JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

Registered in U.S. Patent Office.

VOLUME 99, NUMBER 17 AUGUST 17, 1977

Analytical Potentials from "ab Initio" Computations for the Interaction between Biomolecules. 1. Water with Amino Acids

E. Clementi,* F. Cavallone, and R. Scordamaglia

Contribution from the Società Montedison, Istituto Ricerche "G. Donegani", 28100 Novara, Italy. *Received May 14, 1976*

Abstract: The general problem of the determination of the structure of water molecules surrounding biological molecules has been investigated with a method capable of providing an adequate simulation of liquid water (at room temperature) and ion pairs in aqueous solution (at room temperature). SCF-LCAO-MO computations are presented here for 21 amino acids interacting with one molecule of water located at different positions and orientations around each amino acid. The 1690 computed total energies, E, for the water-amino acid complex have been fitted with an analytical potential of the form $E = \Sigma_i \Sigma_{j \neq i} (-A_{ij}{}^{ab}/r_{ij}{}^6 + B_{ij}{}^{ab}/r_{ij}{}^{12} + C_{ij}{}^{ab}/r_{ij}) + E(\text{amino acid}) + E(\text{water})$, where i and j designate two atoms, one in the amino acid and the second in the water molecule, a is an index that distinguishes the electronic environment of an atom in the amino acid, b is an index that distinguishes between either a hydrogen or the oxygen atom in the water molecule, A, B, and C are fitting constants, E(amino acid) and E(water) are the total energies of the amino acid and of the water molecule, respectively, at infinite separation from the other. The standard deviation of the fit (comparing the 1690 SCF-LCAO-MO energies with those obtained by the above analytical potential) is 0.6 kcal/mol.

It is well known that the amino acids constitute essential building blocks of many macromolecules of biological interest. It is equally well known that most biological molecules interact in aqueous solution at temperatures distributed in a relatively narrow range centered approximately around room temperature. A problem of interest in this context is the determination of water's structural organization around macromolecules and the complementary problem, namely the determination of the most stable configurations of macromolecules, when in solution at room temperature.

In a number of papers (see below) we have started some preliminary study aimed at solving some aspects of the above questions. Since we have selected to use as research tools those techniques broadly designated as numerical simulations, preliminary tests have been performed in order to check the validity of the approach. Our techniques can be summarized into three consecutive steps. *First* we compute the interactions of a selected molecule, M, with a molecule of water placed at a sufficiently large number of positions and orientations relative to M, as to give a reasonable sampling of the potential surface representing the interactions of the "M-water" complex. During the first step, M is considered as rigid, namely, we study only a single conformation of M interacting with water. The interaction can be computed in different ab initio approximations, either in the self-consistent-field molecular orbital (SCF-MO) approximation, with a relatively small basis set, or in the SCF-MO approximation with a sufficiently large basis set as to nearly reach the Hartree-Fock limit,[1] or in some more refined model (like using perturbation[2] or configuration interaction, CI, techniques[3]) introduced to take into account the quantum mechanical many-body correction (correlation energy correction, for short). The decision on the specific ab initio technique to be chosen among the above ones depends mainly on the accuracy one wishes to reach. In the *second step*, we fit the interaction energy of M with water with some relatively simple analytical expression; there are many ways to perform a fitting, either using well-known models, like the Lennard-Jones potentials, or some electrostatic potentials or some series expansions in terms of the distances between the atoms of M and those of the water molecule. In this step one obtains an analytical potential that reproduces the computed interaction energies. For "low resolution" type work, a Lennard-Jones potential (possibly with some additional term) has proven to be adequate;[4] clearly, this is not the case for "high resolution" work especially if one wishes to accurately reproduce interactions that have more than one minimum along one of the axes defining the M-H_2O conformational space.[5] The *third step* makes use of the analytical potentials previously obtained and determines the structural organization of many molecules of water (representing the solvent) around M for some selected temperature, using Monte Carlo techniques.[6]

Clearly, there are many alternative ways to reach the same goal, namely the structural organization of water as solvent, around M. For example, one could obtain the structural organization of water molecules around M by *direct* experimentation. To our knowledge, however, such an approach can be brought to its conclusion in a shorter time and in a more definite manner, if theoretical data, like those reported here,

Table I. Geometry,[a] Charges, MO Valency State for Glycine[b]

Atom	Class	x	y	z	Charge	MOVS
O(2)	9	2.835	2.101	1.389	−0.50	0.56
O(1)	10	−1.320	3.200	1.631	−0.38	0.40
N	11	2.700	2.032	−3.749	−0.55	0.92
C′	5	0.673	2.824	0.416	0.46	1.41
C(A)	8	0.580	3.276	−2.431	−0.33	1.07
H(1)	1	2.597	2.270	−5.723	0.24	0.28
H(2)	1	4.433	2.698	−3.125	0.25	0.28
H(A1)	2	0.689	5.303	−2.786	0.18	0.27
H(A2)	2	−1.194	2.587	−3.213	0.22	0.30
H(O2)	4	2.607	1.920	3.160	0.40	0.32

[a] The geometry has been taken from P.-G. Jonsson and A. Kvick, *Acta Crystallogr., Sect. B,* **28,** 1827 (1972). [b] Total energy −281.779 au.

Table II. Geometry,[a] Charges, MO Valency State for Alanine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	8.302	1.971	6.866	−0.42	0.38
O(2)	9	5.076	4.360	8.391	−0.55	0.52
N	11	7.473	3.218	2.030	−0.55	0.91
C′	5	6.389	3.299	6.575	0.50	1.36
C(A)	8	5.434	3.765	3.884	−0.13	1.11
C(B)	6	3.131	2.137	3.306	−0.59	1.01
H(A)	2	4.967	5.767	3.690	0.21	0.28
H(1)	1	8.101	1.384	2.190	0.26	0.29
H(2)	1	8.939	4.472	2.307	0.25	0.28
H(B1)	3	3.635	0.157	3.342	0.20	0.28
H(B2)	3	1.663	2.449	4.695	0.20	0.29
H(B3)	3	2.394	2.579	1.453	0.22	0.29
H(O2)	4	5.851	3.929	9.952	0.41	0.32

[a] The geometry has been taken from M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *J. Am. Chem. Soc.,* **94,** 2657 (1972). [b] Total energy −320.661 au.

would be available. In addition, the experimental difficulties in obtaining the structural organization of water around M either by x-ray or neutron diffraction (at room temperature or nearby) are nearly as remarkable as those one encounters in solving the previously explained three steps. Alternatively, in the field of numerical simulations, one can construct the interaction pair potentials either making use of experimental data[6] or semiempirical methods.[7] The choice among such alternatives probably depends on natural inclination and personal experience. Personally, we favor the use of ab initio computations over the use of experimental data (in the phase that leads to the construction of an analytical potential), since it is rather difficult to select, and often even find (especially in the repulsive region of the potential), a sufficient sample of experiments of direct relevance and equal reliability to infer the potential functions, at least in the case of biomolecules in solution.

Before attempting to consider biological molecules, we felt it was advisable *to test the feasibility of our approach* on a number of simpler cases that can be compared with accurate experimental data. Thus, we have considered as M: (1) the molecule of water; (2) separated ions; and (3) ion pairs. The study of the structural organization of water around water, i.e., the study of the structure of liquid water, has been presented starting with a Hartree-Fock interaction potential,[8] with a Hartree-Fock potential corrected by an empirical inclusion of dispersion forces,[9] with a Hartree-Fock potential corrected by a perturbation expansion[2] and by an extensive configuration interaction potential.[3] For each potential we have compared the computed and the experimentally derived pair correlation functions,[10] $g_{\mathrm{O\text{-}O}}$, $g_{\mathrm{O\text{-}H}}$, and $g_{\mathrm{H\text{-}H}}$, and, more recently, the computed and observed x-ray and neutron diffraction intensities.[11] The comparison nicely indicates, as expected, that the more accurate the potential (the higher the "resolution" of our technique), the nearer the agreement; for the case of a CI potential, the agreement is nearly within experimental error. An additional test of the quality of the water-water potential has been a comparison of computed and experimental second virial coefficients; again, we conclude that such a potential is likely to be the best one currently available in literature and a quite reliable one for the task we have undertaken.[12]

When M is a single ion, F^- or Cl^-,[13] Li^+,[14] Na^+, or K^+,[15] or an ion pair,[16] the water organization around it and other quantities like the coordination numbers, the radii of the first and second solvation shell, have been computed and compared to experimental values. Again, the agreement is satisfactory. (We note that we have considered the solvation of ions mainly because the solvent of interest for biological molecules is often not pure water, but water with ions.)

In this paper we report on the interaction of a number of amino acids with water; these results have been reported at several meetings[17] and full details of this work are available in seven technical reports.[18] In Tables I-XXI, we report the x, y, and z coordinates of the atoms constituting the 21 amino acids analyzed in this work;[18] additional data on the amino acids are given in the Tables I-XXI and the corresponding explanation is postponed till later in this paper. The geometry (and other characteristics) of the amino acids are illustrated

Table III. Geometry,[a] Charges, MO Valency State for Valine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	15.972	10.125	8.884	−0.40	0.38
O(2)	9	14.534	6.355	7.448	−0.55	0.52
N	11	11.462	12.281	9.221	−0.50	0.91
C′	5	14.249	8.734	8.085	0.50	1.38
C(A)	8	11.653	9.897	7.716	−0.13	1.08
C(B)	7	10.963	10.369	4.916	−0.17	1.06
C(G2)	6	9.956	7.969	3.640	−0.60	1.99
C(G1)	6	13.289	11.478	3.454	−0.57	1.01
H(1)	1	11.952	11.900	11.165	0.24	0.28
H(2)	1	9.552	12.994	9.129	0.23	0.27
H(G11)	3	12.779	11.796	1.503	0.19	0.29
H(G12)	3	14.842	10.155	3.544	0.18	0.28
H(G13)	3	13.851	13.245	4.308	0.21	0.29
H(G21)	3	10.225	6.690	2.072	0.21	0.30
H(G22)	3	8.091	8.795	3.531	0.19	0.26
H(G23)	3	10.144	6.943	5.394	0.19	0.26
H(B)	2	9.477	11.768	4.870	0.19	0.26
H(A)	2	10.282	8.587	8.471	0.17	0.26
H(O2)	4	16.257	5.920	7.708	0.42	0.32

[a] The geometry has been taken from K. Torii and Y. Iitaka, *Acta Crystallogr., Sect. B*, **26**, 1317 (1970). [b] Total energy −398.355 au.

