Ab Initio Pair Potentials for the Interactions between Aliphatic Amino Acids

J. A. Sordo,*† M. Probst,‡ G. Corongiu, S. Chin, and E. Clementi

Contribution from the IBM Corporation, Data Systems Division, Department 48B/MS 248, Kingston, New York 12401. *Received August 18, 1986*

Abstract: A pair potential describing the potential hypersurface between interacting aliphatic amino acids (without sulfur) is presented. This pair potential has been derived entirely from ab initio calculations at the Hartree-Fock level. Almost two thousand SCF calculations have been performed and used as input for a nonlinear least-squares fitting in order to obtain the parameters for the atom-atom analytical pair potential.

The rigorous theoretical study of biological systems is still far removed from the present day computational facilities.[1] In fact, the theoretical background is available but its practical implementation for large molecular systems demands both very powerful machines and very long computing times. In order to obtain information on these kinds of systems, some approximate treatment must be used.[2] Some of the many semiempirical methods available in the literature can be used; however, many problems still remain when using such methods. In fact, in spite of the remarkable reduction in computation time when passing from ab initio to semiempirical methods, this factor continues to be a very serious drawback when dealing with large systems. A very interesting alternative is provided by analytical pair potentials.[3-10] Once an analytical expression representing the potential hypersurface is available, the interaction energy between very large systems can be evaluated relatively fast. In addition, an analytic expression brings about the possibility of using statistical mechanics, thus providing the ability to predict thermodynamic properties of the system under study.[11,12]

The parameters included in the analytical pair potential may be obtained from experimental data (empirical pair potentials),[5,6] from semiempirical calculations (semi-empirical pair potentials),[3,4] or from calculations based on first principles (ab initio pair potentials).[7,8] (There are also pair potentials that cannot be assigned just to one of these three classes (see, for example, ref 9 and 10).)

One major advantage of the latter pair potentials compared to the others is that information concerning any point on the potential hypersurface is accesible from calculations but not always from experiments.[13] In this paper we present an ab initio pair potential for the interaction between aliphatic amino acids (without sulfur). This research is a continuation of the work on ab initio analytical pair potentials for the interaction between water molecules[14-16] and between amino acids and water[7,8] that has been systematically developed at this laboratory. In addition, work is currently in progress to derive the corresponding analytical pair potentials for the amino acids containing sulfur, as well as for the nonaliphatic amino acids.

At the completion of this work, a rather comprehensive set of analytical pair potentials for amino acids and for amino acids and water will have been derived. This set of pair potentials will provide a very powerful tool for theoretical studies of systems of biological interest.

At the present time, the application of pair potentials, like those presented in this paper, to perform MC and MD simulations on systems containing molecules of biological interest has been shown to be a very useful tool in providing structural and energetic information of such systems.[17-34] The development of a pair potential for the interaction between amino acids will facilitate additional studies of these type of systems, like, for example, the 3-D structure of proteins, a subject of fundamental importance in biochemistry.

† Permanent address: Departamento de Quimica Fisica, Universidad de Oviedo, Oviedo, Principado de Asturias, Spain.

‡ Present address: Institute of Inorganic and Analytical Chemistry, University of Innsbruck, A-6020, Innsbruck, Austria.

Table I. Charges (au) from Mulliken Population Analysis for Atoms in Alanine as Calculated from Monomer (ALA) and from a Very Stable Conformation in the Dimer (ALA-ALA) (see ref 7 for notation)

atom	q_i(ALA)	molecule 1 q_i(ALA-ALA)	molecule 2 q_i(ALA-ALA)
O(1)	8.4333	8.4870	8.4866
O(2)	8.5632	8.5778	8.5783
N	7.6104	7.6098	7.6098
C′	5.4843	5.4600	5.4610
C(A)	6.1401	6.1413	6.1413
C(B)	6.6379	6.6391	6.6390
H(A)	0.7737	0.7682	0.7683
H(1)	0.7153	0.7170	0.7171
H(2)	0.7275	0.7273	0.7274
H(B1)	0.7886	0.7895	0.7895
H(B2)	0.7814	0.7776	0.7777
H(B3)	0.7660	0.7634	0.7635
H(O2)	0.5783	0.5412	0.5412

The procedure for obtaining the analytical pair potentials from ab initio calculations is well-defined in previous works.[7,8] However,

(1) Clementi, E. *Lecture Notes in Chemistry*; Springer-Verlag: New York, 1980; Vol. 19, also references therein.
(2) Pullman, A., Ed. *From Diatomics to Biopolymers*; Wiley: New York, 1978.
(3) Pullman, A. *Fortschr. Chem. Forsch.* **1972**, *31*, 45.
(4) Melly, B.; Pullman, A. *C. R. Acad. Sci.* **1972**, *274*, 1371.
(5) Momamy, A.; McGuire, R. F.; Burguess, A. W.; Scheraga, H. A. *J. Chem. Phys.* **1974**, *78*, 1595; **1975**, *ibid 79*, 2361. Shipman, L. L.; Burgess, A. W.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 543, 854.
(6) Jorgensen, W. L; Gao, J. *J. Am. Chem. Soc.* **1986**, *90*, 2174, and references therein.
(7) Clementi, E.; Cavallone, F.; Scordamaglia, R. *J. Am. Chem. Soc.* **1977**, *99*, 5531, 5545.
(8) Bolis, G.; Clementi, E. *J. Am. Chem. Soc.* **1977**, *99*, 5550.
(9) Fraga, S. *J. Comput. Chem.* **1982**, *3*, 329; *Comput. Phys. Commun.* **1983**, *29*, 351.
(10) Sordo, J. A.; Klobukowski, M.; Fraga, S. *J. Am. Chem. Soc.* **1985**, *107*, 7569.
(11) Romano, S.; Clementi, E. *Gazz. Chim. Ital.* **1978**, *108*, 319.
(12) Romano, S.; Clementi, E. *Int. J. Quantum Chem.* **1978**, *14*, 839; **1980**, *17*, 1007.
(13) Snir, J.; Nemenoff, R. A.; Scheraga, H. A. *J. Chem. Phys.* **1978**, *82*, 2498.
(14) Matsuoka, O.; Clementi, E.; Yoshimine, M. *J. Chem. Phys.* **1976**, *64*, 1351.
(15) Clementi, E.; Corongiu, G. *Int. J. Quantum Chem.: Quantum Biology Symposium* **1983**, *10*, 31.
(16) Detrich, J.; Corongiu, G.; Clementi, E.; *Chem. Phys. Lett.* **1984**, *112*, 426.
(17) Bolis, G.; Clementi, E.; Scheraga, H. A.; Wertz, H.; Tosi, C. *J. Am. Chem. Soc.* **1983**, *105*, 355.
(18) Venanzi, C. A.; Weinstein, H.; Corongiu, G.; Clementi, E. *Int. J. Quantum Chem.* **1982**, *9*, 355.
(19) Clementi, E.; Corongiu, G.; Jonsson, B.; Romano, S. *J. Chem. Phys.* **1980**, *72*, 260.
(20) Corongiu, G.f Fornili, S. L.; Clementi, E. *Int. J. Quantum Chem.* **1983**, *10*, 277.
(21) Ranghino, G.; Clementi, E.; Romano, S. *Biopolymers* **1983**, *22*, 1449.
(22) Ragazzi, M.; Ferro, D. R.; Clementi, E. *J. Chem. Phys.* **1979**, *70*, 1040.
(23) Bolis, G.; Clementi, E. *Chem. Phys. Lett.* **1981**, *82*, 147.
(24) Bolis, G.;Corongiu, G.; Clementi, E. *Chem. Phys. Lett.* **1982**, *86*, 299.

