(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 –NH₂ 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 H₂O 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, $E(\text{base})$ and $E(\text{water})$ 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

Table I. Geometry,[a] Charges, and Molecular Orbital Valency States for Adenine[b]

Atom	Class	x	y	z	Charge	MOVS
N(1)	12	4.457	2.596	0.000	−0.33	0.90
N(2)	12	2.176	−1.256	0.000	−0.32	0.94
N(3)	12	−2.416	3.516	0.000	−0.29	0.89
N(4)	15	−2.457	−0.751	0.000	−0.47	1.33
N(5)	11	2.280	6.384	0.000	−0.60	1.28
C(1)	24	2.280	3.852	0.000	0.25	1.52
C(2)	19	0.000	2.588	0.000	−0.01	1.14
C(3)	20	0.000	0.000	0.000	0.21	1.30
C(4)	25	4.301	0.086	0.000	0.05	1.41
C(5)	25	−3.798	1.373	0.000	0.05	1.22
H(1)	16	6.091	−0.968	0.000	0.20	0.28
H(2)	16	−3.119	−2.521	0.000	0.38	0.33
H(3)	16	−5.876	1.373	0.000	0.21	0.30
H(4)	1	0.696	7.413	0.000	0.33	0.30
H(5)	1	3.865	7.413	0.000	0.32	0.30

[a] The geometry has been taken from M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959). [b]Total energy −462.841 au.

Table II. Geometry,[a] Charges, and Molecular Orbital Valency States for Guanine[b]

Atom	Class	x	y	z	Charge	MOVS
O	27	8.535	7.333	0.000	−0.36	0.35
N(1)	15	9.288	3.040	0.000	−0.49	1.22
N(2)	12	5.899	0.122	0.000	−0.38	0.92
N(3)	12	3.158	6.191	0.000	−0.25	0.91
N(4)	15	1.748	2.159	0.000	−0.46	1.34
N(5)	11	9.997	−1.260	0.000	−0.62	1.28
C(1)	24	8.353	0.666	0.000	0.42	1.58
C(2)	20	4.334	2.089	0.000	0.21	1.30
C(3)	19	5.156	4.544	0.000	−0.05	1.09
C(4)	26	7.747	5.166	0.000	0.37	1.67
C(5)	25	1.161	4.603	0.000	0.02	1.20
H(1)	16	−0.804	5.280	0.000	0.20	0.30
H(2)	16	0.559	0.690	0.000	0.38	0.33
H(3)	1	9.461	−3.072	0.000	0.35	0.31
H(4)	1	11.871	−1.013	0.000	0.31	0.30
H(5)	16	11.161	3.286	0.000	0.35	0.31

[a]The geometry has been taken from M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959). [b] Total energy −537.459 au.

Table III. Geometry,[a] Charges, and Molecular Orbital Valency States for Cytosine[b]

Atom	Class	x	y	z	Charge	MOVS
O	27	15.682	−0.460	0.000	−0.40	0.33
N(1)	15	19.039	2.159	0.000	−0.48	1.24
N(2)	12	14.826	3.649	0.000	−0.36	0.83
N(3)	11	14.140	7.920	0.000	−0.60	1.28
C(1)	26	16.471	1.706	0.000	0.46	1.66
C(2)	24	15.768	5.980	0.000	0.26	1.50
C(3)	25	18.318	6.522	0.000	−0.31	1.26
C(4)	25	19.995	4.524	0.000	−0.01	1.43
H(1)	1	12.264	7.689	0.000	0.33	0.30
H(2)	1	14.692	9.727	0.000	0.31	0.29
H(3)	16	18.995	8.487	0.000	0.19	0.28
H(4)	16	22.048	4.849	0.000	0.23	0.30
H(5)	16	20.228	0.690	0.000	0.37	0.31

[a]The geometry has been taken from M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959). [b]Total energy −391.161 au.

of biochemical interest, namely amino acids and DNA bases; as known, the amino acids are components of proteins; the four bases, presented here, are components of DNA.