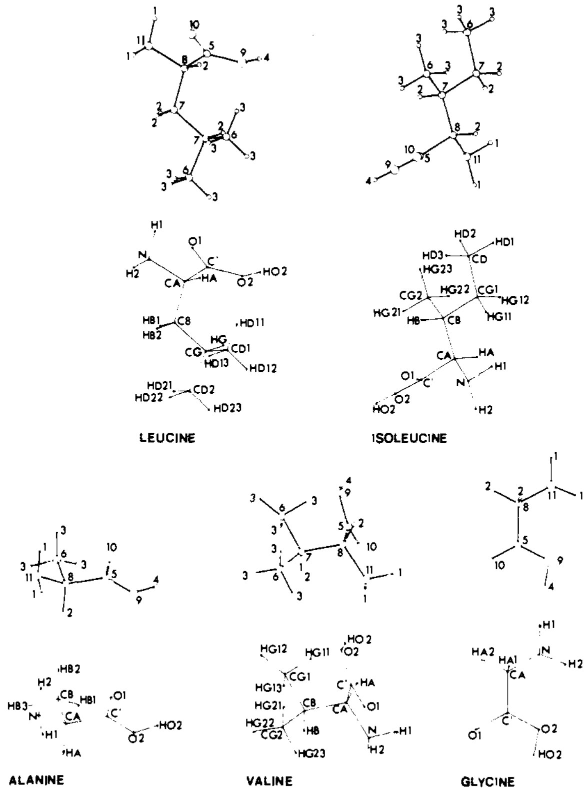


Figure 1. Geometry and identification code for the atoms. Each amino acid is described two times. The first time, the atom is identified with the same alphanumerical identification code adopted in the Tables I-V reporting the molecular geometry. The second time, each atom is identified by a numerical code that designates the class to which the atom belongs; such class code is given, in addition, in the Tables I-V reporting the molecular geometry. See ref 18a for additional information.

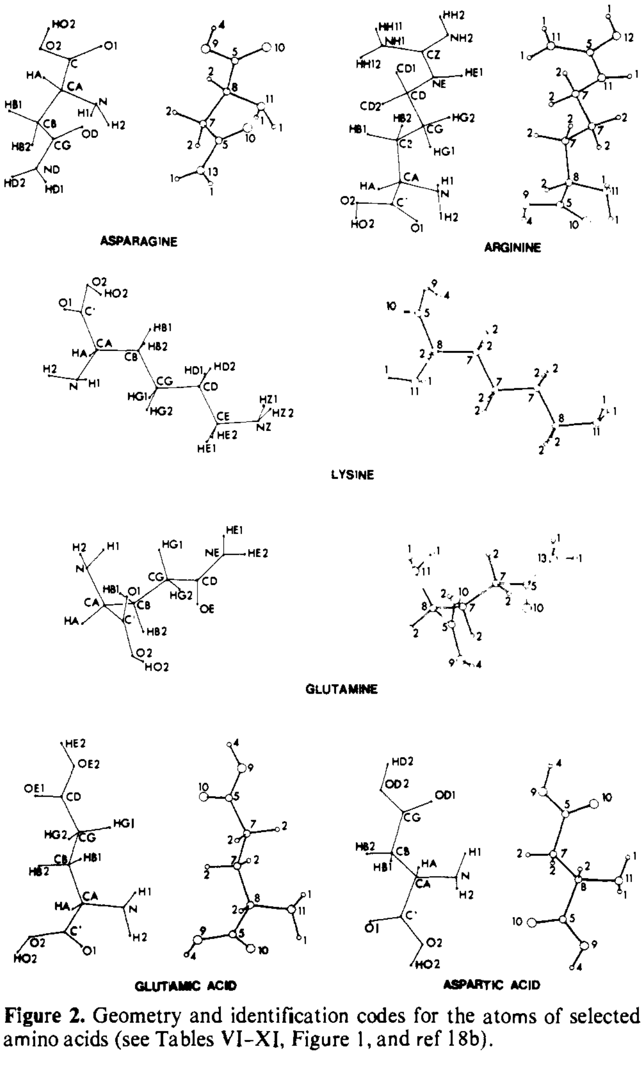


Figure 2. Geometry and identification codes for the atoms of selected amino acids (see Tables VI-XI, Figure 1, and ref 18b).

in Figures 1-5.

Results and Discussion

Computation of the Interaction Energy. The computations we shall discuss below are not as accurate as those performed for the water-water interaction,[2,3] the reason being that for the case of water we aimed our work at a duplication and verification, via simulation, of accurate x-ray and neutron

Table IV. Geometry,[a] Charges, MO Valency State for Leucine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	10.418	5.759	−1.406	−0.41	0.38
O(2)	9	13.771	7.422	0.510	−0.53	0.57
N	11	7.538	9.774	−0.638	−0.57	0.90
C′	5	11.430	7.392	−0.101	0.53	1.44
C(A)	8	9.911	9.583	0.872	−0.17	1.05
C(B)	7	9.175	9.376	3.666	−0.47	0.94
C(G)	7	11.308	9.443	5.589	−0.21	1.05
C(D1)	6	10.187	9.039	8.217	−0.69	0.92
C(D2)	6	12.710	11.956	5.490	−0.68	0.92
H(A)	2	10.905	11.090	0.648	0.21	0.27
H(B1)	2	8.044	10.761	4.042	0.23	0.29
H(B2)	2	8.291	7.782	3.896	0.25	0.29
H(G)	2	12.497	8.106	5.222	0.23	0.29
H(D11)	3	9.240	7.417	8.277	0.23	0.30
H(D12)	3	11.526	9.118	9.524	0.23	0.30
H(D13)	3	8.911	10.427	8.654	0.22	0.29
H(D21)	3	13.397	12.269	3.766	0.23	0.29
H(D22)	3	14.059	12.019	6.794	0.23	0.30
H(D23)	3	11.524	13.445	5.870	0.22	0.29
H(1)	1	6.493	11.213	−0.041	0.26	0.29
H(2)	1	7.929	9.913	−2.473	0.27	0.29
H(O2)	4	14.950	5.820	0.255	0.39	0.33

[a] The geometry has been taken from L. Golic and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **28,** 1265 (1972). [b] Total energy −437.030 au.

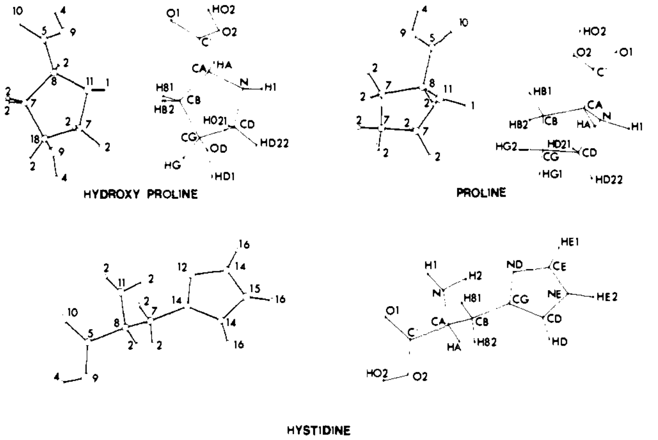


Figure 3. Geometry and identification codes for the atoms of selected amino acids (see Tables XII-XIV, Figure 1, and ref 18c).

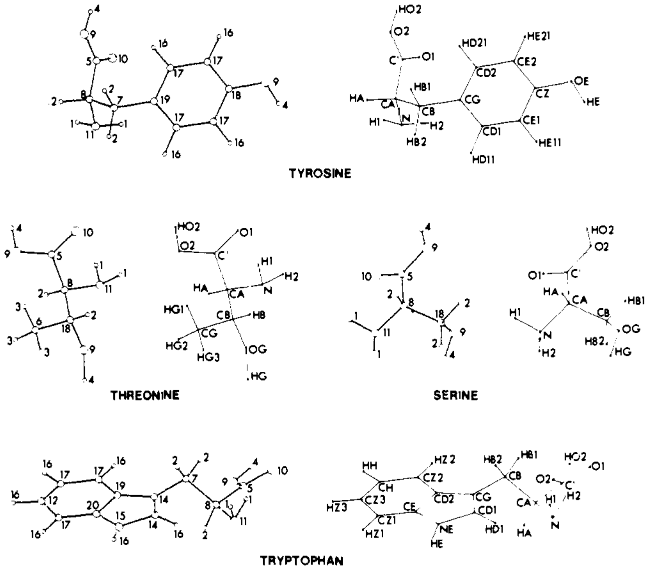


Figure 4. Geometry and identification codes for the atoms of selected amino acids (see Tables XV-XVIII, Figure 1, and ref 18d).

diffraction experiments.[19] For the case of macromolecules in water, the situation is qualitatively very different; indeed, only macromolecules in crystals and not in solution can now be tackled experimentally by diffraction techniques. Our computations are likely the needed ground work to obtain structural determinations with about 0.3 Å resolution. In the *first step,* described above, the accuracy of the SCF-LCAO-MO computations is likely sufficient to yield structural data with resolution in the range 0.20-0.15 Å; however, the errors introduced in *step two* and later in *step three* will decrease our reliability to the above given limit. Increased efficiency in the computer program in *step three* (by a factor of 5 to 10) could, however, decrease our errors that are mainly due to the practical impossibility to carry out the Monte Carlo computations to the same level of accuracy as for the case of water-water and water-ions interaction. As stated above, the interactions for the five amino acids and one molecule of water are computed in the SCF-LCAO-MO approximation. The molecular orbitals are expanded in a linear combination of contracted Gaussian functions, each one centered at a nucleus of the system;[20] in turn, the contracted Gaussian functions are a prefabricated (on the basis of atomic computations) linear combination of primitive Gaussian functions.[20] The primitive Gaussian functions are specified by giving the orbital exponent of the radial part as well as the angular part (type); the type, the orbital exponent, and the contraction coefficients for each contracted function are given in ref 18.