0002-7863/87/1509-1702$01.50/0 © 1987 American Chemical Society

an important point arises as a result of the greater complexity (degrees of freedom) of the molecules considered in this paper: the configurational space for these types of systems becomes very prolix. Therefore, since the reliability of an ab initio pair potential depends strongly on the completeness of the configurations sampled and selected for the fitting procedure, special consideration has been exercised in this work on this regard (see next section).

Method

The procedure for obtaining an ab initio pair potential may be divided into four steps: (a) selection of a functional form, (b) selection of conformations, (c) ab initio calculations, and (d) fitting procedure.

Analytical Pair Potential. As in our previous work[7,8] we used an analytical pair potential of the form

$$V_{MN} = \sum_i \sum_{i>j} -\left(A_{ij}^{ab}\right)^2/r_{ij}^6 + (B_{ij}^{ab})^2/r_{ij}^{12} + C_{ij}^{ab}q_iq_j/r_{ij} \quad (1)$$

where V_{MN} is the total interaction energy (kcal/mol) between molecules M and N, i is the ith atom in the molecule M belonging to the class a, j is the jth atom in molecule N belonging to the class b, and r_{ij} is the distance in Å between the two atoms i and j; and A_{ij}^{ab}, B_{ij}^{ab}, and C_{ij}^{ab} are the fitting parameters describing the interaction between atoms in class a and in class b. Actually, on the basis of our previous experience, the parameter C_{ij}^{ab} has been kept constant and equal to 1 in all cases.

q_i and q_j represent the charges associated to atoms i and j, respectively. These are not fitting parameters but they have a clear physical meaning. Mulliken population analysis (MPA) provides appropriate values for such charges. However, it is well-known that the values provided by MPA depend, for a given basis set, on the conformation considered. In order to observe the variation in the charges in different conformations, the corresponding MPAs have been computed for all the conformations considered in this work. By comparing the MPA for many different conformations, it was observed that the estimation of the charges q_i, q_j from the MPA on the monomers is a reasonable approximation to those computed in the pairs of amino acids.

In fact, in Table I we collect, as an example, the values of the charges q_i for all the atoms in alanine as calculated from the MPA on the monomer as well as from the MPA on the dimer, this last in a conformation in which the interaction energy was very strong (−16 kcal/mol), and therefore the differences between the two sets of MPAs are expected to be the greatest (of course, the sets become identical when there is no interaction between both amino acids). From Table I it can be concluded that taking the charges from MPA on the monomer is a good choice.

The different classes of atoms for the amino acids have been adopted from our previous work,[7,8] with the exception that the old classes 2 and 3 (see the first of the papers in ref 7) corresponding to different aliphatic hydrogens have been unified in the present work. In fact, hydrogens belonging to classes 2 and 3 exhibit a quite similar chemical behavior (as measured by means of the values of both partial charges and molecular orbital valency values).[7] This point was fully confirmed by carrying out two different fits. In the first one the classes were exactly as defined in our previous work,[7,8] and in the second one the same class was used to represent all the aliphatic hydrogens appearing in the amino acids. Both fits led practically to the same results. Table II collects the classes of atoms considered in this work.

Table II. Notation for the Classes of Atoms Considered in This Work

OXYGEN	
OCBL	carbonylic oxygen: R_2—C=O
OCBX	hydroxylic oxygen: R—COOH, R—OH
NITROGEN	
NNH2	aminic nitrogen: R—NH_2
CARBON	
CALP	α carbon: R_2—CH—NH_2
CCBX	carboxylic carbon: R—COOH
CCH3	aliphatic carbon: R—CH_3
CCH2	aliphatic carbon: R_2—CH_2
HYDROGEN	
HNH2	aminic hydrogen: R—NH_2
HCBX	hydroxylic hydrogen: R—COOH, R—OH
HALI	aliphatic hydrogen: R_3—CH, R_2—CH_2, R—CH_3

Selection of Conformations. As stated above, the selection of the conformations used in the fitting procedure is a very crucial point in this work. This has been pursued with two goals in mind. First, there are relevant chemical groups in the aliphatic amino acids not containing sulfur; these are the carboxylic and aminic groups (present in all these amino acids), and the alcohol group (present only in some of them). Some of the interactions involving these groups give rise to very stable conformations, for example, those in which the two carboxylic groups, one on each interacting molecule, form a double hydrogen bond, or those in which the aminic group in one molecule interacts with the OH in the carboxylic group of the other molecule, and so forth. Therefore, in order to ensure an adequate description of the local minima regions on the hypersurface, it is necessary to include such conformations into the fit. The generation of these special conformations has been performed by using graphics facilities. These facilities allow one to recover the final coordinates of two or more molecules once an appropriate set of movements (rotations and/or translations) have been performed (in order to achieve a given relative orientation of the molecules).

On the other hand, large portions of the configurational space still remain uncovered. In particular, there are many conformations in which the interaction between the above-mentioned chemical groups are not directly involved. Several procedures have been designed to systematically generate these other conformations. In such procedures one of the molecules is kept fixed with the center of mass at the origin, and then a sphere of radius r (centered at the origin) is constructed and a spherical grid formed by a set of regularly spaced points is defined on its surface. The initial radius, r, is increased step by step (increments of 0.5 Å, up to 4.5 Å, are taken) and equivalent spherical grids are defined on each one of the generated spherical surfaces. Then the center of mass of the second molecule is positioned on each one of the grid points and a rotation defined by three Eulerian angles is performed on each molecule. (This procedure is quite similar to that described in ref 10.) The starting radius, r, is chosen in such a way that the distance between the two nearest atoms (one belonging to the first molecule and the other belonging to the second molecule) is on the order of the length of a hydrogen bond ($\simeq$2.5 au). The three Eulerian angles are systematically chosen in order to account for the largest number of different relative orientations between both molecules.

