Theoret. chim. Acta (Berl.) 20, 31—40 (1971)

An Iterative Extended Hückel Calculation for Some Amino Acids Containing Sulphur and Selenium

GIULIANO COLOMBETTI* AND CARLO PETRONGOLO

Laboratorio di Chimica Quantistica ed Energetica Molecolare del C.N.R., Via Risorgimento 35, 56100 Pisa, Italy**

Received April 21, 1970/July 22, 1970

Semiempirical iterative extended Hückel calculations are performed for the dipolar ions of cystine, cysteine, methionine and their selenium analogues. An attempt has been made to interprete the wave function of cysteine in terms of localized molecular orbitals.

On a analisé les ions dipolaires de la cystine, cystéine, methionine et leurs analogues séléniates avec la méthode semiempirique de Hückel étendue itérative. On a chérché à intérpreter la fonction d'onde pour la cystéine dans le schème des orbitales moleculaires localisées.

Semiempirische iterative Rechnungen wurden nach der erweiterten Hückeltheorie für die dipolaren Ionen Cystin, Cystein, Methionin und ihre Selen-Homologen durchgeführt. Für den Fall des Cysteins wurde die Wellenfunktion auf lokalisierte Orbitale transformiert.

1. Introduction

At the present time, the lack of quantum calculations on the amino acids (mainly due to the vast amount of necessary numerical computations) do not make a strict description of the electronic structure of these compounds. Del Re, Pulman, and Yonezawa [1] have extensively examined the charge distribution and the proton chemical shift of 25 α-amino acids in their neutral, dipolar, cationic and anionic forms, by using a simple method proposed by Del Re [2]. In 1969 the extended Hückel method was used by some authors [3, 4] to study the conformations of peptides. Their results have encouraged us to examine the electronic features of the dipolar ions of the cystine, cysteine, methionine and their selenium analogues by using the iterative extended Hückel (IEH) method [5, 6]. In addition, for the cysteine we made an analysis in terms of localized molecular orbitals strictly derived from the delocalized molecular orbitals. We obtained an approximate electronic picture which, even if very crude, conforms to chemical intuition and is in reasonable agreement with the more sophisticated *a priori* calculations on simpler molecules.

2. Outline of the Method

In the IEH framework utilized here the diagonal matrix elements of the effective one-electron hamiltonian are evaluated as function of $q(\alpha)$, the atomic

---

* Laboratorio per lo studio delle proprietà fisiche di Biomolecole e Cellule, Via Valdagno 9, 56100 Pisa, Italy.

** Work performed with the financial support of the Consiglio Nazionale delle Ricerche.

charge on the atom α, according to the expression

$$-H_{ii}(\alpha) = \mathrm{VOIP}_i(\alpha) = A_i(\alpha)\cdot q^2(\alpha) + B_i(\alpha)\cdot q(\alpha) + C_i(\alpha) \tag{1}$$

where $H_{ii}(\alpha)$ is the Coulomb integral between the atomic orbitals χ_i on the atom α, $\mathrm{VOIP}_i(\alpha)$ is the ionization potential of an electron on the valence orbital χ_i on the atom α, and the constants $A_i(\alpha)$, $B_i(\alpha)$ and $C_i(\alpha)$ are taken from Basch [7]. The off-diagonal matrix elements are approximated by the Wolfsberg-Helmholz formula

$$H_{ij} = \tfrac{1}{2} k (H_{ii} + H_{jj}) S_{ij} \tag{2}$$

where S_{ij} is the overlap integral between the atomic orbitals χ_i and χ_j, and k is assumed to be 1.75 as proposed by Hoffmann. The **H**-matrix, calculated via (2), transforms properly under a rotation of the coordinates. An iterative procedure for the evaluation of Coulomb integrals was performed according to the following expression:

$$H_{ii}^{(n)}(\alpha) = H_{ii}^{(n-1)}(\alpha)\cdot(1-\lambda) - \lambda\,[A_i(\alpha)\cdot q^2(\alpha)^{(n-1)} + B_i(\alpha)\cdot q(\alpha)^{(n-1)} + C_i(\alpha)] \tag{3}$$

where the overscript (n) and $(n-1)$ refer respectively to n^{th} and $(n-1)^{\text{th}}$ iteration, and the λ parameter is assumed to be 0.1, as proposed by Carroll [6]. Input data are the $A_i(\alpha)$, $B_i(\alpha)$ and $C_i(\alpha)$ constants and a set of atomic charges; the iterative process is repeated until convergence of the atomic charges is achieved. All the valence electrons are considered, and the atomic basis functions are Slater type orbitals whose exponents are those proposed by Clementi [8] (best atom ζ), save the hydrogen $1s$ function whose exponent is 1.2, a more appropriate value for hydrogen atom in molecule, than 1.0. The overlap and dipole one-electron integrals are exactly calculated for all the atomic orbitals by using a program of C. Guidotti and O. Salvetti of the Institute of Chemical Physics of Pisa.

3. Results of IEH Calculations

From the cristallographic structure of the dipolar ions of the compounds containing sulphur [9–11], the atomic cartesian coordinates were deduced with respect to an orthogonal left-handed reference axes system. Because of the lack of experimental crystallographic data on selenium amino acids, we were forced to assume for these compounds the same geometry as their sulphur analogues. Fig. 1 shows the atom numbering convention.

Some results of our calculations, performed to the self-consistent charge configuration, are reported in Table 1. The iterative process is performed until $[q(\alpha)^{(n)} - q(\alpha)^{(n-1)}] < 0.005$ for all the atoms; the convergency is reached in 10–12 iterations. The first column of the table shows the sum over the valence electrons of the **H**-matrix eigenvalues. Except for cystine and Se-cystine, this sum decreases when going from sulphur to selenium compounds. In the second and third columns are respectively reported the calculated and some experimental values of ionization potentials (IP). As is seen by inspection of these last colums, the agreement between our calculated values and experimental IP is sufficiently good; there is also, as expected, an IP decrease when going from sulphur to selenium amino acids. These agreements are essentially due to the

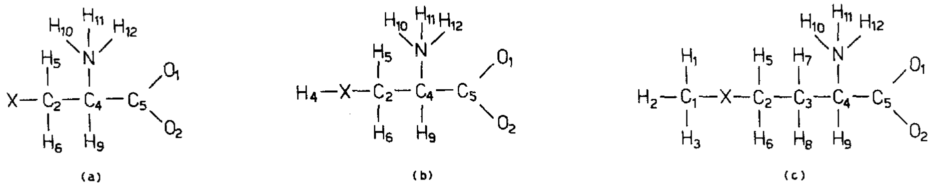


Fig. 1. Atoms numbering convention: a) cystine (Se-cystine) (only half molecule) – b) cysteine (Se-cysteine) – c) methionine (Se-methionine)

iterative procedure. As a matter of fact, as a result of Hoffmann [5] calculations and of ours, the IP in the non iterative method, are about two eV greater than the observed values. The last column of Table 1 report the expectation values of the dipole moments. We have calculated these values, instead of in the point-charge approximation generally used in the semiempirical methods, by means of the expectation value of a one-electron observable in the MO-LCAO framework:

$$\begin{aligned}\langle \mu \rangle_{\text{el}} &= \langle \Psi | \, e \sum_{i}^{\text{el}} r_i \, | \Psi \rangle = 2 \sum_{i}^{\text{MO}} \langle \phi_i | \, e r_i \, | \phi_i \rangle \\ &= 2 \sum_{i}^{\text{MO}} \sum_{r}^{\text{AO}} c_{ir}^{+} \sum_{s}^{\text{AO}} c_{si} \langle \chi_r | \, e r \, | \chi_s \rangle \end{aligned} \tag{4}$$

because it was possible for us to calculate exactly the dipole atomic integrals of (4).