The molecule, M, is considered fixed in space (with coordinates given in Tables I-XXI) and the water molecule is

Table V. Geometry,[a] Charges, MO Valency State for Isoleucine[b]

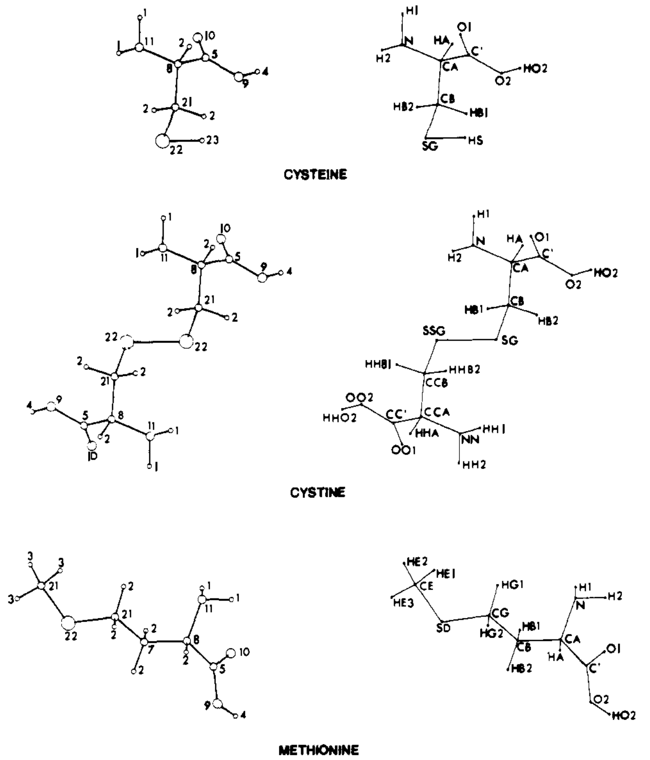
Atom	Class	x	y	z	Charge	MOVS
O(2)	9	11.414	5.430	9.769	−0.54	0.53
O(1)	10	12.923	9.392	8.907	−0.41	0.37
N	11	16.254	3.561	8.916	−0.50	0.91
C′	5	13.103	6.999	8.857	0.50	1.38
C(A)	8	15.369	5.839	7.445	−0.12	1.10
C(B)	7	14.634	5.076	4.741	−0.18	1.08
C(G1)	7	16.932	3.888	3.294	−0.37	1.06
C(D)	6	16.344	3.200	0.539	−0.58	1.02
C(G2)	6	13.594	7.367	3.315	−0.58	1.00
H(1)	1	16.764	4.129	10.809	0.23	0.27
H(2)	1	17.880	2.744	7.990	0.23	0.27
H(D1)	3	18.006	2.397	−0.332	0.20	0.29
H(D2)	3	15.788	4.880	−0.478	0.19	0.28
H(D3)	3	14.824	1.840	0.487	0.20	0.29
H(G21)	3	13.083	6.813	1.419	0.19	0.28
H(G22)	3	15.020	8.825	3.237	0.19	0.28
H(G23)	3	11.944	8.081	4.284	0.20	0.28
H(G11)	2	17.511	2.195	4.276	0.18	0.27
H(G12)	2	18.472	5.227	3.314	0.18	0.27
H(B)	2	13.156	3.672	4.854	0.20	0.28
H(A)	2	16.879	7.210	7.363	0.18	0.27
H(O2)	4	10.071	6.405	10.453	0.42	0.32

[a] The geometry has been taken from K. Torii and Y. Iitaka, *Acta Crystallogr., Sect. B,* **27,** 2237 (1971). [b] Total energy −437.293 au.

placed at different distances around M. For a given position of the oxygen atom of H_2O, one or more orientations for the water molecule are considered ("orientations" in this context indicate, for a given position of the oxygen of H_2O, the hydrogen atoms of H_2O might point toward, away, or otherwise relative to M). For each position and orientation of the water molecule we have computed the total energy of the system M + H_2O designated as E(M,W); by subtracting from E(M,W) the energy of the water molecule, E(W), at infinite distance from M and the energy of the molecule M, E(M), at infinite distance from the water molecule, we obtain the interaction energy, I(M,W) between M and water, i.e., $I(\mathrm{M,W}) = E(\mathrm{M,W}) - E(\mathrm{M}) - E(\mathrm{W})$. The values of E(M) for the amino acids considered in this paper are given (in atomic units) at the bottom of Tables I-XXI. In such tables the last two columns report the computed net charge, NCH, on each atom, as well as the value of the molecular orbital valency state, MOVS. The first quantity, NCH, is obtained from the electron population analysis, proposed by Mulliken.[21] It represents the gain or the loss in the electronic population of an atom, when in the molecule M, relative to its population when isolated; therefore, the NCH constitutes an index of the electron density rearrangement that follows the formation of the molecule from separated atoms. The second quantity, MOVS, proposed by Clementi,[22] represents the energy of an atom when in the molecule M, relative to its energy when in the atomic ground state.

The MOVS constitute an index of the energetic variations that follow the formation of a molecule from separated atoms. Both quantities are further constrained by their definition within the SCF-LCAO-MO framework and therefore are, to some extent, basis set dependent. Both quantities are related to the hybridization of the atoms in M,[23] and can be computed in a fast, simple way once that an SCF-LCAO-MO solution has been obtained (in the next section of this paper we shall describe in detail the need to compute the NCH and the MOVS).

For the water molecule, we have taken a standard geometry and the basis set chosen yields a value of $E(\mathrm{W}) = 75.733\,058$ au.


CYSTEINE

CYSTINE

METHIONINE

Figure 5. Geometry and identification codes for the atoms of selected amino acids (see Tables XIX-XXI, Figure 1, and ref 18e).

Analytical Expression for the Interaction Energy. As stated in the first section of this paper, the fitting of the computed interaction energies is intended to provide the interaction potentials needed as input to the Monte Carlo simulation.

In the Monte Carlo procedure, many configurations (typically 0.5 × 10^6 to 2.0 × 10^6) of the water molecules (corre-

Table VI. Geometry,[a] Charges, MO Valency State for Arginine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	7.091	9.353	21.244	−0.41	0.38
O(2)	9	7.398	13.432	20.046	−0.55	0.52
N	11	11.568	7.995	19.198	−0.58	0.87
N(E)	11	4.744	8.355	11.515	−0.45	1.06
N(H1)	11	2.102	11.819	9.426	−0.61	1.04
N(H2)	12	1.290	7.388	8.700	−0.50	0.82
C′	5	8.182	11.079	20.096	0.51	1.38
C(A)	8	10.564	10.521	18.587	−0.12	1.10
C(B)	7	9.997	10.757	15.761	−0.37	1.04
C(G)	7	7.939	8.995	14.792	−0.41	1.02
C(D)	7	6.570	10.108	12.596	−0.28	1.09
C(Z)	5	2.701	9.143	9.823	0.34	1.43
H(A)	2	11.926	12.000	19.090	0.18	0.27
H(1)	1	11.454	7.718	21.081	0.27	0.28
H(2)	1	13.402	7.978	18.810	0.26	0.28
H(B1)	2	9.539	12.727	15.397	0.19	0.27
H(B2)	2	11.743	10.425	14.772	0.20	0.28
H(G1)	2	8.762	7.230	14.212	0.21	0.28
H(G2)	2	6.654	8.506	16.208	0.22	0.29
H(D1)	2	5.567	11.728	13.305	0.19	0.27
H(D2)	2	7.808	10.841	11.226	0.19	0.27
H(E)	1	5.215	6.481	11.429	0.31	0.31
H(H2)	1	3.111	13.239	10.196	0.28	0.31
H(H12)	1	0.608	12.290	8.297	0.28	0.31
H(H11)	1	−0.023	7.950	7.423	0.22	0.29
H(O2)	4	5.896	13.529	21.024	0.42	0.32

[a] The geometry has been taken from M. S. Lehmann, J. J. Verbist, W. C. Hamilton, and T. F. Koetzle, *J. Chem. Soc., Perkin Trans,* **2,** 133 (1973). [b] Total energy −600.620 au.