Finally, further additional conformations are generated from feedback provided by the fitting procedure itself (see the subsection below on Fitting Procedure).

Following the above schemes a total of almost two thousand conformations were generated.

Ab Initio Calculations. The magnitude of the problem forced us to use standard 7/3(SZ) basis sets.[36] Indeed, we had to

(25) Fornili, S. L.; Vercauteren, D. P.; Clementi, E. *J. Biomolec. Struct. Dyn.* **1984**, *1*, 1281.
(26) Kim, K. S.; Vercauteren, D. P.; Welti, M.; Chin, S.; Clementi, E. *Biophys. J.* **1985**, *47*, 327.
(27) Carozzo, L.; Corongiu, G.; Petrongolo, C. *J. Chem. Phys.* **1978**, *68*, 787.
(28) Clementi, E.; Corongiu, G. *J. Chem. Phys.* **1980**, *72*, 3979.
(29) Clementi, E.; Corongiu, G. *Biopolymers* **1979**, *18*, 2431.
(30) Clementi, E.; Conongiu, G. *Int. J. Quantum Chem.* **1979**, *16*, 897.
(31) Clementi, E.; Corongiu, G. *Chem. Phys. Lett.* **1979**, *60*, 175.
(32) Clementi, E.; Corongiu, G. *Biopolymers* **1982**, *21*, 175.
(33) Clementi, E.; Corongiu, G. *Proceedings of the second SUNYA Conversation in the Discipline Biomolecular Stereodynamics*; Sharma, R. H., Ed.; Adenine Press: New York, 1982; Vol. I.
(34) Kim, K. S.; Clementi, E. *J. Am. Chem. Soc.* **1985**, *102*, 227.
(35) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
(36) Gianolio, L.; Pavani, R.; Clementi, E. *Gazz. Chim. Ital.* **1978**, *108*, 181.

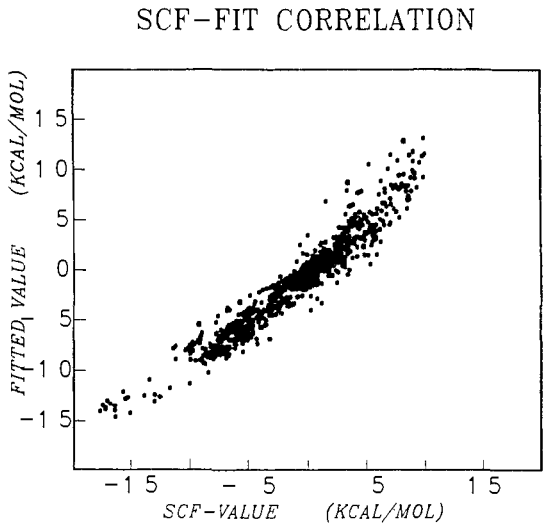

Figure 1. Correlation between interaction energies (corresponding to points included into the fitting) calculated via Hartree–Fock (with BSSE) and those derived from the parameters in Table III.

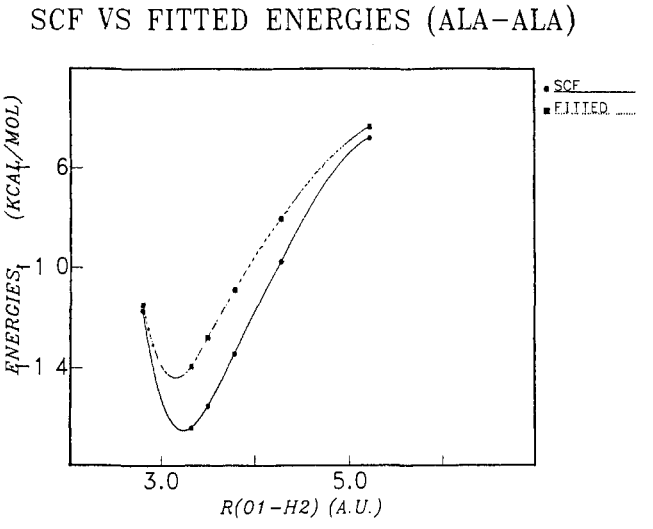

Figure 2. Potential curves from Hartree–Fock (kcal/mol) (with BSSE) (full line) and from parameters in Table III (dashed line) for alanine–alanine interacting through the carboxylic groups forming double-hydrogen-bonded associations. R(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

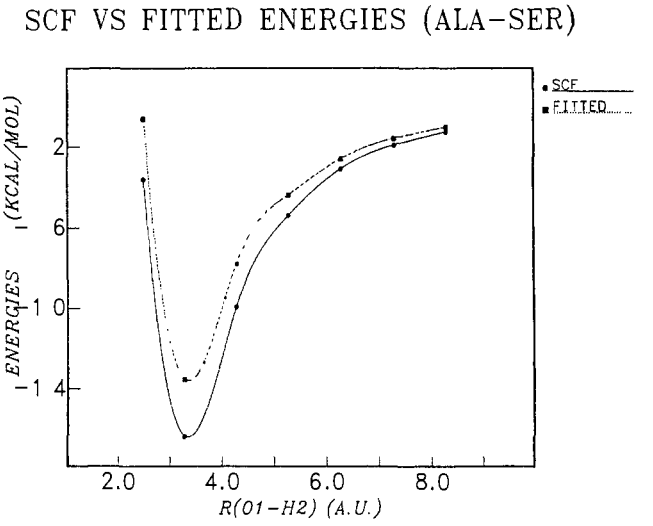

Figure 3. Potential curves from Hartree–Fock (kcal/mol) (with BSSE) (full line) and from parameters in Table III (dashed line) for alanine–serine interacting through the carboxylic groups forming double-hydrogen-bonded associations. R(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

perform almost two thousand SCF calculations at the Hartree–Fock level for systems with an average of more than 100 electrons, the typical average time being about 90 min of CPU time for each SCF calculation (integral evaluation, self-consistency and—see below—basis set superposition correction) on a FPS164. It is well-known that a minimal basis set gives errors due to the basis set superposition (BSSE). In order to minimize these errors the counterpoise (CP) method was applied to every SCF point. A discussion on the limitations of the 7/3(SZ) basis sets in constructing pair potentials can be found in the section Results and Discussion at the end of this paper.

Since the two amino acids alanine and serine contain all the classes of atoms collected in Table II, the pairs alanine–alanine, alanine–serine, and serine–serine were selected as the pairs on which the SCF calculations were performed. However, due to the chemical similarities among all amino acids these parameters should be transferable to other amino acids. As expected,[8] transferability of the parameters of the analytical pair potential (based on the concept of *class of atoms*[7]) was quite good. This is described below.

