such a condition is apparently operative in the carboxylic acid hydrogen bond and probably also in the amide hydrogen bond. We do not know ϵ for the CH···O hydrogen bond but only an estimate (Sasisekharan, 1959; Ramachandran *et al.*, 1967; Krimm *et al.*, 1967) of the closest distance of approach, *viz.*, $r_{C\cdots O}$ = 3.1–3.2 A, which is noticeably smaller than the expected value of 3.8 A (Scott and Scheraga, 1965). For this case, we can use the known c value for the dispersion forces, and treat the repulsive parameters as adjustable. Because we have only one condition (*viz.*, we "know" only the value of r_m), we must use a one-parameter potential for $U_{H\cdots B}(r)$, and we choose the "6–12" form for simplicity, and omit $S(r)$. With c = 135 (kcal/mole) A^6, r_{CH} = 1.00 A (corresponding to $r_m = r_{H\cdots O}$ = 2.05 A), we obtain

$$d^* = cr_m^6/2 = 4900 \text{ (kcal/mole) A}^{12} \qquad (14)$$

as a guess. The energy becomes

$$\epsilon = (1/2)(c/r_m^6) = 0.93 \text{ kcal/mole} \qquad (15)$$

Thus, this is a weak hydrogen bond since no electrostatic contribution was included in this calculation.

In summary, we have provided a procedure to compute $U_{H\cdots B}(r)$, an empirical part of a hydrogen-bond potential. This function may be used in combination with nonbonded and electrostatic interaction energies in calculations of polypeptide conformation. The method used here, and the results obtained therefrom, seem reasonable. However, it must be emphasized that the hydrogen-bond potential is only as reliable as the assumptions made in its computation. It is hoped that crystal structure calculations, now in progress, will provide a more refined potential function of the form presented here.

References

Berthod, H., and Pullman, A. (1965), *J. Chim. Phys. 62*, 942.

Bradley, D. F., Lifson, S., and Honig, B. (1963), Electronic Aspects of Biochemistry, New York, N. Y., Academic, p 77.

Curl, R. F., Jr. (1959), *J. Chem. Phys. 30*, 1529.

Del Re, G. (1958), *J. Chem. Soc.*, 4031.

Del Re, G. (1963a), *Theoret. Chim. Acta 1*, 188.

Del Re, G. (1963b), Electronic Aspects of Biochemistry, New York, N. Y., Academic, p 221.

Krimm, S., Kuroiwa, K., and Rebane, T. (1967), *in* Conformation of Biopolymers, Vol. 2, Ramachandran, G. N., Ed., New York, N. Y., Academic, p 439.

Kurland, R. J., and Wilson, E. B. (1957), *J. Chem. Phys. 27*, 585.

Pimentel, G. C., and McClellan, A. L. (1960), The Hydrogen Bond, San Francisco, Calif., W. H. Freeman.

Pullman, A., and Pullman, B. (1963), Quantum Biochemistry, New York, N. Y., Interscience, pp 321 and 661.

Ramachandran, G. N., Ramakrishnan, G., and Venkatachalam, C. M. (1967), *in* Conformation of Biopolymers, Vol. 2, Ramachandran, G. N., Ed., New York, N. Y., Academic, p 429.

Sasisekharan, V. (1959), *Acta Cryst. 12*, 897.

Scheraga, H. A., Scott, R. A., Vanderkooi, G., Leach, S. J., Gibson, K. D., Ooi, T., and Némethy, G. (1967), *in* Conformation of Biopolymers, Vol. 1, Ramachandran, G. N., Ed., New York, N. Y., Academic, p 43.

Scott, R. A., and Scheraga, H. A. (1965), *J. Chem. Phys. 42*, 2209.