Table VII. Geometry,[a] Charges, MO Valency State for Asparagine[b]

Atom	Class	x	y	z	Charge	MOVS
O(2)	9	−0.027	0.889	0.007	−0.56	0.51
O(1)	10	−1.257	−3.223	−0.149	−0.39	0.38
O(D)	10	1.208	−1.967	5.312	−0.42	0.36
N(D)	13	1.128	1.187	8.217	−0.63	1.02
N	11	−4.388	−2.893	3.830	−0.54	0.92
C′	5	−1.313	−1.075	0.802	0.52	1.42
C(A)	8	−3.221	−0.540	2.889	−0.11	1.20
C(G)	5	0.174	−0.035	6.190	0.41	1.39
C(B)	7	−2.219	1.042	5.071	−0.43	0.99
H(1)	1	−3.277	−3.746	5.212	0.26	0.28
H(1)	1	−6.172	−2.540	4.576	0.24	0.28
H(D1)	1	2.643	0.512	9.079	0.33	0.32
H(D2)	1	0.301	2.785	8.801	0.31	0.31
H(A)	2	−4.741	0.643	2.055	0.20	0.28
H(B1)	2	−1.767	2.959	4.327	0.21	0.28
H(B2)	2	−3.571	1.391	6.538	0.20	0.28
H(O2)	4	0.942	0.372	−1.413	0.41	0.32

[a] The geometry has been taken from M. Ramanadham, S. K. Kikka, and R. Chidambaram, *Acta Crystallogr., Sect. B,* **28,** 3000 (1972). [b] Total energy −487.852 au.

sponding to different positions and/or orientations) are randomly chosen and weighted, using the Boltzmann factor; therefore, it is essential to use a *simple* and *computationally fast* analytical form for the potential, since for each configuration it is required to recompute the quantity E(M,W), in addition to the water–water interaction.[24] On the base of the conformational study on sugar-phosphate-sugar interaction,[25] we have decided to use a Lennard-Jones type potential. We note that this potential has been used by several authors in describing intermolecular interaction for many biological systems including amino acids.[6,26] In the sugar-phosphate-sugar work, we have attempted to vary the 6–12 exponents of the Lennard-Jones potential into a pair of m–n exponents where m and n can be not only integers different from 6–12, respectively, but even fractional values. We concluded that the values 6–12 are nearly as good as those we have systematically analyzed. However, it was clear that the interactions of an oxygen atom with a different atom could not be described with a single potential, but by two different potentials, one of which represents the situation of an oxygen of the PO_4^- group, and

Table VIII. Geometry,[a] Charges, MO Valency State for Lysine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	1.457	3.059	−3.412	−0.39	0.38
O(2)	9	5.084	1.476	−5.108	−0.54	0.55
N	11	2.778	2.150	1.345	−0.55	0.91
N(Z)	11	3.923	−10.137	4.074	−0.56	0.91
C′	5	3.507	1.925	−3.248	0.50	1.36
C(A)	8	4.265	0.858	−0.659	−0.15	1.08
C(B)	7	3.826	−1.990	−0.638	−0.37	1.05
C(G)	7	4.400	−3.255	1.897	−0.36	1.06
C(D)	7	4.100	−6.110	1.736	−0.38	1.03
C(E)	8	4.260	−7.360	4.305	−0.28	1.09
H(1)	1	2.880	4.083	1.069	0.26	0.28
H(2)	1	0.918	1.591	1.315	0.26	0.29
H(A)	2	6.256	1.259	−0.325	0.18	0.27
H(B1)	2	1.863	−2.388	−1.136	0.19	0.28
H(B2)	2	5.070	−2.791	−2.089	0.17	0.27
H(G1)	2	6.333	−2.788	2.514	0.18	0.28
H(G2)	2	3.145	−2.557	3.369	0.20	0.28
H(D1)	2	2.273	−6.566	0.911	0.18	0.27
H(D2)	2	5.567	−6.889	0.506	0.17	0.27
H(E1)	2	6.067	−7.012	5.198	0.19	0.28
H(E2)	2	2.739	−6.685	5.519	0.19	0.28
H(Z1)	1	2.335	−10.525	3.068	0.25	0.28
H(Z2)	1	5.499	−10.913	3.262	0.24	0.28
H(O2)	4	6.486	0.580	−4.434	0.41	0.31

[a] The geometry has been taken from T. F. Koetzle, M. S. Lehmann, J. J. Verbist, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **28**, 3207 (1972). [b] Total energy −492.117 au.

Table IX. Geometry,[a] Charges, MO Valency State for Glutamine[b]

Atom	Class	x	y	z	Charge	MOVS
C′	5	14.907	8.505	7.153	0.49	1.39
C(A)	8	12.424	9.768	6.085	−0.12	1.12
C(B)	7	10.348	7.771	5.953	−0.38	1.03
C(G)	7	11.005	5.518	4.306	−0.42	1.01
C(D)	5	9.016	3.475	4.365	0.42	1.40
N	11	12.809	11.001	3.571	−0.55	0.91
N(E)	13	9.246	1.704	2.593	−0.63	1.02
O(1)	10	16.648	8.273	5.480	−0.40	0.36
O(2)	9	14.897	7.866	9.548	−0.54	0.53
O(E)	10	7.323	3.450	5.952	−0.42	0.37
H(1)	1	13.853	9.851	2.366	0.25	0.28
H(2)	1	11.120	11.353	2.704	0.24	0.28
H(A)	2	11.870	11.262	7.398	0.21	0.28
H(B1)	2	9.987	7.129	7.899	0.21	0.29
H(B2)	2	8.607	8.660	5.266	0.19	0.28
H(G1)	2	11.335	6.070	2.346	0.19	0.28
H(G2)	2	12.739	4.648	4.966	0.20	0.28
H(E2)	1	10.726	1.702	1.391	0.30	0.31
H(E1)	1	8.080	0.216	2.553	0.32	0.32
H(O2)	4	16.494	7.40	9.929	0.42	0.33

[a] The geometry has been taken from T. F. Koetzle, M. N. Frey, M. S. Lehmann, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **29**, 2571 (1973). [b] Total energy −526.726 au.

the second a different oxygen (e.g., of the sugar fragment).

This finding is extended and systematized in this paper. The interaction energy $I(\mathrm{M,W})$ is fitted by the following expression:

$$I(\mathrm{M,W}) = \sum_i \sum_{j \neq i} I_{ij}{}^{ab}(\mathrm{M,W})$$
$$= \sum_i \sum_{j \neq i} (-A_{ij}{}^{ab}/r_{ij}{}^6 + B_{ij}{}^{ab}/r^{12} + C_{ij}{}^{ab} q_i q_j / r_{ij})$$

and where A_{ij}, B_{ij}, and C_{ij} are fitting constants, r_{ij} is the distance between an atom i on M and an atom j on the water molecule, q_i and q_j are the net charges, NCH, for the atoms i and j, a and b are indices that not only distinguish between atoms of different Z value (e.g., an hydrogen atom from a carbon atom) but also, within a group of atoms of equal atomic number Z, differentiate its environmental conditions in the molecule, *according to some quantitative criteria*, obtained directly from the SCF-MO study of M. We note that, physi-

Table X. Geometry,[a] Charges, MO Valency State for Glutamic Acid[b]

Atom	Class	x	y	z	Charge	MOVS
O(2)	9	4.253	10.307	23.389	−0.53	0.52
O(1)	10	2.042	6.744	23.932	−0.41	0.39
O(E1)	10	4.083	9.984	13.747	−0.41	0.38
O(E2)	9	6.751	7.406	11.764	−0.54	0.54
N	11	5.450	3.981	21.250	−0.53	0.90
C′	5	3.763	7.943	22.970	0.52	1.41
C(A)	8	5.539	6.769	21.000	−0.12	1.10
C(B)	7	4.679	7.687	18.379	−0.39	1.03
C(G)	7	6.443	7.054	16.186	−0.42	1.02
C(D)	5	5.630	8.264	13.810	0.52	1.44
H(O2)	4	2.838	11.089	24.429	0.39	0.28
H(E2)	4	6.279	8.279	10.197	0.40	0.28
H(1)	1	5.651	3.476	23.129	0.25	0.28
H(2)	1	6.995	3.195	20.258	0.23	0.27
H(A)	2	7.452	7.388	21.399	0.18	0.27
H(B1)	2	4.453	9.716	18.465	0.22	0.28
H(B2)	2	2.855	6.914	18.003	0.22	0.29
H(G1)	2	8.357	7.709	16.668	0.21	0.28
H(G2)	2	6.633	4.982	15.871	0.22	0.29

[a] The geometry has been taken from A. Sequeira, H. Rajagopal, and R. Chidambaram, *Acta Crystallogr., Sect. B,* **28,** 2514 (1972). [b] Total energy −546.504 au.

Table XI. Geometry,[a] Charges, MO Valency State for Aspartic Acid[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	7.577	−0.524	11.055	−0.41	0.37
O(2)	9	9.104	−2.528	7.547	−0.54	0.52
O(D1)	10	−0.522	−2.436	6.969	−0.42	0.38
O(D2)	9	−1.277	−4.754	10.387	−0.55	0.54
N	11	4.615	−2.827	5.277	−0.65	0.86
C′	5	7.298	−1.616	8.980	0.52	1.41
C(A)	8	4.583	−1.854	7.972	−0.16	1.07
C(B)	7	2.939	−3.474	9.678	−0.57	0.88
C(G)	5	0.214	−3.512	8.866	0.55	1.55
H(1)	1	3.004	−2.887	4.708	0.32	0.30
H(2)	1	5.350	−4.444	5.207	0.29	0.29
H(A)	2	3.903	−0.121	7.868	0.26	0.30
H(B1)	2	3.013	−2.850	11.361	0.29	0.31
H(B2)	2	3.531	−5.196	9.713	0.26	0.30
H(D2)	4	−3.016	−4.687	9.898	0.42	0.33
H(O2)	4	10.669	−2.212	8.369	0.42	0.32

[a] The geometry has been taken from S. T. Rao, *Acta Crystallogr., Sect. B,* **29,** 1718 (1973). [b] Total energy −507.465 au.

Table XII. Geometry,[a] Charges, MO Valency State for Proline[b]

Atom	Class	x	y	z	Charge	MOVS
C′	5	10.362	12.748	2.204	0.52	1.41
C(A)	8	9.759	10.011	2.835	−0.13	1.06
C(B)	7	9.521	9.528	5.700	−0.38	1.03
C(G)	7	7.942	7.161	5.710	−0.40	1.01
C(D)	7	6.493	6.989	3.272	−0.30	1.06
N	11	7.353	9.214	1.683	−0.39	0.90
O(1)	10	9.175	13.964	0.594	−0.41	0.38
O(2)	9	12.209	13.670	3.578	−0.55	0.51
H(1)	1	7.631	8.652	−0.259	0.25	0.28
H(D1)	2	4.486	7.100	3.625	0.20	0.28
H(D2)	2	6.906	5.232	2.320	0.18	0.27
H(G1)	2	6.640	7.211	7.281	0.21	0.28
H(G2)	2	9.153	5.530	5.904	0.20	0.28
H(B1)	2	11.342	9.225	6.569	0.21	0.29
H(B2)	2	8.579	11.069	6.651	0.20	0.29
H(A)	2	11.253	8.831	2.099	0.18	0.27
H(O2)	4	12.439	15.386	3.106	0.41	0.32

[a] The geometry has been taken from Y. Mitsui, M. Tsuboi, and Y. Iitaka, *Acta Crystallogr., Sect. B,* **25,** 2182 (1969). [b] Total energy −397.270 au.