Fitting Procedure. The fitting procedure was carried out in several steps:

1. The interaction energies of the configurations generated according to the procedure described in a previous subsection (on Selection of Conformations) were fitted to the analytical pair potential in eq 1 and a set of parameters $\{A_{ij}^{ab},B_{ij}^{ab}\}$ was obtained (recall C_{ij}^{ab} is kept fixed and equal to 1.0).
2. With the above set of parameters the following procedure was used to calculate a number of isoenergy surfaces. One amino acid was kept fixed at the origin and the C^{α} of the second amino acid was moved to the intersection points of a two-dimensional grid. The size of such a grid was chosen to be 8 × 8 (au). For each grid point (a total of 1024 points were considered), we minimized the interaction energy with respect to the three rotational degrees of freedom of the second amino acid.
3. We searched for local minima of the energy in the isoenergy map. Some of such local minima conformations corresponded to those already used in the fitting procedure (those giving strong stabilization energies), but some others were new. SCF calculations for new local energy minima that had not previously been used in the fitting procedure were performed, and these points were included in the fitting procedure.
4. The fitting in (3) was carried out to convergence, giving the final parameters for the analytical pair potential.

Results and Discussion

Table III shows the values of the parameters of the analytical pair potential (see Table II for notation).

The standard deviation obtained from the fitting procedure was 1.08 kcal/mol; this is good considering that almost two thousand points have been fitted to a relatively simple functional form (eq 1). Figure 1 shows the correlation between the SCF calculations and the fitted values as calculated with eq 1 and the parameters in Table III. The regression coefficient was 0.93, and therefore the correlation may be classified as good.

Figures 2, 3, and 4 exhibit further evidence of the general reliability of our analytical pair potential. As described previously, some of the conformations were generated in order to describe those kinds of interactions between functional groups of special relevance in amino acids (COOH, NH_2, and OH). For some of these conformations a series of SCF points at varying intermolecular distances was computed (keeping the relative orientations of the two molecules fixed). The resulting potential energy curves may be compared with those provided by the analytical pair potential. Figures 2, 3, and 4 show the corresponding results for alanine and serine when both molecules interact through the carboxylic groups (see Figure 5, insets a to b). From both

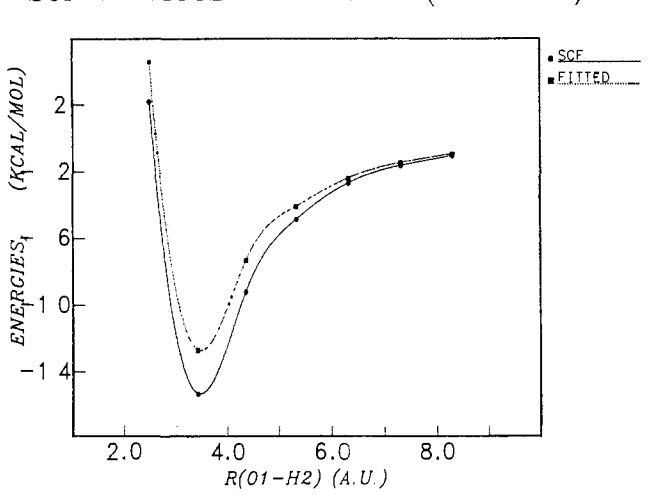

Figure 4. Potential curves from Hartree–Fock (kcal/mol) (with BSSE) (full line) and from parameters in Table III (dashed line) for serine–serine interacting through the carboxylic groups forming double-hydrogen-bonded associations. *R*(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

qualitative and quantitative points of view, the analytical pair potential provides a reasonably good description of such potential curves.

As a check on the reliability of the analytical pair potential a systematic search for local minima between pairs of amino acids was performed. From a chemical point of view it is easy to make an a priori prediction on the existence of a set of conformations in which a large stabilization should be expected because of the formation of hydrogen bonding. Samples of such conformations involving the interaction between COOH, NH_2, and OH groups for the pairs alanine–alanine, alanine–serine, and serine–serine were included into the fitting. The question which arises is whether or not the analytical pair potential is able to reproduce these types of interactions for other pairs of amino acids.

Figure 5 collects a representative sample of the local minima conformations for some of these other pairs of amino acids as obtained from the analytical pair potential proposed in this work. Indeed many of the conformations computed to be the most stable ones are those one might expect on the basis of chemical intuition. For all the pairs, the structures involving a double hydrogen bond (–COOH···HOOC–) between carboxylic groups in amino acids are the most stable ones (Figure 5, insets a to f). This agrees with the well-known fact of the dimeric structure observed in carboxylic acids.

However, the above-mentioned associations are not the only ones providing stable dimers. In fact, the high basicity of the aminic nitrogen allows for the possibility of very stable conformations involving hydrogen bonding between this nitrogen in one of the amino acids and the carboxylic hydrogen in the other-($-H_2N$···HOOC–). Conformations g and h (Figure 5) are examples of this kind of association and are predicted with the proposed pair potentials.

The existence of the alcohol group in some of the amino acids considerably increases the number of stable dimeric associations as shown in Figure 5 (insets i to l). Such associations are of two classes. On one hand are those in which the carboxylic group in one of the amino acids interacts with both aminic and alcohol groups in the other amino acid. These associations may also be classified as double-hydrogen-bonding associations, but they are not as strong as in the case of the associations involving two carboxylic groups (Figure 5, insets a to f) because of steric reasons. Thus, for example, in the case of the interaction between threonine and valine (see Figure 5, inset k), the spatial conformation of the hydrogen in the alcohol group of threonine prevents the formation of a double hydrogen bond involving the carboxylic group in valine and both aminic and alcohol groups in threonine. The interaction energy for this last type of association is on the order of −10 kcal/mol and should be compared with a value of about −14 kcal/mol corresponding to the interaction energy between the two carboxylic groups in a double-hydrogen-bonded association (Figure 5, insets a to j).

On the other hand, the interaction between the aminic and alcohol groups in one of the amino acids and the same groups in the other amino acid also gives rise to very stable conformations (Figure 5, inset l).