The experimental values of this observable were estimated for the acqueous solutions of the dipolar ions under the hypothesis of a free rotation about all the valence bonds [12, 13]. From data of dielectric constants in water, Kulkani Jatkar [12] furnish a value of 12.5 D for all the α amino acids. From the work of Greenstein [13], a value of approximately 12.8 and 13.3 D for the dipolar ions of cysteine and methionine respectively can be deduced from the dielectric constants increment per mole of solute in water at 25° C, and a value of 15.4 D for both these compounds if a free rotation about all the valence bonds is assumed. Although the calculation refers to the gaseous state of the amino acids and their

Table 1. *Total electronic energies* $\left(2\sum_{i}^{\text{MO}} \varepsilon_i\right)$, *first ionization potentials and expectation values of the dipole moments*

Molecule	$2\sum_{i}^{\text{MO}} \varepsilon_i{}^{a}$	IP^{b}	$\text{IP}_{\text{exp}}{}^{b}$	μ^{d}
Cystine	—51.6402	9.08		2.86
Se-Cystine	—51.5093	8.85		3.33
Cysteine	—26.5085	8.89		13.58
Se-Cysteine	—26.5570	8.87		13.48
Methionine	—33.8910	8.61	8.9 (8.63)^{c}	12.85
Se-Methionine	—33.9362	8.59	8.4 (8.29)^{c}	12.98

[a] Atomic units.
[b] Electron Volt units.
[c] Svee, H. J., Junk, G. A.: J. Amer. chem. Soc. **89**, 790 (1968).
[d] Debye units.

geometry has been deduced from crystallographic data, our calculated dipole moments are not too different from the above quoted references for cysteine, methionine and their selenium analogues. The situation is totally different in the cases of cystine and Se-cystine. For these compounds the calculated dipole moments are much lower than that estimated by Greenstein (24 D). This feature is essentially due to the difference in the structure of cystine (and Se-cystine) in its crystal state and in its acqueous solution. In the first case the molecule is more symmetric than in water and the free rotation about the valence bonds is forbidden and, therefore, the statistical treatment of Eyring, which supports the data of Greenstein, is not applicable.

To save space, we do not report here the values of the coefficients and of the eigenvalues of the molecular orbitals, however some remarkable features are present: unoccupied molecular orbitals also have negative eigenvalues and the eigenvalues of the amino acids containing sulphur are very close to their selenium analogues; at last many eigenvectors of cystine and Se-cystine are degenerate. The first characteristic is peculiar to the extended Hückel method; this hypothesis is confirmed by inspection of several other authors results [6, 14, 15]. To be noted, is that in all the molecules the first filled MO has its major atomic contributions on the C_4 and N, the second filled MO is mainly an $NC_5O_1O_2$ type, and all the last three filled MO are extended on the oxigen atoms. Finally, the lowest empty MO is mainly a $C_5O_1O_2$ type. In cystine and Se-cystine, for each of these eigenvalues, are present two degenerate eigenvectors each extended on both half molecules.

Mulliken Population Analysis

In Tables 2, 3, and 4 some results of a Mulliken population analysis are reported. To be noted are the following:

a) The main disagreements between the gross atomic charges calculated by us (Table 2) and those calculated by Del Re [1] are: (i) a greater charge on the nitrogen and α-carbon atom; (ii) a lower charge on the hydrogen atoms of the NH_3 group; (iii) a negative charge on the carbon atom of the COO group. According to these results, besides the charge transfer from the amino group to the carboxil group (Table 3), which leads to the formal charge distribution $(\mathrm{NH}_3^+)-\overset{|}{\underset{|}{\mathrm{C}}}-(\mathrm{COO})^-$ for the dipolar ions of the amino acids, there is also a charge transfer from other atoms, mainly of the α-carbon atom C_4, to the carboxil group where the total negative charge is also shared by the carbon atom C_5.