Table XIII. Geometry,[a] Charges, MO Valency State for Hydroxyproline[b]

Atom	Class	x	y	z	Charge	MOVS
O(2)	9	9.460	8.278	21.723	−0.54	0.52
O(1)	10	6.828	4.966	22.454	−0.42	0.37
O(D)	9	3.818	7.431	13.469	−0.53	0.45
N	11	6.848	9.892	17.686	−0.40	0.91
C′	5	7.483	6.864	21.236	0.51	1.39
C(A)	8	5.878	7.540	18.928	−0.13	1.08
C(B)	7	6.037	5.501	16.894	−0.41	1.00
C(G)	17	6.246	6.883	14.384	0.00	1.14
C(D)	7	7.689	9.274	15.078	−0.31	1.04
H(1)	1	5.393	11.229	17.695	0.26	0.28
H(A)	2	3.952	7.895	19.525	0.21	0.28
H(B1)	2	7.657	4.349	17.190	0.20	0.28
H(B2)	2	4.417	4.293	16.927	0.23	0.30
H(G)	2	7.300	5.788	12.983	0.17	0.27
H(D1)	4	3.966	7.703	11.661	0.36	0.30
H(B1)	2	9.700	8.918	15.147	0.20	0.28
H(B2)	2	7.310	10.812	13.841	0.20	0.29
H(O2)	4	10.286	7.590	23.160	0.41	0.32

[a] The geometry has been taken from T. F. Koetzle, M. S. Lehmann, and W. C. Hamilton, *Acta Crystallogr., Sect. B*, **29**, 231 (1973). [b] Total energy −471.875 au.

Table XIV. Geometry,[a] Charges, MO Valency State for Histidine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	2.200	0.065	7.068	−0.41	0.38
O(2)	9	−1.836	0.142	8.618	−0.55	0.52
N	11	4.031	−2.474	11.080	−0.70	0.83
N(D)	12	5.215	−0.837	15.940	−0.29	0.83
N(E)	15	3.662	−0.007	19.734	−0.46	1.14
C′	5	0.609	−0.235	8.783	0.52	1.41
C(A)	8	1.611	−1.059	11.400	−0.15	1.08
C(B)	7	2.044	1.255	13.101	−0.52	0.95
C(G)	14	3.015	0.551	15.680	0.07	1.15
C(E)	14	5.507	−1.129	18.411	−0.03	1.09
C(D)	14	2.047	1.071	18.008	−0.15	1.07
H(1)	1	4.764	−2.574	12.588	0.34	0.30
H(2)	1	5.126	−1.702	10.020	0.31	0.30
H(E2)	16	3.414	0.000	21.503	0.39	0.35
H(A)	2	0.401	−2.242	12.196	0.24	0,29
H(B1)	2	0.430	2.103	13.230	0.25	0.29
H(B2)	2	3.131	2.546	12.303	0.24	0.29
H(E1)	16	6.887	−2.020	19.114	0.25	0.33
H(D)	16	0.519	2.020	18.436	0.23	0.33
H(O2)	4	−2.196	0.691	6.947	0.42	0.32

[a] The geometry has been taken from P. Eddington and M. H. Harding, *Acta Crystallogr., Sect. B*, **30**, 204 (1974). [b] Total energy −543.309 au.

cally, the r^{-12} term represents the repulsion for the atoms i and j, the r^{-6} term the attractive interaction, the r^{-1} term, the point charge, point charge interaction, attractive or repulsive depending on the sign of q_j and q_j (C_{ij} is always positive). In principle, to each pair of atoms (one on the amino acid, M, the second on water) we could assign a *distinct* pair potential, namely a unique set of the constants A, B, and C. Thus, for example, in the case of the five amino acids, given in Tables I-V and containing a total of ten oxygen atoms, we could construct ten different pair potentials, describing the interactions of the ten oxygens on M with the oxygen on the water, an additional ten different pair potentials, describing the interactions of the ten oxygens on M with one of the hydrogens of water and finally ten additional and different pair potentials describing the interactions of the ten oxygens on M with the second hydrogen on water. To the 30 distinct interaction potentials will correspond 30 distinct sets of fitting constants (A, B, C). However, it is most reasonable to assume that among the 30 sets of parameters (A, B, C) several will be nearly equal

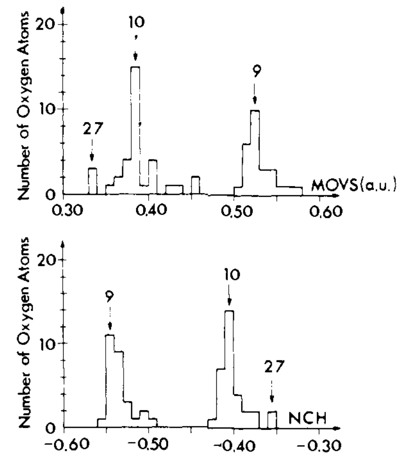


Figure 6. Histograms for the oxygen atoms in the 22 naturally occurring amino acids and four bases of DNA. The units in the abscissa are units of electronic charge for the NCH and atomic units for MOVS.

Table XV. Geometry,[a] Charges, MO Valency State for Tryptophan[b]

Atom	Class	*x*	*y*	*z*	Charge	MOVS
O(2)	9	19.563	0.205	8.010	−0.55	0.53
O(1)	10	6.401	−0.673	10.644	−0.41	0.38
N(E)	15	−0.066	3.270	0.042	−0.46	1.13
N	11	3.129	2.855	8.055	−0.54	0.51
C(Z2)	18	5.463	−0.005	−1.321	−0.20	1.15
C(H)	18	5.445	0.308	−3.968	−0.25	1.11
C(Z3)	18	3.536	1.511	−5.302	−0.22	1.11
C(Z1)	18	1.555	2.650	−4.114	−0.23	1.11
C(E)	20	1.569	2.376	−1.487	0.14	1.12
C(D2)	19	3.456	1.079	−0.070	−0.05	1.10
C(G)	14	2.819	1.199	2.492	−0.04	1.15
C(D1)	14	0.622	2.457	2.569	−0.09	1.10
C(B)	7	4.365	−0.025	4.719	−0.43	1.04
C(A)	8	5.281	1.765	6.792	−0.15	1.08
C′	5	7.188	0.448	8.682	0.51	1.40
H(Z2)	16	6.859	−0.952	−0.311	0.21	0.31
H(H)	16	6.837	−0.437	−5.084	0.21	0.30
H(Z3)	16	3.474	1.639	−7.262	0.21	0.30
H(Z1)	16	0.000	3.606	−5.139	0.19	0.30
H(D1)	16	−0.443	3.184	4.036	0.21	0.31
H(E)	16	−1.615	4.168	−0.292	0.37	0.34
H(B1)	2	5.974	−0.827	4.081	0.23	0.29
H(B2)	2	3.164	−1.342	5.669	0.20	0.28
H(A)	2	6.084	3.387	6.042	0.26	0.30
H(1)	1	1.482	1.936	7.796	0.24	0.22
H(2)	1	2.002	1.369	8.884	0.23	0.20
H(O2)	4	10.366	−0.862	9.210	0.42	0.32

[a] The geometry has been taken from R. A. Pasternak, *Acta Crystallogr.*, **9**, 341 (1950). [b] Total energy −679.361 au.

Table XVI. Geometry,[a] Charges, MO Valency State for Tyrosine[b]

Atom	Class	*x*	*y*	*z*	Charge	MOVS
O(1)	10	2.183	13.293	6.896	−0.42	0.38
O(2)	9	5.659	11.372	8.565	−0.55	0.52
O(E)	9	6.280	23.914	5.176	−0.51	0.54
C′	5	4.237	12.175	6.700	0.50	1.36
C(A)	8	5.222	11.701	4.024	−0.12	1.11
C(B)	7	7.511	13.396	3.438	−0.38	1.06
C(G)	19	7.070	16.180	3.899	−0.04	1.15
C(D1)	18	5.994	17.731	2.050	−0.19	1.14
C(D2)	18	7.799	17.289	6.175	−0.19	1.13
C(E1)	18	5.665	20.311	2.454	−0.26	1.10
C(E2)	18	7.500	19.867	6.602	−0.24	1.11
C(Z)	17	6.460	21.377	4.726	0.21	1.29
N	11	3.158	12.135	2.164	−0.56	0.90
H(1)	1	2.425	13.901	2.304	0.27	0.29
H(2)	1	1.684	10.879	2.452	0.25	0.28
H(A)	2	5.796	9.722	3.851	0.21	0.28
H(B1)	2	9.066	12.752	4.621	0.20	0.28
H(B2)	2	8.082	13.070	1.477	0.21	0.29
H(D1)	16	5.458	16.934	0.237	0.21	0.30
H(D2)	16	8.666	16.127	7.635	0.20	0.30
H(E1)	16	4.837	21.484	0.994	0.18	0.30
H(E2)	16	8.120	20.721	8.362	0.22	0.31
H(E)	4	5.583	24.805	3.707	0.39	0.31
H(O2)	4	4.837	11.798	10.104	0.41	0.32

[a] The geometry has been taken from M. N. Frey, T. F. Koetzle, M. S. Lehmann, and W. C. Hamilton, *J. Chem. Phys.*, **58**, 2547 (1973). [b] Total energy −624.030 au.

in value; therefore, we could group the 30 sets of parameters (*A*, *B*, *C*) into relatively few, however, really distinct sets, each one called "a class". This done, we are in a position to state that a given oxygen atom on M belongs to a class and for each class there is a distinct pair potential. The problems are now the following: (1) how to determine the *number* of classes and (2) how to obtain some objective criterium that will allow one to establish the class to which an atom belongs.