In the first paper of this series,[2] we have provided a rather extended set of references[4] where the method here used to study the solvation of water with biomolecules has been *successfully tested*. In our approach we have kept in mind the following constraints: (a) the practical impossibility to carry out an investigation of *many* molecules of water around a biomolecule, if one selects to use ab initio quantum-mechanical computation; (b) the need to include temperature effects in any realistic study of solutions; and (c) the need to consider the water–water interactions in addition to the consideration of the interaction of a single molecule of water with a biomolecule.

With this in mind we have: (a) computed the water–water interaction to rather high accuracy; (b) computed the water–biomolecule interactions in the SCF–LCAO–MO approxi-

Table IV. Geometry,[a] Charges, and Molecular Orbital Valency States for Thymine[b]

Atom	Class	x	y	z	Charge	MOVS
O(1)	27	0.000	7.463	0.000	−0.40	0.33
O(2)	27	4.188	0.000	0.000	−0.36	0.33
N(1)	15	−2.258	3.854	0.000	−0.47	1.28
N(2)	15	2.305	3.939	0.000	−0.49	1.21
C(1)	26	0.000	5.158	0.000	0.51	1.66
C(2)	25	−2.258	1.304	0.000	−0.05	1.39
C(3)	28	0.000	0.000	0.000	−0.09	1.22
C(4)	26	2.307	1.332	0.000	0.35	1.60
C(5)	6	0.000	−2.910	0.000	−0.58	1.46
H(1)	16	3.903	5.082	0.000	0.36	0.32
H(2)	16	−3.792	5.082	0.000	0.37	0.31
H(3)	16	−3.930	0.134	0.000	0.23	0.29
H(4)	3	−1.672	−3.627	−0.965	0.19	0.26
H(5)	3	0.000	−3.627	1.930	0.20	0.26
H(6)	3	1.672	−3.627	−0.965	0.22	0.28

[a] The geometry has been taken from M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959). [b]Total energy −449.833 au.

mation with a basis set sufficiently large as not to give gross errors;[2] (c) obtained a set of pair potentials that when used in conjunction with the water–water potential allows a rather accurate description of the interactions of *many* molecules of water with a biomolecule; and (d) devised a rather flexible form of the pair potential (for the molecule M and water) such that several aspects of this work performed on a few specific molecules might be easily transferred to the study of the interaction between similar molecules.

A number of different approaches have been presented in literature, and we refer to a recent review paper on the subject,[5] and to the references given in the first paper of this series.[2] We think that some of the other methods[5] adopted today to study the water solvation in biomolecules possibly present the limitation of not being adapted to statistical mechanical extensions (it is known that the internal energy alone is a parameter totally insufficient to describe the thermodynamics of solution), nor to be constructed in such a way as to allow easy transferability.[4] In the review paper by Pullman and Pullman[5] one can find a rather detailed description of one water molecule at a few positions and orientations around adenine, one of the four bases studied in this work; a detailed comparison with this work is not feasible, however, since the input geometries are not available for the computation published by the Pullmans.

Results and Discussion

Computation of the Interaction Energy. As stated above the interactions of the four bases of DNA 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.[6] The basis set used here is given elsewhere;[2] we note that this basis set is very similar to the one we selected some time ago to describe the electronic structure of the four bases.[7]

For the four molecules, adenine, cytosine, guanine, and thymine, we have selected the geometry given in ref 8. In Tables I-IV the geometry is reported by presenting the x, y, and z coordinates of the atoms (in atomic units). In addition, in these tables we report a code name for each atom (and other quantities explained later).

Each base, M, is considered fixed in space and the water molecule, W, is placed at different positions and orientations around M. For each position and orientation of the water molecule we have computed the total energy of the system M + W, 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 finite distance from the water molecule, we obtain the interaction energy, $I(M,W)$, between M and water, namely $I(M,W) = E(M,W) − E(M) − E(W)$. The value of $E(M)$ for the four bases considered here is given at the bottom of Tables I-IV (in atomic units).