b) A negative charge is present except in the Se-cystine molecule, on all the sulphur and selenium atoms, as the *a priori* calculations for molecules containing sulphur illustrate [16, 17].

c) All the hydrogen atoms are positively charged, which charge increases in the order: hydrogen bonded to primary, secondary, tertiary carbon and to nitrogen.

d) Examination of the *s* and *p* gross atomic orbital populations (Table 4) reveals a net promotion out of *s* into *p* atomic orbitals, due to their hybridization. The order of the $s \to p$ promotion is, quite naturally, carbon > nitrogen > oxigen ≃ sulphur (or selenium).

Table 2. *Gross atomic charges*

	Cystine[a]	Se-Cystine[a]	Cysteine	Se-Cysteine	Methionine	Se-Methionine
X	−0.0089	0.0298	−0.0279	−0.0055	−0.0525	−0.0207
C_1					−0.0294	−0.0349
C_2	0.0272	0.0182	0.0304	0.0236	−0.0050	−0.0090
C_3					0.0169	0.0144
C_4	0.1195	0.1127	0.1271	0.1232	0.1149	0.1129
C_5	−0.0301	−0.0291	−0.0228	−0.0250	−0.0138	−0.0148
N	0.1819	0.1766	0.2020	0.2000	0.1941	0.1934
O_1	−0.4683	−0.4669	−0.4592	−0.4610	−0.4634	−0.4642
O_2	−0.4666	−0.4653	−0.4708	−0.4725	−0.4846	−0.4854
H_1					0.0211	0.0191
H_2					0.0207	0.0180
H_3					0.0212	0.0187
H_4			0.0020	0.0095		
H_5	0.0429	0.0383	0.0423	0.0390	0.0279	0.0262
H_6	0.0455	0.0405	0.0425	0.0390	0.0262	0.0245
H_7					0.0361	0.0350
H_8					0.0350	0.0339
H_9	0.0564	0.0519	0.0565	0.0544	0.0633	0.0626
H_{10}	0.1682	0.1658	0.1594	0.1587	0.1572	0.1569
H_{11}	0.1566	0.1540	0.1595	0.1589	0.1574	0.1571
H_{12}	0.1760	0.1735	0.1590	0.1579	0.1568	0.1565

[a] Only half molecule.

Table 3. *Total group charges for the* NH_3, C_4 *and* COO

	Cystine	Se-Cystine	Cysteine	Se-Cysteine	Methionine	Se-Methionine
NH_3	0.6827	0.6699	0.6799	0.6755	0.6655	0.6639
NH_3 and C_4	0.8022	0.7826	0.8070	0.7987	0.7804	0.7768
COO	−0.9650	—0.9643	−0.9528	−0.9585	−0.9618	−0.9644

Table 4. *Total gross* AO *populations (s and p)*

	Cystine[a]		Se-Cystine[a]		Cysteine		Se-Cysteine		Methionine		Se-Methionine	
	s	*p*	*s*	*p*	*s*	*p*	*s*	*p*	*s*	*p*	*s*	*p*
X	1.6649	4.3441	1.6520	4.3183	1.6210	4.4070	1.6443	4.3613	1.6215	4.4310	1.6447	4.3760
C_1									1.1228	2.9066	1.1076	2.9273
C_2	1.1311	2.8417	1.1184	2.8634	1.1397	2.8299	1.1291	2.8473	1.1545	2.8505	1.1395	2.8695
C_3									1.1263	2.8568	1.1279	2.8577
C_4	1.1657	2.7147	1.1684	2.7189	1.1558	2.7171	1.1574	2.7194	1.1458	2.7393	1.1458	2.7412
C_5	1.1654	2.8647	1.1657	2.8635	1.1600	2.8629	1.1609	2.8642	1.1293	2.8845	1.1297	2.8850
N	1.2907	3.5273	1.2899	3.5365	1.3085	3.4895	1.3087	3.4913	1.3104	3.4954	1.3106	3.4960
O_1	1.6158	4.8524	1.6178	4.8491	1.6082	4.8510	1.6084	4.8526	1.5502	4.9132	1.5504	4.9139
O_2	1.6078	4.8588	1.6097	4.8556	1.6255	4.8453	1.6257	4.8469	1.5847	4.8999	1.5849	4.9006

[a] Only half molecule.