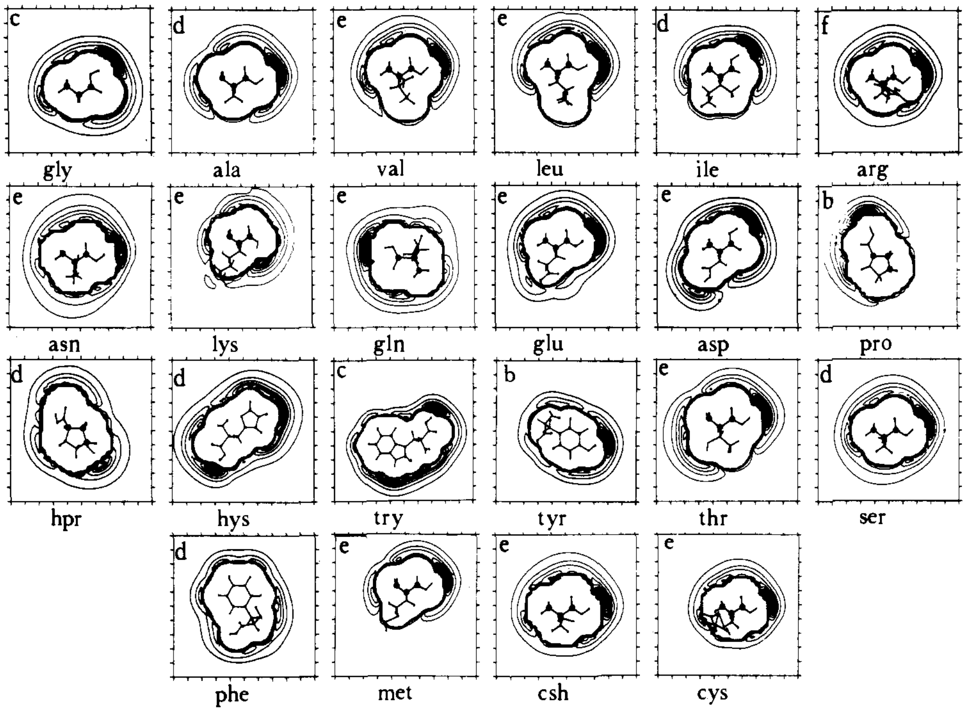


Figure 7. Interaction energy contours describing the interaction of a molecule of water with amino acids. The interval between contours is of 1 kcal/mol. See the text for additional explanation.

Table XVII. Geometry,[a] Charges, MO Valency State for Threonine[b]

Atom	Class	*x*	*y*	*z*	Charge	MOVS
O(1)	10	14.342	3.170	4.536	−0.42	0.38
O(2)	9	12.936	3.109	0.464	−0.55	0.51
O(G)	9	6.086	2.147	5.125	−0.41	0.42
N	11	10.626	0.203	6.234	−0.49	0.90
C′	5	12.748	2.685	2.900	0.50	1.35
C(A)	8	10.263	1.571	3.797	−0.15	1.07
C(B)	17	8.174	3.562	4.188	0.00	1.18
C(G)	6	7.472	4.893	1.773	−0.60	0.97
H(A)	2	9.954	0.102	2.526	0.23	0.29
H(B)	2	8.540	5.162	5.393	0.18	0.27
H(G1)	3	6.006	6.263	2.152	0.18	0.28
H(G2)	3	6.793	3.527	0.417	0.21	0.29
H(G3)	3	9.113	5.843	1.018	0.20	0.28
H(1)	1	11.320	1.502	7.645	0.24	0.28
H(2)	1	11.976	−1.304	5.973	0.23	0.27
H(G)	4	3.987	3.305	5.199	0.24	0.26
H(O2)	4	14.535	3.867	0.161	0.41	0.32

[a] The geometry has been taken from D. P. Shoemaker, J. Donohue, V. Shomaker, and R. B. Coray, *J. Am. Chem. Soc.*, **72,** 2328 (1950). [b] Total energy −434.095 au.

Table XVIII. Geometry,[a] Charges, MO Valency State for Serine[b]

Atom	Class	*x*	*y*	*z*	Charge	MOVS
O(1)	10	3.189	6.944	−1.571	−0.41	0.38
O(2)	9	6.521	8.809	0.419	−0.55	0.52
O(G)	9	8.385	2.913	−0.373	−0.52	0.45
N	11	2.961	2.935	1.358	−0.54	0.89
C′	5	4.853	7.003	0.099	0.52	1.41
C(A)	8	4.953	4.814	1.988	−0.14	1.08
C(B)	17	7.556	3.614	2.068	−0.18	1.13
H(A)	2	4.499	5.562	3.875	0.17	0.26
H(2)	1	1.245	3.878	1.314	0.26	0.28
H(1)	1	2.881	1.591	2.791	0.23	0.27
H(B1)	2	8.887	5.012	2.811	0.21	0.29
H(B2)	2	7.558	1.993	3.354	0.17	0.27
H(G)	4	7.790	1.198	−0.754	0.36	0.30
H(O2)	4	6.233	10.053	−0.842	0.42	0.32

[a] The geometry has been taken from M. N. Frey, M. S. Lehmann, T. F. Koetzle, and W. C. Hamilton, *Acta Crystallogr., Sect. B,* **29,** 876 (1973). [b] Total energy −395.270 au.

Table XIX. Geometry,[a] Charges, MO Valency State for Cystine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	10	6.623	2.652	4.243	−0.41	0.40
O(2)	9	4.099	4.782	6.862	−0.54	0.51
N	11	6.049	5.961	0.467	−0.54	0.90
C′	5	5.326	4.448	4.744	0.53	1.42
C(A)	8	4.946	6.659	2.950	−0.13	1.09
C(B)	21	2.191	7.437	2.730	−0.50	0.97
S(G)	22	0.008	4.954	1.926	0.11	0.19
H(2)	1	5.698	7.391	−0.849	0.24	0.28
H(1)	1	7.970	5.727	0.535	0.26	0.29
H(A)	2	6.048	8.245	3.703	0.18	0.27
H(B1)	2	1.560	8.184	4.567	0.22	0.29
H(B2)	2	1.993	9.013	1.369	0.19	0.28
H(O2)	4	4.421	3.369	8.062	0.40	0.32
O(O1)	10	−6.623	2.652	−4.243	−0.41	0.40
O(O2)	9	−4.099	4.782	−6.862	−0.54	0.51
N(N)	11	−6.049	5.961	−0.467	−0.54	0.90
C(C′)	5	−5.326	4.448	−4.744	0.53	1.42
C(CA)	8	−4.946	6.659	−2.950	−0.13	1.09
C(CB)	21	−2.191	7.437	−2.730	−0.50	0.97
S(SG)	22	−0.008	4.954	−1.926	0.11	0.19
H(H2)	1	−5.698	7.391	0.849	0.24	0.28
H(H1)	1	−7.970	5.727	−0.535	0.26	0.29
H(HA)	2	−6.048	8.245	−3.703	0.18	0.27
H(HB1)	2	−1.560	8.184	−4.567	0.22	0.29
H(HB2)	2	−1.993	9.013	−1.369	0.19	0.28
H(HO2)	4	−4.421	3.369	−8.062	0.40	0.32

[a] The geometry has been taken from D. D. Jones, I. Bernal, M. N. Frey, and T. F. Koetzle, *Acta Crystallogr., Sect. B*, **30**, 1220 (1974).
[b] Total energy −1433.387 au.

Table XX. Geometry,[a] Charges, MO Valency State for Cysteine[b]

Atom	Class	x	y	z	Charge	MOVS
S(G)	22	0.008	4.954	1.926	0.06	0.22
O(1)	10	6.623	2.652	4.243	−0.41	0.40
O(2)	9	4.099	4.782	6.862	−0.54	0.51
N	11	6.049	5.961	0.467	−0.54	0.90
C′	5	5.326	4.448	4.744	0.53	1.41
C(A)	8	4.946	6.659	2.950	−0.13	1.09
C(B)	21	2.191	7.437	2.730	−0.50	0.95
H(2)	1	5.698	7.391	−0.849	0.24	0.28
H(1)	1	7.970	5.727	0.535	0.26	0.29
H(A)	2	6.048	8.245	3.703	0.18	0.27
H(B1)	2	1.560	8.184	4.567	0.20	0.27
H(B2)	2	1.993	9.013	1.369	0.19	0.28
H(SG)	23	0.019	4.954	4.477	0.06	0.22
H(O2)	4	4.421	3.369	8.062	0.40	0.32

[a] The geometry has been taken from D. D. Jones, L. Bernal, M. N. Frey, and T. F. Koetzle, *Acta Crystallogr., Sect. B*, **30**, 1220 (1974).
[b] Total energy −717.233 au.

Initially, we have considered *two criteria* for each atom, its net charge, NCH, and its molecular orbital valency state, MOVS, in order to classify it into the classes identified by the indices *a* and *b*. There are many alternatives to such criteria, for example the use of the hybridization. A drawback with the hybridization, as a discriminant, is that it cannot be unequivocally reduced to a single numerical value (for example, the hybrid $2s^m2p^n$ has two parameters, m and n), especially if we consider the possibility of charge transfer. In addition, since both NCH and MOVS are related to the hybridization distribution of the electrons in an atom,[23] the use of NCH and MOVS, as criteria, most likely will result in a classification of the atoms (with some Z) not too different from the one obtainable by considering the hybridization.