In Tables I-IV, the last two columns report the computed net charges, NCH, following the formalism proposed by Mulliken,[9] and the molecular orbital valency state energy (MOVS), following the formalism proposed by Clementi.[10] Such quantities represent the gain (or the loss) in the electronic population of an atom (NCH) and the energy variation for the same atom (MOVS) characterizing the electronic rearrangement in the formation of a molecule when one starts from its component atoms (considered in the ground state atomic configuration). Therefore the quantity NCH and MOVS can be used as quantum mechanical indices, that in a simple way partly described the electronic environment of an atom when in a molecule.

We have considered 368 positions or orientations of the water molecule with the four bases, more specifically 82, 73, 88, and 125 for adenine, guanine, cytosine, and thyamine, respectively.[11] In such computations we have included cases yielding strongly repulsive interactions as well as strong and moderate attractions. The distances from one atom of water to one atom in the base varies from a minimum of about 3.5 au to a maximum of about 20 au.[11]

Analytical Expression of the Interaction Energy. As done in the previous paper,[2] we express the interaction energy $I(M,W)$ as the sum of pair potentials between all possible pairs of atoms, one on M (designated by the index "i") and the second on water (designated by the index "j"). In addition we select another set of indices "a" and "b" to further characterize the pair of atoms on the two interacting molecules. Such additional characterizations differentiate for a given atom, within a group of atoms of equal atomic number Z, the specific electronic environment of that atom in a molecule.

For the interaction energy we write:

$$I(M,W) = \sum_i \sum_j I_{ij}{}^{ab}(M,W)$$

where the pair potential is expressed as

$$A_{ij}{}^{ab} = -A_{ij}{}^{ab}(r_{ij})^{-6} + B_{ij}{}^{ab}(r_{ij})^{-12} + C_{ij}{}^{ab}q_iq_j(r_{ij})^{-1}$$

and where the A, B, and C quantities are fitting constants; the quantities q_i and q_j are the net charges, NCH, for the atom

Table V. 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	0.999 966 806 + 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	0.631 822 125 + 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
24	0.167 215 482 + 04	0.576 040 719 + 06	0.100 573 744 + 01
25	0.149 723 616 + 02	0.241 460 602 + 06	0.999 616 608 + 00
26	0.229 376 206 + 02	0.211 941 961 + 06	0.996 846 467 + 00
27	0.233 567 953 + 02	0.341 006 609 + 06	0.994 164 132 + 00
28	0.102 073 144 + 02	0.360 925 835 + 05	0.108 784 823 + 01

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

Table VI. 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
24	0.943 062 698 + 03	0.970 811 281 + 06	0.998 235 397 + 00
25	0.649 988 854 + 01	0.173 948 033 + 06	0.995 092 936 + 00
26	0.974 940 002 + 01	0.189 976 014 + 06	0.997 791 417 + 00
27	0.207 783 979 + 03	0.296 326 648 + 04	0.100 600 804 + 01
28	0.649 988 854 + 01	0.649 861 953 + 06	0.856 379 583 + 00

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

i and the atom j, respectively. The indices "a" and "b" constitute the "classes" within a group of atoms of equal atomic number Z.