4. Localization of the Molecular Orbitals

As is well known, the molecular orbitals which derive from a MOLCAO method are delocalized over all the atoms of a given molecule and, therefore, do not have an immediate chemical meaning. Since the total wave function is invariant with respect to a unitary transformation of the molecular orbitals, we transformed the MO, localizing them by using the intrinsic Boys method [18] which maximizes the sum of the square distances of the MO charge centroids.

In the present paper we have only analyzed the electronic structure of the cysteine molecule obtained by this same localization method. A schematic representation of the localized molecular orbitals (LMO) of the cysteine is reported in Fig. 2. The LMO can be divided as lone-pair and bond orbitals, respectively localized on one and two atoms. Some interesting features are present, in good accord with chemical evidence: the two sulphur lone-pair orbitals are essentially tetrahedral *sp*^{3} hybrids with the charge centroids symmetrically placed with respect to the H_4SC_2 plane; in the NH_3 group the symmetry group C_{3v} of the ammonia molecule is preserved, i.e. the charge centroids of the three equivalent NH bond orbitals and the nitrogen nucleus are piramid-like with a triangular equilateral base; the electronic structure of the carboxil group is represented by four CO banana bonds, whose charge centroids are symmetrically placed with respect to the COO plane, and two lone-pair orbitals on each oxigen atom, whose charge centroid are in the COO plane and symmetrically placed with respect to the CO axis.

The localization degree of a molecular orbital can be evaluated by the overlap integral $\langle\lambda|\chi\rangle$ between the LMO λ and its normalized meaningful part[1]. With these values the goodness of the localization procedure is confirmed. As a matter of fact for all the lone-pair orbitals, the overlaps' values are greater or equal than 0.99, and for other orbitals are greater than 0.96 (the smaller overlaps are those of the CC and CO bond orbitals).

In Table 5 are reported the parallel and perpendicular components, with respect to the internuclear axis, of the localized orbitalic dipole moments [17] (the first column shows the symbols of the various LMO: for example S is a lone-pair orbital on the sulphur atom and SC_2 a bond orbital between the sulphur and carbon atoms). The electronic contributions are calculated by the atomic dipole integrals (see formula 4). For the lone-pair moments and the perpendicular components the absolute value are reported.

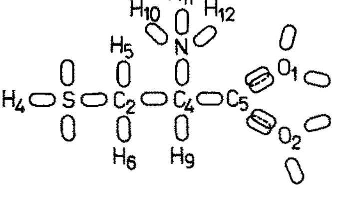


Fig. 2. Schematic representation of the localized orbitals of the cysteine

---

[1] For greater details see Ref. [20].

Table 5. *Cysteine molecule: localized orbitalic dipole moments*

λ	$\mu_{\parallel}{}^{a}$	$\mu_{\perp}{}^{a}$
S	4.53	—
S′	4.55	—
O_4	2.85	—
O'_1	2.81	—
O_2	2.87	—
O'_2	2.84	—
SC_2	−0.10	0.09
C_2C_4	−0.26	0.53
C_4C_5	0.28	0.20
C_4N	−1.50	0.04
C_5O_1	−1.17	2.82
$C_5O'_1$	−1.18	2.79
C_5O_2	−1.29	2.75
$C_5O'_2$	−1.23	2.78
SH_4	−2.49	0.
C_2H_5	−2.31	0.35
C_2H_6	−2.32	0.37
C_4H_9	−2.25	0.05
NH_{10}	−1.30	0.
NH_{11}	−1.31	0.
NH_{12}	−1.30	0.