Therefore, Figure 5 shows that the analytical pair potential developed in this paper provides a good general description of the kinds of interactions with give rise to stable associations of pairs of amino acids, even for those pairs of amino acids not explicitly included into the fitting procedure. As an additional check on the transferability of our potential, SCF energies of 100 conformations in which at least one of the interacting amino acids was not one of those used in determining the parameters in Table

Table III. Parameters for the Analytical Pair Potential

classes		A^{ab}	B^{ab}
OCBL	OCBL	−0.001 799 460 1	449.235 381 266 0
OCBL	OCBX	−15.345 274 744 0	515.041 778 191 3
OCBX	OCBX	10.261 556 200 6	501.220 456 128 5
OCBL	NNH2	−0.177 726 715 6	652.336 932 089 9
OCBX	NNH2	0.005 498 848 2	585.890 844 871 3
NNH2	NNH2	0.395 454 129 4	1 950.126 018 635 9
OCBL	CCBX	−27.729 179 710 4	38.890 146 199 9
OCBX	CCBX	−0.055 403 175 9	272.899 545 203 3
NNH2	CCBX	0.314 399 390 9	−46.388 803 002 0
CCBX	CCBX	−0.010 563 110 0	−22.944 389 778 3
OCBL	CALP	−0.029 150 540 1	−661.244 978 845 9
OCBX	CALP	0.132 184 197 0	1 055.576 673 080 1
NNH2	CALP	0.340 914 367 0	1 184.310 493 966 6
CCBX	CALP	0.118 900 543 8	−3 837.195 654 376 4
CALP	CALP	0.079 620 506 9	−1 835.478 278 030 7
OCBL	CCH3	−0.005 441 829 2	−850.663 777 895 7
OCBX	CCH3	12.223 554 155 3	1 070.634 876 015 6
NNH2	CCH3	−0.141 557 349 7	1 693.936 740 266 6
CCBX	CCH3	0.070 935 811 3	−2 346.955 424 952 1
CALP	CCH3	−0.074 479 091 2	91.469 391 842 3
CCH3	CCH3	3.903 548 794 3	117.700 794 892 4
OCBL	HALI	0.010 010 493 6	58.134 799 917 1
OCBX	HALI	−0.005 735 569 1	37.113 378 369 6
NNH2	HALI	−7.614 766 263 0	−20.001 003 395 6
CCBX	HALI	−0.001 264 778 2	20.000 936 764 6
CALP	HALI	0.037 196 224 8	−140.816 763 212 5
CCH3	HALI	−0.043 511 600 3	254.781 125 123 6
HALI	HALI	−0.030 382 600 3	21.120 281 088 1
OCBL	HNH2	−0.033 345 321 6	85.529 497 451 3
OCBX	HNH2	0.000 355 946 0	20.001 000 189 0
NNH2	HNH2	19.697 011 690 7	20.000 999 980 4
CCBX	HNH2	−0.010 435 936 0	202.060 509 487 4
CALP	HNH2	0.035 053 111 6	289.872 395 700 1
CCH3	HNH2	−0.077 877 241 9	20.208 014 094 1
HALI	HNH2	0.003 626 365 3	22.391 090 214 2
HNH2	HNH2	−0.004 926 831 9	34.971 067 775 4
OCBL	HCBX	−7.639 262 099 2	20.001 000 058 4
OCBX	HCBX	−8.606 426 125 3	29.457 176 368 7
NNH2	HCBX	13.618 825 546 0	34.665 737 559 1
CCBX	HCBX	0.010 794 369 9	−285.519 789 786 7
CALP	HCBX	−0.093 292 054 7	197.874 008 996 1
CCH3	HCBX	3.062 174 707 3	−181.481 688 992 8
HALI	HCBX	−0.010 248 546 0	20.086 001 265 6
HNH2	HCBX	−0.008 103 837 6	−60.081 114 177 1
HCBX	HCBX	0.007 015 236 2	−59.518 851 602 5
OCBL	CCH2	0.280 673 051 2	−1 364.697 472 322 5
OCBX	CCH2	−0.065 899 877 4	1 253.498 257 774 4
NNH2	CCH2	0.035 118 173 7	−1 603.233 710 697 5
CCBX	CCH2	−0.553 225 087 1	−39.642 237 838 0
CALP	CCH2	−0.423 428 200 7	−2 927.372 817 832 2
CCH3	CCH2	−3.593 580 046 1	794.421 686 157 7
HALI	CCH2	0.250 995 768 1	21.624 935 402 8
HNH2	CCH2	−0.015 663 698 8	28.154 087 041 0
HCBX	CCH2	14.596 150 170 0	−23.116 800 677 8
CCH2	CCH2	0.011 061 349 0	−1 910.133 719 301 3