We have made use of the computed values of NCH and MOVS to classify the atoms into classes (see the previous papers of this series[2]). The classes are distinguished by a number and in Tables V and VI we report the entire set of constants A, B, and C needed to fit the computed interaction energies for amino acids and bases. The bases studied in this work required an extension of our previous classification[2] to five additional
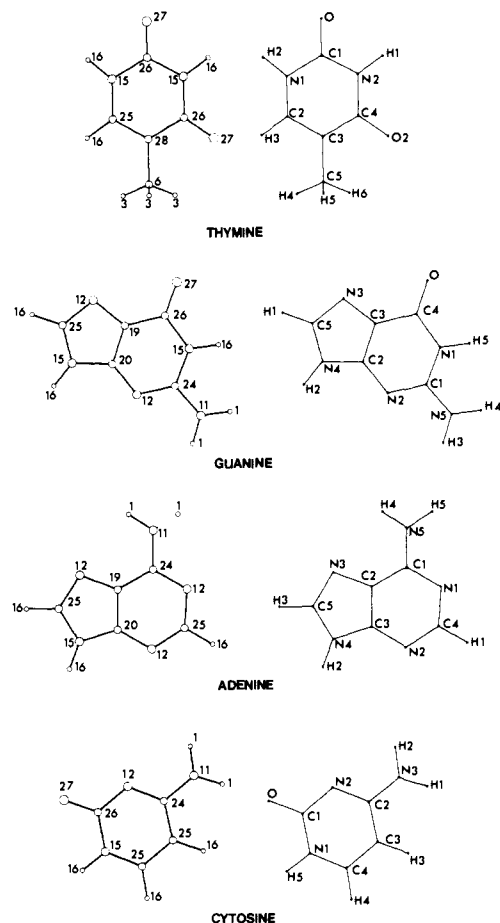
Figure 1. Geometry and identification code for the atoms in the bases. Each base is described twice, once using the alphanumerical identification code reported in Tables 1-1V, the second time using the "class" identification reported in Tables 1-1V.

Figure 2. Interaction energy contour maps for adenine and guanine with water (the letters a, b, and c identify the plane of the molecule and planes 4.0 and 6.0 au above the molecular plane, respectively).

classes (classes 24-28). The correspondence between atom and class can be obtained from Figure 1, where we give the class for each atom and the code name used in Tables I-IV.

The carbon atom in the aromatic ring, connected to a carbon and a nitrogen atom (both in the ring) and to a terminal $-NH_2$ group, is specified by class 24. A carbon atom in a ring, connected to two nitrogen atoms (in the ring) is specified by class 25. A carbon atom in a ring, connected to two carbon atoms and to a terminal hydrogen, is specified by class 26; if the same carbon atom is connected to the terminal group $-CH_3$, then the carbon atom is specified by class 28. The oxygen atom connected to a ring structure, but not part of it (see, for example, thyamine, in Figure 1), is specified by class 27.

A detailed analysis of the pair potentials presented in this paper is given elsewhere.[11] However, we provide in Figures 2 and 3 the isoenergy contour maps representing the interaction of one molecule of water with the four bases.

For each base we have selected three planes; one is the main molecular plane, the others are planes parallel to the first, but displaced by 4.0 and 6.0 au. The interval between successive contours is 1.0 kcal/mol. The maps are obtained by placing the oxygen of a molecule of water at each intersection of a square grid (of dimension 60 × 60) placed onto each of the planes. The hydrogens are allowed to assume the orientation that corresponds to an energy minimum (a detailed description on the construction of these maps is available in the first paper of this series). The position of the minima in the planes considered are very easy to identify (the outermost contour corresponds to the zero energy contour). In the Monte Carlo study the position of many molecules of water surrounding each base will provide
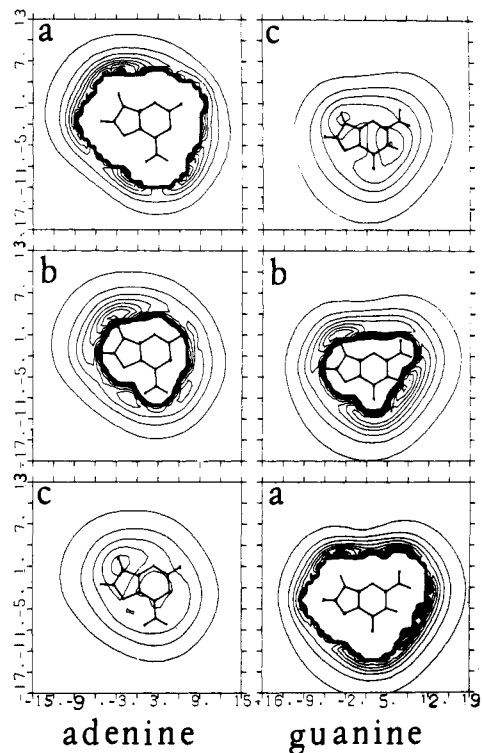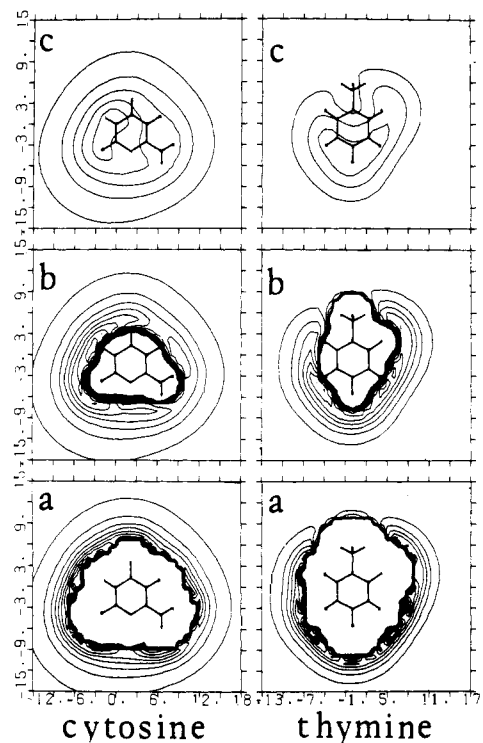