Let us consider, as an example, the NCH values and the MOVS values for the oxygen atoms and for the hydrogen atoms in the 22 naturally occurring amino acids and for the four bases of DNA. In Figure 6, we report a hystogram where in the ordinate we report the number of times an oxygen assumes a given value either for the NCH (in units of electrons) or for the MOVS (in atomic units); the NCH values and the MOVS values are given on the abscissa. The hystogram for the oxygen atoms (Figure 6) reveals the existence of three peaks (two well defined, the third less defined) for NCH and of two distinct peaks for the MOVS. In the figure, the peaks are distinguished by the indices 9, 10, and 27. If we consider the oxygens present in the 22 naturally occurring amino acids and in the four bases of DNA, we immediately realize that these

Table XXI. Geometry,[a] Charges, MO Valency State for Methionine[b]

Atom	Class	*x*	*y*	*z*	Charge	MOVS
S	22	0.306	1.528	3.633	0.21	0.20
O(2)	9	−4.937	1.726	13.260	−0.55	0.52
O(1)	10	−1.638	−0.693	14.122	−0.43	0.38
N	11	1.782	2.718	12.282	−0.51	0.91
C′	5	−2.556	1.084	13.007	0.51	1.37
C(A)	8	−0.882	2.718	11.309	−0.13	1.11
C(B)	7	−0.918	1.501	8.542	−0.38	1.05
C(G)	21	0.684	2.984	6.676	−0.51	0.97
C(E)	21	2.718	3.109	1.904	−0.74	0.94
H(G1)	2	2.644	2.934	7.240	0.21	0.28
H(G2)	2	0.063	4.927	6.609	0.20	0.28
H(B1)	2	−2.844	1.454	7.867	0.20	0.28
H(B2)	2	−0.205	−0.409	8.652	0.20	0.28
H(A)	2	−1.604	4.626	11.246	0.21	0.29
H(1)	1	2.502	0.808	12.294	0.24	0.28
H(2)	1	1.813	3.470	14.179	0.23	0.28
H(E1)	3	3.676	4.433	3.127	0.21	0.29
H(E2)	3	1.885	4.112	0.334	0.21	0.29
H(E3)	3	4.059	1.740	1.201	0.21	0.29
H(O2)	4	−5.696	0.578	14.414	0.41	0.32

[a] The geometry has been taken from A. McL. Mathieson, *Acta Crystallogr.*, **5**, 332 (1952). [b] Total energy −794.990 au.

Table XXII. Pairwise Interaction Constants for Molecule–Water Complexes: Interaction with Oxygen Atoms of Water[a]

Class	*A*	*B*	*C*
1	0.227 048 409 + 03	0.319 285 464 + 04	0.998 202 369 + 00
2	0.344 879 279 + 01	0.332 071 268 + 04	0.999 966 606 + 00
3	0.344 879 279 + 01	0.704 513 690 + 04	0.998 410 933 + 00
4	0.389 675 373 + 03	0.307 583 102 + 04	0.100 093 916 + 01
5	0.242 698 627 + 03	0.570 163 312 + 04	0.999 941 483 + 00
6	0.102 073 144 + 02	0.294 559 020 + 06	0.999 955 520 + 00
7	0.110 689 075 + 03	0.276 483 570 + 06	0.999 073 327 + 00
8	0.102 073 144 + 02	0.345 947 762 + 06	0.100 345 021 + 01
9	0.112 916 672 + 02	0.305 269 293 + 06	0.999 645 844 + 00
10	0.102 073 144 + 02	0.261 569 141 + 06	0.100 092 629 + 01
11	0.124 076 189 + 02	0.631 8286125 + 06	0.999 096 863 + 00
12	0.220 215 490 + 02	0.200 655 234 + 06	0.996 603 809 + 00
13	0.318 320 038 + 03	0.615 968 625 + 06	0.100 005 582 + 01
14	0.126 181 303 + 03	0.130 406 014 + 06	0.101 622 795 + 01
15	0.161 787 306 + 04	0.626 337 617 + 06	0.100 060 296 + 01
16	0.233 495 670 + 03	0.792 550 488 + 04	0.100 239 015 + 01
17	0.102 073 144 + 02	0.102 467 840 + 06	0.993 570 738 + 00
18	0.102 073 144 + 02	0.520 450 148 + 06	0.999 982 804 + 00
19	0.199 325 260 + 03	0.136 718 561 + 05	0.109 999 999 + 01
20	0.655 663 666 + 03	0.911 956 562 + 06	0.101 060 666 + 01
21	0.229 295 538 + 03	0.155 507 061 + 05	0.100 122 450 + 01
22	0.139 064 225 + 02	0.174 304 908 + 07	0.995 094 240 + 00
23	0.254 667 215 + 03	0.369 866 099 + 05	0.105 506 256 + 01

[a] The interaction energy is expressed in kcal/mol; the distances are expressed in angstroms.

appear either in the -OH group or in the -CO unit of the -COOH group, or in the C═O group where the C is part of an aromatic ring structure; these three types of oxygens correspond to the three distributions designated with the three indices 9, 10, and 27, respectively. Thus, both the NCH and the MOVS can be used to: (a) define how many "classes" we must assign to the atoms of given *Z* value; and (b) to determine that a given atom (among those of given *Z* value) belongs to a given "class".

The final existence of "class" is then verified in the energy fitting procedure: if the set of constants (*A*, *B*, *C*) for any two classes is nearly equal, then there is little reason for keeping the two classes as two distinct ones. Recently, we have been experimenting with a rather new technique that not only helps in defining the classes, but present several advantages to obtain an accurate fit in a relatively short time. This technique decomposes the SCF-LCAO-MO energy directly into pairwise potentials, using an energy decomposition algorithm derived from the bond energy analyses.[22] A rather extended account of this new energy decomposition is available elsewhere.[27]

The *classes* obtained in our study on the amino acids are identified by a numeral from 1 to 23 and such "indices" are reported in Tables I-XXI. The *constants A*, *B*, and *C* for the 23 classes are given in Tables XXII and XXIII. With these

Table XXIII. Pairwise Interaction Constants for Molecule-Water Complexes: Interaction with Hydrogen Atoms of Water[a]

Class	*A*	*B*	*C*
1	0.207 521 847 + 01	0.173 850 511 + 04	0.100 088 759 + 01
2	0.824 252 534 + 01	0.145 269 011 + 04	0.100 000 000 + 01
3	0.130 720 375 + 01	0.280 374 992 + 03	0.998 704 314 + 00
4	0.159 769 344 + 01	0.463 124 774 + 04	0.100 093 982 + 01
5	0.356 004 411 + 01	0.184 585 125 + 06	0.999 759 927 + 00
6	0.356 004 411 + 01	0.389 186 260 + 05	0.100 027 880 + 01
7	0.190 353 591 + 02	0.618 960 681 + 04	0.100 000 000 + 01
8	0.117 872 994 + 03	0.235 846 649 + 04	0.100 101 370 + 01
9	0.845 311 871 + 02	0.870 600 693 + 03	0.100 005 104 + 01
10	0.163 284 536 + 03	0.762 464 880 + 04	0.999 389 425 + 00
11	0.141 507 664 + 03	0.277 916 666 + 04	0.100 009 498 + 01
12	0.491 524 463 + 02	0.486 883 081 + 04	0.998 375 371 + 00
13	0.347 660 562 + 01	0.682 350 397 + 04	0.100 299 819 + 01
14	0.968 582 392 + 01	0.130 161 777 + 06	0.100 080 614 + 01
15	0.441 930 092 + 03	0.127 116 438 + 05	0.100 051 089 + 01
16	0.356 528 398 + 01	0.104 377 336 + 04	0.998 747 133 + 00
17	0.356 004 411 + 01	0.121 598 151 + 06	0.992 949 978 + 00
18	0.356 004 411 + 01	0.404 495 615 + 05	0.100 236 642 + 01
19	0.356 004 411 + 01	0.157 257 820 + 06	0.874 550 834 + 00
20	0.385 323 176 + 02	0.177 231 436 + 06	0.981 899 209 + 00
21	0.604 173 851 + 01	0.545 685 400 + 05	0.100 149 776 + 01
22	0.259 413 876 + 03	0.918 035 120 + 04	0.101 107 869 + 01
23	0.102 959 207 + 03	0.776 556 488 + 04	0.102 515 377 + 01

[a] The interaction energy is expressed in kcal/mol; the distances are expressed in angstroms.

constants we can reproduce the 1690 computed interactions (amino acid and water) with an accuracy of about 0.6 kcal/mol (for more details on the standard deviation, see ref 18).

Conclusions

We neglect to present a detailed analysis of the interaction potentials, since they are available elsewhere.[18]

As previously stated, we are interested in using the obtained potentials (Tables XXII and XXIII) to learn how the water arranges around amino acids.

A partial answer (partial since here we do not consider the water-water interaction nor the temperature effects) is provided by the contour diagrams presented in Figure 7. For each amino acid we have selected a plane either defined by the three atoms N, C′, and CA (see Tables I-XXI) characterizing the H_2N-CHR-COOH backbone, or by the aromatic ring of the amino acid, when present. Onto each plane we have constructed a square grid of points (120 × 120) and at each grid point we have fixed the oxygen atom of a molecule of water: the position for the hydrogens has been obtained by minimizing the energy of interaction with the amino acids. Thus, for each grid point (the grid interval is 0.5 Å) we know the energy interaction for a water molecule, whose hydrogens are optimally rotated to ensure the minimum of interaction energy. We have neglected the contours corresponding to repulsive interactions higher than 3 kcal/mol; the interval between each contour is of 1 kcal/mol. The outermost contour is the zero energy contour and provides an *immediate,* even if approximate, *estimate of the volume of water that is expected to be perturbed by the amino acid.* The well-defined positions of the minima point out where, approximately, a molecule of water will tend to place itself and thus provides a first approximate idea of the number of molecules of water bound to the amino acid.