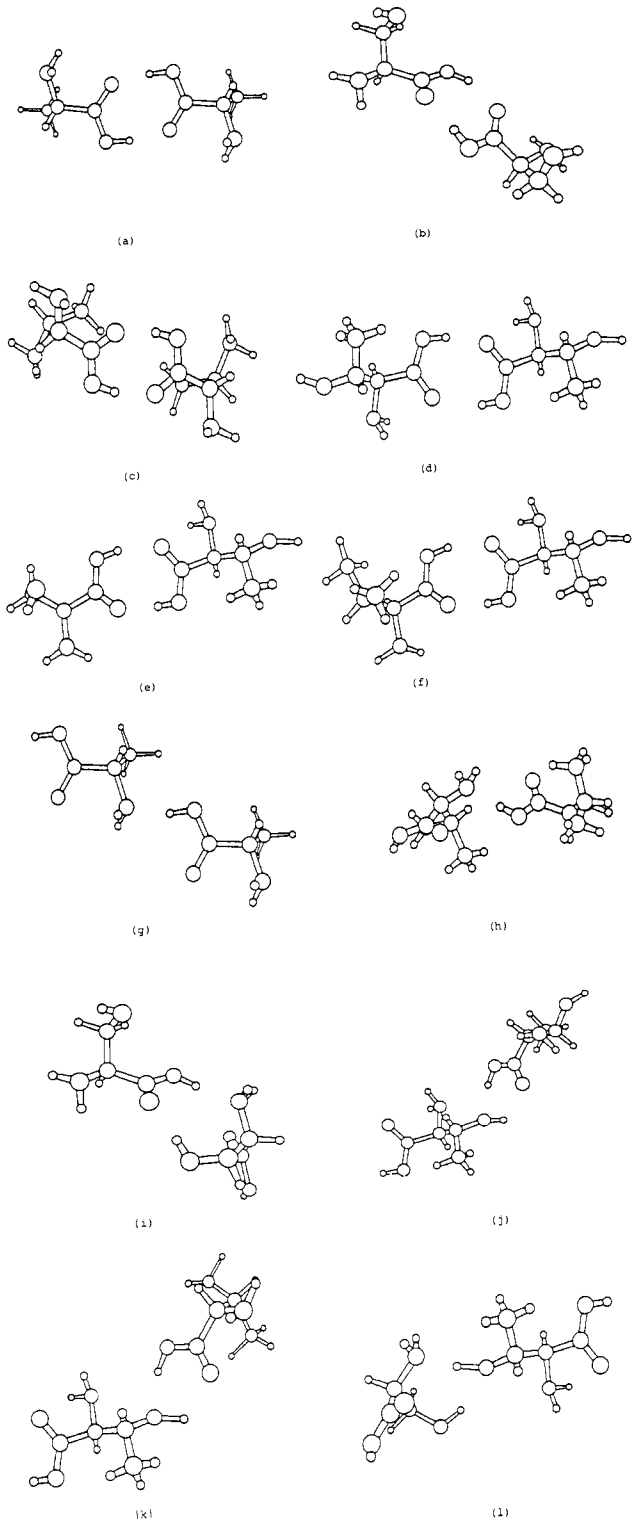

Figure 5. Conformations corresponding to local minima for several pairs of amino acids as obtained from parameters in Table III. Alanine–alanine (insets a, g), serine–serine (insets b, i), valine–valine (insets c, h), threonine–threonine (insets d, j), serine–threonine (insets e, l), valine–threonine (insets f, k).

III were computed. The correlation between the SCF values and the fitted values is shown in Figure 6. The correlation is quite good (the standard deviation for these additional calculations is 0.58 kcal/mol) and is indeed a strong support to the concept of *class of atoms* around which all of this research is being developed. In other words, our parameters reproduce the hypersurface of other chemically equivalent molecules not originally included in the derivation of these parameters. Therefore, the parameters appear to be quite transferable and reliable.

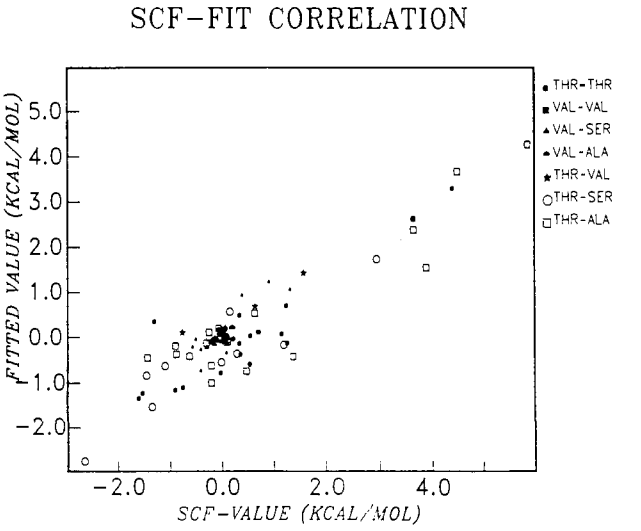

Figure 6. Correlation between interaction energies (corresponding to points not included into the fitting) calculated via Hartree–Fock (with BSSE) and those derived from the parameters in Table III.

Table IV. Total Energies, Monomer Energies (with CP correction to BSSE) (au), and Interaction Energies (kcal/mol) for Three Selected Geometries of the System Alanine–Alanine as Calculated with Different Basis Sets

STO-3G	ΔE	7/3(SZ)	ΔE	9/5(DZ)	ΔE
		Geometry 1			
−635.403 04	0.7	−641.384 08	−11.7	−643.360 98	−9.4
−317.702 01		−320.682 73		−321.673 06	
−317.702 18		−320.682 68		−321.673 00	
		Geometry 2			
−635.402 28	−6.1	−641.388 97	−16.4	−643.367 19	−15.5
−317.696 23		−320.681 42		−321.671 27	
−317.696 37		−320.681 40		−321.671 21	
		Geometry 3			
−635.399 46	−6.5	−641.386 77	−15.5	−643.366 42	−15.4
−317.694 47		−320.681 02		−321.670 92	
−317.694 62		−320.681 03		−321.670 92	

Basis Sets Limitations. As mentioned in a previous section, the calculations have been carried out at the SCF level with use of the 7/3(SZ) *minimal* basis sets. In this section we shall address questions concerning the accuracy (degree of confidence) of the results reported in this work.

In previous work aimed at developing analytical pair potentials, special care was exercised in examining the dependence of the results on the quality of the basis sets used.[17-19] The analysis of the results supported the use of 7/3(SZ)-type minimal basis sets in the derivation of pair potentials. In addition, subsequent applications of these pair potentials in studies of structural and energetic properties of several chemical and biochemical systems via MC and MD simulations have been reported in the last few years. The agreement found with available experimental data supports the use of 7/3(SZ) basis sets for constructing such potentials,[20-34] particularly in dealing with systems of such complexity as to make totally unfeasible the use of larger basis sets. However, this conclusion needs case by case validation.

In order to confirm the validity of the 7/3(SZ) basis sets in this study of the interactions between amino acids, SCF calculations for several hydrogen-bonded geometries for the systems alanine–alanine, alanine–serine, and serine–serine were performed with use of two other basis sets: the *subminimal* STO-3G basis sets developed by Pople et al.[35] and 9/5 basis sets of *double-ζ* quality.[36]

Tables IV, V, and VI collect the total energies for these associations, the monomer energies, and the interaction energies. In Figures 7, 8, and 9 we plot the corresponding potential curves

Table V. Total Energies, Monomer Energies (with CP correction to BSSE) (au), and Interaction Energies (kcal/mol) for Three Selected Geometries of the System Alanine–Serine as Calculated with Different Basis Sets

STO-3G	ΔE	7/3(SZ)	ΔE	9/5(DZ)	ΔE
		Geometry 4			
−709.212 27	7.2	−715.991 57	−3.6	−718.157 45	−3.0
−391.521 49		−395.302 54		−396.479 63	
−317.702 17		−320.683 31		−321.673 06	
		Geometry 5			
−709.220 83	−6.9	−716.007 32	−16.4	−718.172 83	−16.2
−391.513 30		−395.299 78		−396.476 06	
−317.696 55		−320.681 38		−321.670 92	
		Geometry 6			
−709.211 22	−5.4	−715.991 99	−9.9	−718.162 91	−11.6
−391.508 19		−395.297 02		−321.474 77	
−317.694 44		−320.679 17		−321.669 66	

Table VI. Total Energies, Monomer Energies (with CP correction to BSSE) (au), and Interaction Energies (kcal/mol) for Three Selected Geometries of the System Serine–Serine as Calculated with Different Basis Sets

STO-3G	ΔE	7/3(SZ)	ΔE	9/5(DZ)	ΔE
		Geometry 7			
−783.029 74	11.8	−790.602 46	2.2	−792.954 58	2.4
−391.522 93		−395.303 35		−396.479 81	
−391.525 60		−395.302 66		−396.478 63	
		Geometry 8			
−783.039 20	−6.4	−790.624 13	−15.3	−792.976 17	−15.1
−391.513 95		−395.299 87		−396.476 07	
−391.515 02		−395.299 85		−396.476 07	
		Geometry 9			
−783.025 09	−5.0	−790.609 16	−9.2	−792.967 08	−10.8
−391.508 39		−395.297 18		−396.474 92	
−391.508 79		−395.297 37		−396.474 91	