Figure 3. Interaction energy contour maps for cytosine and thymine with water (the letters a, b, and c identify the plane of the molecule and planes 4.0 and 6.0 au above the molecular plane, respectively).

a realistic representation of the bases in solution; however, the maps of Figures 2 and 3 provide a representation that, even if preliminary, we think is useful. As previously noted[2] these maps provide us with a very clear representation on the structure of the DNA's basis as seen by water. We stress that:

(i) the structures of the bases *as seen by water* are quite different from the structures of the base *as seen*, for example, *by x rays;* and (ii) concept like hydrophobicity (see, for example, Figure 3, insert a) can be put on more quantitative grounds because of the availability of potentials of the type given in Tables V and VI.

References and Notes

(1) This work has been partially presented at the International Symposium of Theoretical Chemistry, Boulder, Col., June 1975, and at the VI Simposio dei Chimici Teorici di Lingua Latina, Arles, France, September 1975.
(2) E. Clementi, F. Cavallone, and R. Scordamaglia, *J. Am. Chem. Soc.*, preceding paper in this issue.
(3) G. Bolis and E. Clementi, *J. Am. Chem. Soc.*, following paper in this issue.
(4) E. Clementi, "Lecture Notes in Chemistry", Vol. 2, Springer-Verlag, Heidelberg, W. Germany, 1976.
(5) B. Pullman and A. Pullman, *Q. Rev. Biophys.*, **7**, 4 (1975).
(6) E. Clementi and D. R. Davies, *J. Comput. Phys.*, **1**, 223 (1966); E. Clementi, *Int. J. Quantum Chem., Symp.*, **No.** 1, 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 Users Manual (July 1971). E. Clementi, *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 2942 (1972).
(7) E. Clementi, J. M. André, M. Cl. André, D. Klint, and D. Hahn, *Acta Phys. Acad. Sci. Hung.*, **27**, 493 (1968). The same basis set has been used to study the hydrogen bond between base pairs: E. Clementi, J. Mehl, and W. von Niessen, *J. Chem. Phys.*, **54**, 508 (1971); in this work the basis set has been compared to a larger basis set.
(8) M. Spencer, *Acta Crystallogr.*, **12**, 59, 66 (1959).
(9) R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833, 1841, 2338, 2343 (1955). The electronic charges we have computed for water are 8.6221 for the oxygens atom and 0.6889 for each one of the hydrogen atoms. The dependence of the NCH on the basis set has been analyzed for the case of water by E. Clementi and H. Popkie, *J. Chem. Phys.*, **57**, 1077 (1972).
(10) E. Clementi, *J. Chem. Phys.*, **46**, 3842 (1967); E. Clementi and A. Routh, *Int. J. Quantum. Chem.*, **6**, 525 (1972); H. Popkie and E. Clementi, *J. Chem. Phys.*, **57**, 4870 (1972).
(11) The geometry of the complex "amino 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 Table I–IV. The coordinates of the water molecule are given in a set of tables, where 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. These tables are available as a Montedison Technical Report, DDC-766, February 1976; a copy of the report can be secured by request to the authors of this paper.