The positions of the nuclei of the amino acids are projected from the geometries given in Tables I-XXI onto the N, C′, CA plane, to allow an easier interpretation of the contour maps. The regions of hydrophobicity are very evident, since for such regions we observe only repulsive contours. *These maps show how the amino acids are "seen" by water, a rather basic information if one is interested in protein structure in solution.*[27] *These structures are clearly not the same as those structures seen by x rays.*[30]

We are of the opinion that Figure 7 represents one of the most accurately determined representation of biomolecules interacting with water. (A large number of contour diagrams is available in ref 18; see in addition ref 27.) Finally, we note that the rather drastic economic cost of these computations has been reduced by over a factor of 10 by making use of a particular feature of the IBMOL program, called ADD,[29] where the integrals for the molecule M and those for the molecule of water are separately computed once for ever: therefore, for each computation of the M-water complex, only those integrals specifically requested to represent the interaction of M with water need to be computed.

Acknowledgment. The work here presented required (among other things) a heavy effort in the preparation of "inputs", because of the unusually large number of computations performed. We are very much indebted to Drs. L. Barino, R. Barsotti, A. Martellani, R. Pavani, G. Ranghino, C. Tosi, and to Mr. Bolis for the collaboration on this tedious, but essential part of work. Their specific contributions are available elsewhere as documented in a series of technical reports (see ref 18).

References and Notes

(1) See, for example, E. Clementi, *J. Chem. Phys.*, **46**, 3851 (1967); in this study two molecules interacting are considered as "one single system of nuclei and electrons". This approach is often referred to as "supermolecule approach", a designation, however. that we accept with reservation, since the usual terminology like stable and unstable complex, inner and outer complex, suffices.
(2) See, for example, G. C. Lie and E. Clementi, *J. Chem. Phys.*, **62,** 2195 (1975).
(3) See, for example, O. Matsuoka, E. Clementi and M. Yoshimine, *J. Chem. Phys.*, in press.
(4) See, for example, O. Matsuoka, C. Tosi, and E. Clementi, *Int. J. Quantum Chem.*, submitted for publication.
(5) Different types of fitting can be found in H. Popkie and E. Clementi, *J. Chem. Phys.*, **57**, 1077 (1972); H. Kistenmacher, H. Popkie, and E. Clementi, *ibid.*, **59**, 5842 (1973).
(6) See, for example, A. Momamy, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975); **78**, 1595 (1974); L. L. Shipman, A. W. Burgess, and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 543, 854 (1975).

(7) See, for example, A. Pulmann, *Fortschr. Chem. Forsch.*, **31**, 45 (1972); B. Mely and A. Pulmann, *C. R. Acad. Sci.* **274**, 1371 (1972).
(8) H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **61**, 546 (1974).
(9) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **60**, 4455 (1974).
(10) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **59**, 3 (1973).
(11) G. C. Lie, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, **64**, 2314 (1976).
(12) G. C. Lie and E. Clementi, *J. Chem. Phys.*, **64**, 5308 (1976).
(13) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **58**, 5627 (1973).
(14) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **61**, 799 (1974).
(15) H. Kistenmacher, H. Popkie, and E. Clementi, *J. Chem. Phys.*, **58**, 1689 (1973).
(16) R. O. Watts, J. Fromm, and E. Clementi, *J. Chem. Phys.*, **61**, 2550 (1974); **62**, 1388 (1975); E. Clementi, R. Barsotti, J. Fromm, and W. Watts, *Theor. Chim. Acta*, **43**, 101 (1976). E. Clementi and R. Barsotti, "Study of the Structure of Molecular Complexes. 15. Accurate Coordination Numbers for Ions in Solution", to be submitted for publication.
(17) International Symposium of Theoretical Chemistry, Boulder, Col., June 1975; VI Simposio Chimici Teorici di Lingua Latina, Arles, France, Sept 1975.
(18) See the series of the reports "Analytical Potentials from "Ab Initio" Computations for the Interaction between Biomolecules", Technical Report I to VII, Istituto "G. Donegani", Via del Lavoro, 4, Novara, Italy, February 1976. Each Technical Report represents a study on a few of the amino acids considered in this paper; the reference's notation (18a, (18b, . . . , 18g) will be used to refer to the individual technical reports: DDC-761, DDC-762, DDC-763, DDC 764 and 765, respectively. The geometry of the amino acids is obtained from experimental geometries given in the ref 18 of the seven technical reports referred to above. Since the experimental geometries refer to amino acids in the zaitterion form, we have introduced the necessary geometry corrections in the $-NH_2$ and $-COOH$ groups, thus obtaining the coordinates reported in Tables I–XXI.
(19) A. H. Narten, *J. Chem. Phys.*, **56**, 5681 (1972).
(20) E. Clementi and D. R. Davies, *J. Comput. Phys.*, **1**, 223 (1966).
(21) R. S. Mulliken, *J. Chem. Phys.*, **23**, 1883, 1841, 2338, 2343 (1955). The charges we have used for H_2O are -8.6221 for the oxygen atom, and 0.6889 for each one of the hydrogen atoms.
(22) (a) E. Clementi, *J. Chem. Phys.*, **46**, 3842 (1967); (b) E. Clementi and A. Routh, *Int. J. Quantum Chem.*, **6**, 525 (1972); (c) H. Popkie and E. Clementi, *J. Chem. Phys.*, **57**, 4870 (1972).
(23) In ref 22a, this point is discussed in detail with some numerical application to methane and fluoro derivatives of methane.
(24) Ideally, the fit should reproduce exactly the tabulated values of *E*(M,W); practically, the fit will always retain a standard deviation small (as much as possible) but finite. In addition, the fit should be more accurate in the attractive regions than in the repulsive regions (typically, we neglect interaction energies repulsive by more than 10 kcal/mol), since the Boltzmann factor (in the Monte Carlo computation) will practically neglect highly repulsive situations.
(25) O. Matsuoka, C. Tosi, and E. Clementi, "Conformational Studies on Polynucleotide Chains. I. Description of the Hartree-Fock Energies for Nonbonded Interactions with Lennard-Jones Potential", *Biopolymers*, in press; C. Tosi, E. Clementi, and O. Matsuoka, "Conformational Studies on Polynucleotide Chains. II. Analysis of Steric Interactions and Derivation of Potential Functions for Internal Rotations", *Biopolymers*, in press; C. Tosi, E. Clementi, and O. Matsuoka, "Conformational Studies on Polynucleotide Chains. III. Intramolecular Energy Maps and Comparison with Experiments", *Biopolymers*, in press.
(26) W. K. Olson and P. J. Flory, *Biopolymers*, **11**, 1, 25, 57 (1972).
(27) E. Clementi, "Lecture Notes in Chemistry", Vol. II, Springer-Verlag, Heidelberg, W. Germany, 1976.
(28) The geometry of the complex animo acid–water is defined by giving the *x, y, z* coordinates of each atom (relative to a selected Cartesian frame of axes). The coordinates for the atoms in each amino acid are given in Tables I–V. The coordinates of the water molecule are given in the appendix of ref 18a–18g. These tables can be obtained by writing to the authors of this paper. In these tables we report a code number for each computation, the *x, y, z* coordinates of the two hydrogen atoms and of the oxygen atom and the computed total energy (in atomic units) for the amino acid–water complex.
(29) E. Clementi, *Int. J. Quantum Chem.*, **15**, 307 (1967). See, in addition, E. Clementi and J. Mehl, Special IBM Technical Reports: (a) IBMOL-V Program's System Manual (June 1975); (b) IBMOL-V Program's User Guide (July 1971). The program version used in this work is IBMOL-VI, an improved version of IBMOL-V. The documentation for such a version is in progress.
(30) NOTE ADDED IN PROOF. The interaction of water with an enzyme has been recently computed and similar conclusions are obtained; see, E. Clementi, G. Ranghino, and R. Scordamaglia, "Intermolecular Potentials: Interaction of Water with Lysozyme", *Chem. Phys. Lett.*, in press.

Analytical Potentials from "ab Initio" Computations for the Interaction between Biomolecules. 2. Water with the Four Bases of DNA[1]

R. Scordamaglia, F. Cavallone, and E. Clementi*

Contribution from the Società Montedison, Istituto Ricerche "G. Donegani", 28100 Novara, Italy.

Received August 30, 1976

Abstract: SCF-LCAO-MO computations are presented for the four bases of DNA interacting with one molecule of water located at different positions and orientations around each base. The 368 computed total energies, E, for the water-base complex have been fitted with an analytical potential of the form $E = \sum_i \sum_{j\neq i} (-A_{ij}{}^{ab}/r_{ij}{}^6 + B_{ij}{}^{ab}/r_{ij}{}^{12} + C_{ij}{}^{ab}/r_{ij}) + E(\text{base}) + E(\text{water})$, where i and j designate two atoms, one on the base and the other one in the water molecule, a is an index that distinguishes the electronic environment of an atom in the base, b is an index that distinguishes between either a hydrogen or an oxygen atom in the water molecule, A, B, and C are fitting constants, E(base) and E(water) are the total energies of the base and of the water molecule, respectively, at infinite separation from one another. The overall standard deviation of the fit (comparing the SCF-LCAO-MO energies with those obtained by the above analytical potential) is 0.68 kcal/mol. In this work we have made use of recently reported potentials describing the interaction of water with 21 amino acids.

In the previous paper[2] of this series we have reported pair potentials representing the interaction energy of a water molecule with 21 naturally occurring amino acids. In this paper we extend our work by reporting the pair potentials for the interaction of water with the four bases of DNA. In a following paper,[3] we shall examine the overall quality of these potentials, taking the phenylalanine (the last one of the series of naturally occurring amino acids) as an example.

The potentials are obtained by fitting the computed interaction energies (in the SCF-LCAO-MO approximation) between a molecule M (amino acid or bases of DNA) and a molecule of water, placed at a large number of positions and orientations around M (a total of over 2000 positions and orientations have been computed in the SCF-LCAO-MO approximation to ensure we have a sufficient sample of the potential energy surfaces representing the interaction of the molecules M with water).

We have concentrated our effort on a few building blocks