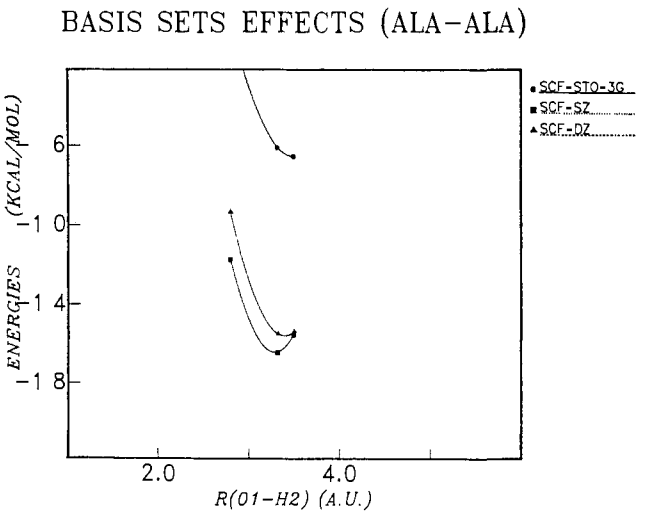

Figure 7. Potential curves from Hartree–Fock (kcal/mol) as calculated with different basis sets for alanine–alanine interacting through the carboxylic groups forming double-hydrogen-bonded associations. *R*(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

as calculated with use of STO-3G, 7/3(SZ), and 9/5(DZ) basis sets.

It is clear from Figures 7–9 that 7/3(SZ) and 9/5(DZ) basis sets provide comparable interaction energies from both qualitative and quantitative points of view. The STO-3G basis sets show substantial deviations, even if they are able to provide some qualitative information (i.e., except for Figure 7, the relative order of the interaction energies remains the same as the order calculated with 9/5(DZ) basis sets). However, quantitative accuracy was

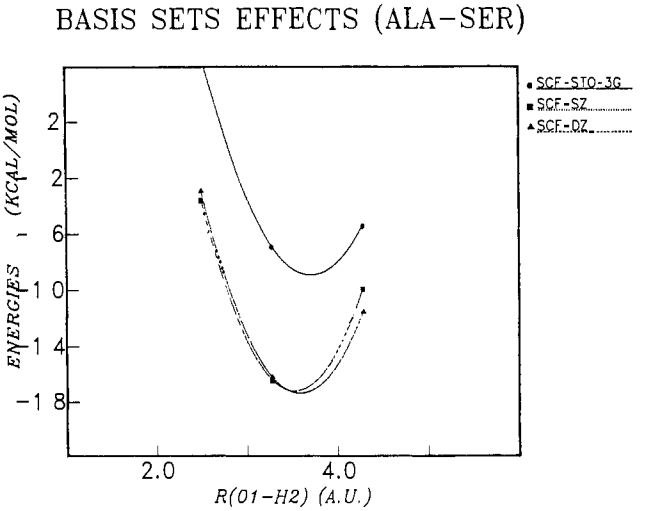

Figure 8. Potential curves from Hartree–Fock (kcal/mol) as calculated with different basis sets for alanine–serine interacting through the carboxylic groups forming double-hydrogen-bonded associations. *R*(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

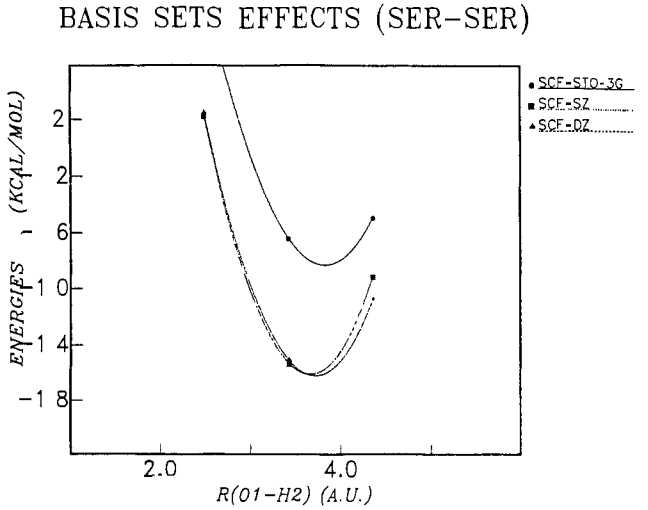

Figure 9. Potential curves from Hartree–Fock (kcal/mol) as calculated with different basis sets for serine–serine interacting through the carboxylic groups forming double-hydrogen-bonded associations. *R*(O1–H2) represents the distance between the carbonyl oxygen in one of the molecules and the hydrogen belonging to the OH group in the second molecule.

Table VII. Atomic Energies (au) for the Basis Sets Discussed in This Work

element	STO-3G	7/3(SZ)	9/5(DZ)	HF
H	−0.466 58	−0.499 28	−0.499 94	−0.5
C	−37.198 39	−37.615 99	−37.685 19	−37.688 619
N	−53.719 01	−54.284 41	−54.395 35	−54.400 934
O	−73.804 15	−74.627 84	−74.800 40	−74.809 398

not achieved with the STO-3G bases. We note that this is what might be expected by examining the atomic energies (see Table VII) for the different basis sets (HF energies have been included for comparison). The improvement of the 7/3(SZ) *minimal* basis sets over the STO-3G *subminimal* basis sets is very clear. As is known, this is a direct consequence of the care exercised in balancing and optimizing (the contraction coefficients were variationally optimized) the 7/3(SZ) basis sets[36] (in general, the 7/3(SZ) basis sets are somewhat better than the Slater single-ζ and not too different from the Slater double-ζ basis sets[17,36]). Molecular calculations reported in Tables IV–VI and Figures 7–9 fully corroborate this point.

All the above clearly confirms previous findings concerning the usefulness of the 7/3(SZ) basis sets in creating a library of analytical pair potentials for associations involving molecules of biological interest.

Before finishing this section, let us make some comments on the CP method used in this work to correct for the error arising from the superposition of the basis sets.

Doubts have been reported in the literature concerning the use of the CP method to correct for the BSSE.[37,38] However, other recent literature also gives some examples in which the use of the CP method results in much improved interaction energies when calculated with small basis sets.[39-41] Therefore, at this moment, a definitive conclusion about the validity of the use of the CP method cannot be determined.[42] Furthermore, most of the above studies[37-41] have been carried out on small systems, and we feel that extrapolation to larger systems is not straightforward. In fact, on the basis of the origin of the BSSE it seems reasonable to infer that as long as the number of atoms *directly involved* in the molecular association grows, an increase in the BSSE should be expected.