[a] Debye units.

From a similar analysis one can deduce a more strict picture of the electronic distribution than from the population analysis. As a matter of fact, the parallel and perpendicular components of an ODM give a measure of the polarity and of the bending of the bond respectively; a minus sign for $\mu_{\parallel}$ represents a moment vector toward the first atom of the bond, i.e. the centroid of the electronic distribution is nearest of the second atom.

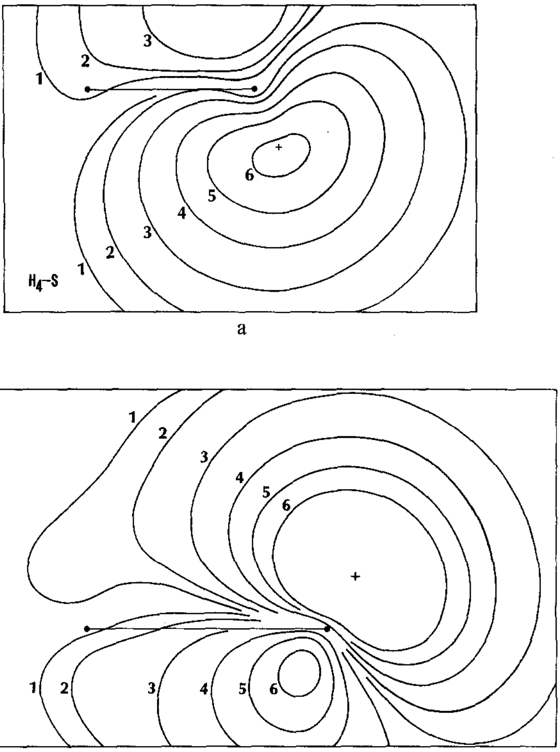
The absolute value of the parallel component for the CX bonds ($X \neq H$) increases with the electronegativity of the X atom, and its direction furnishes a charge centroid shift which is in good agreement with the electronic charge transfer obtained from a Mulliken population analysis. For a direct comparison, in the first column of Table 6 are reported the directions and values of the electronic charge transfers with respect to a state of isolated atoms and of N^+ and C^- ions, and in the following columns are reported the charge centroids shifts and the values of the parallel components of the ODM. An inversion is present in the SC_2 bond orbital, but in this case the small values of the electronic charge transfer and of $\mu_{\parallel}$ do not permit a significative comparison. The values of $\mu_{\parallel}$ for the LMO of the cysteine are in reasonable agreement with the results of *a priori* calculations on simpler molecules [17, 19, 20], save for the CH and NH bonds (and probably for the SH bond), which $\mu_{\parallel}$ are overestimated in the IEH method. This fact may give a measure of the transferability of the relevant part of a LMO, in spite of our semiempirical wave function.

Inspection of the values of the perpendicular component reveals that the greater degree of bending is present, as is natural, on the CO banana bonds. The other bond orbitals are linear or only sligthly bended.

Table 6. *Electronic charge transfer (from the Mulliken population analysis), charge centroids shifts and parallel component of* ODM *(from the* LMO *analysis)*

Electronic charge transfer		Charge centroid shift	$\lvert\mu_\parallel\rvert^a$
$C_5 \rightarrow O_2$	0.4699	$C_5 \rightarrow O_2$	1.29
$C_5 \rightarrow O_1$	0.4592	$C_5 \rightarrow O_1$	1.17
$C_4 \rightarrow N$	0.3201	$C_4 \rightarrow N$	1.50
$C_2 \rightarrow C_4$	0.0893	$C_2 \rightarrow C_4$	0.26
$C_5 \rightarrow C_4$	0.0472	$C_5 \rightarrow C_4$	0.28
$C_2 \rightarrow S$	0.0259	$S \rightarrow C_2$	0.10

[a] Debye units.