Analytical Potentials from "ab Initio" Computations for the Interaction between Biomolecules. 3. Reliability and Transferability of the Pair Potentials[1a]

G. Bolis[1b,c] and E. Clementi*[1d]

Contribution from the Università degli Studi di Milano, Facoltà di Fisica, 20100 Milan, Italy, and the Società Montedison, Istituto Ricerche "G. Donegani", 28100 Novara, Italy. Received August 30, 1976

Abstract: The pair potentials reported in the previous papers of this series have been analyzed using phenylalanine as a test case. For this amino acid (the last one of the naturally occurring amino acids to be studied) we have computed (in the SCF–LCAO–MO approximation) the interaction with a molecule of water placed at 75 different positions (or orientations) relative to phenylalanine. The computed interaction energies can be duplicated to a reasonable level of accuracy by making use of the previously reported pair potentials (with an error of about 1 kcal/mol); therefore, we conclude that our pair potentials are, to some extent, transferable to molecules that are chemically similar to those previously studied. We have discussed problems related to: (1) more refined selections of the points needed to sample the multi-dimensional interaction energy surface; and (2) the need to select a more physical form for the pair potential.

In the previous papers of this series we have presented a number of pair potentials to describe the interaction energy of a molecule of water (W) with several molecules (M); in particular we have considered 21 out of the 22 naturally occurring amino acids[2] and the four bases of DNA.[3]

In this paper we shall discuss some aspects related to the *reliability* of the pair potentials for those molecules we have considered; we shall, in addition, investigate the possibility of using the potentials to describe the interaction energy for molecules not explicitly considered in our work (i.e., *transferability* of the pair potentials). As a test case for the transferability we shall use the amino acid phenylalanine, previously not studied in this series. In Table I and in Figure 1 we report the geometry of phenylalanine; since the geometry given in ref 4 refers to the zwitterionic form, we have introduced in Table 1 and Figure 1 the necessary coordinate variations for the $-NH_2$ and $-COOH$ groups. In Table I, we give a code name for each atom, a "class" specification (see ref 2 and 3 and following discussions), the *x, y, z* coordinates for each atom (in atomic units), the computed net charges,[5] NCH, and the computed molecular orbital valency state energies,[6] MOVS.

The basis set used to compute the phenylalanine molecules, the water molecules, and the phenylalanine–water complexes has been discussed in ref 2. The total energy for the water molecules, $E(W)$, and for the phenylalanine, $E(M)$, are $E(W) = -75.733\ 057$ and $E(M) = -549.373\ 069$ au, respectively.

For the complex we have considered 75 different positions or orientations of the water relative to phenylalanine. In Table II we report a running index (1–75) to identify the computation of the complex, the *x, y, z* coordinates of the atoms for water (the phenylalanine is kept fixed in space at the position specified in Table I), the total energy of the complex $E(M,W)$ expressed in atomic units, the interaction energy $I(M,W) = E(M,W) - E(M) - E(W)$, expressed in kilocalories per mole, and in the last two columns two different approximations to $I(M,W)$, discussed later in this paper.

In the preceding papers of this series we have not reported the equivalent of Table II, for obvious reasons of space (about 2000 cases have been studied); however, such tables are available to the interested reader as specified in ref 2 and 3. For these computations we have used a computer program, named IBMOL.[7]