The above discussion combined with previous experience in using 7/3(SZ) basis sets (see, for example, ref 17 and 43 for a discussion concerning this point) have been the reasons to include the CP correction in the present work.

Conclusions

By performing an extensive number of SCF calculations at the Hartree–Fock level, an ab initio analytical pair potential for the computation of the interaction energies between aliphatic amino acids (without sulfur) has been derived. Because of the complexity of the potential hypersurfaces for the interaction between pairs of amino acids, special attention has been focused on the generation of the conformations from which the parameters of the analytical pair potential have been obtained.

Since parameters for the interactions between amino acids and water, as well as for the interaction between water molecules, are already available in the literature, the present-day potential library provides a powerful tool for dealing with a wide variety of problems involving molecules of biological interest.[44] In particular, amino acids are the units from which proteins are built-up. In this regard, the determination of the 3-D structure of proteins is a most important topic that could be tackled by using the analytical pair potential developed in this paper. Some preliminary research is being conducted in order to test the appropriateness of our intermolecular pair potential for dealing with structural problems in proteins involving intramolecular interactions.

Acknowledgment. One of us (J.A.S.) thanks the Ministerio de Educacion y Ciencia (Spain) for a fellowship as well as IBM-Kingston for its very generous hospitality and support.

Registry No. Alanine, 56-41-7; serine, 56-45-1.

(37) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Shaefer, H. F., III *J. Chem. Phys.* **1986**, *84*, 2279.
(38) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.
(39) Greenwood, H. H.; Plant, J. S. *Int. J. Quantum Chem.* **1986**, *30*, 127.
(40) Szczesniak, M. M.; Scheiner, S. *J. Chem. Phys.* **1986**, *84*, 6328.
(41) Bonaccorsi, R.; Cammi, R.; Tomasi, J. *Int. J. Quantum Chem.* **1986**, *29*, 373.
(42) For an extended review on the basis sets we refer to the very recent work by Davidson and Feller: Davidson, E. R.; Feller, D. *Chem. Rev.* **1986**, *86*, 681.
(43) Kolos, W. *Theor. Chim. Acta (Berlin)* **1979**, *51*, 219.
(44) Clementi, E. *J. Phys. Chem.* **1985**, *89*, 4426.

Brønsted Basicity of Atomic Oxygen on the Au(110) Surface: Reactions with Methanol, Acetylene, Water, and Ethylene

Duane A. Outka and R. J. Madix*

Contribution from the Departments of Chemical Engineering and Chemistry, Stanford University, Stanford, California 94305. *Received May 1, 1986*

Abstract: The adsorption and reactions of methanol, acetylene, water, and ethylene were investigated on clean and oxidized Au(110) surfaces by temperature-programmed reaction spectroscopy. All of these molecules are only weakly and molecularly adsorbed on the clean Au(110) surface. Methanol, acetylene, and water, however, react with the oxidized surface. Methanol, activated by 0.25 monolayer of oxygen adatoms, reacts to form water, methyl formate, hydrogen, and carbon dioxide. A stable methoxy intermediate is identified in these reactions. Acetylene reacts to form water and carbon dioxide, and water is more strongly bonded to the Au(110) surface in the presence of oxygen adatoms. Ethylene is the only one of these molecules which does not react with oxygen adatoms on Au(110). This pattern of reactivity parallels that associated with the acidity of these molecules as measured in the gas phase which has been observed on Cu(110) and Ag(110) surfaces. These results complete the studies necessary to demonstrate the Brønsted base character of oxygen adatoms on all of the group 1B metals.

1. Introduction

Atomic oxygen activates copper and silver surfaces toward reaction with a variety of molecules. One of the principal mechanisms for this activation is a Brønsted base reactivity where atomic oxygen abstracts an acidic hydrogen atom from an adsorbed molecule. Among the molecules for which this reaction occurs are formic acid,[1-5] methanol,[6-9] acetylene,[10,11] ethanol,[12] water,[13,14] acetic acid,[15] hydrogen sulfide,[16] hydrogen chloride,[16] and propylene.[17] In the absence of oxygen adatoms, some of these molecules are completely unreactive, particularly on silver surfaces. Despite the diverse functional groups of the molecules involved, the reactivity of oxygen toward these molecules on copper or silver

(1) Sexton, B. A. *Surf. Sci.* **1979**, *88*, 319.
(2) Ying, D. H. S.; Madix, R. J. *J. Catal.* **1980**, *61*, 48.
(3) Barteau, M. A.; Bowker, M.; Madix, R. J. *Surf. Sci.* **1980**, *94*, 303.
(4) Bowker, M.; Madix, R. J. *Surf. Sci.* **1981**, *102*, 542.
(5) Sexton, B. A.; Madix, R. J. *Surf. Sci.* **1981**, *105*, 177.
(6) Wachs, I. E.; Madix. R. J. *J. Catal.* **1978**, *53*, 208.
(7) Wachs, I. E.; Madix. R. J. *Surf. Sci.* **1978**, *76*, 531.
(8) Felter, T. E.; Weinberg, W. H.; Lastushkina, G. Y.; Zhdau, P. A.; Boneskov, G. K.; Hrbek, J. *Appl. Surf. Sci.* **1983**, *16*, 351.
(9) Bowker, M.; Madix, R. J. *Surf. Sci.* **1980**, *95*, 190.
(10) Barteau, M. A.; Madix. R. J. *Surf. Sci.* **1982**, *115*, 355.
(11) Outka, D. A.; Friend, C. M.; Jorgensen, S.; Madix. R. J. *J. Am. Chem. Soc.* **1983**, *105*, 3468.
(12) Wachs, I. E.; Madix. R. J. *Appl. Surf. Sci.* **1978**, *1*, 303.
(13) Bowker, M.; Barteau, M. A.; Madix. R. J. *Surf. Sci.* **1980**, *92*, 528.
(14) Spitzer, A.; Lüth, H. *Surf. Sci.* **1982**, *120*, 376.
(15) Bowker, M.; Madix, R. J. *Appl. Surf. Sci.* **1981**, *8*, 299.
(16) Moroney, L.; Rassias, S.; Roberts, M. W. *Surf. Sci.* **1981**, *105*, L249.
(17) Barteau, M. A.; Madix, R. J. *J. Am. Chem. Soc.* **1983**, *105*, 344.

0002-7863/87/1509-1708$01.50/0 © 1987 American Chemical Society