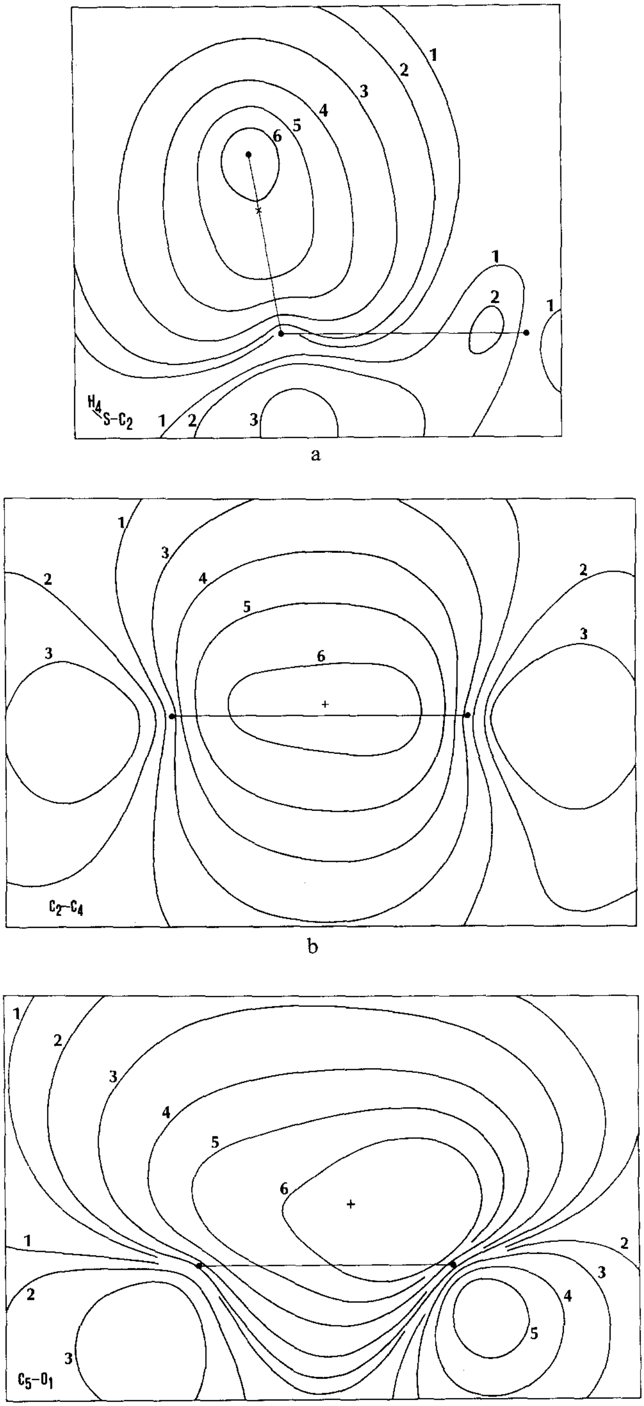


a

b

Fig. 3a and b. Electronic density maps for some localized lone-pair orbitals of cysteine (in the plane containing the charge centroids): a) Sulphur lone-pair – b) oxigen lone-pair. The nuclei and the charge centroids are marked by a dot and a cross respectively. The contour values are as follows: 1 = 0.0005; 2 = 0.002; 3 = 0.01; 4 = 0.04; 5 = 0.1; 6 = 0.2

In order to provide a more visual picture of some LMO, in Fig. 3 and in Fig. 4 are reported the electronic density maps for some localized lone-pair and bond orbitals of cysteine. The bending of C_2C_4 and C_5O_1 bond orbitals is shown by the shift of the charge centroids from the nuclear axis; and



a

b

c

Fig. 4a–c. Electronic density maps for some localized bond orbitals of cysteine: a) SH_4 bond orbital – b) C_2C_4 bond orbital – c) C_5O_1 banana bond. See Fig. 3 for other specifications the shift of the electronic density to H_4 and O_1 atom respectively, is in agreement with the calculated values of the parallel component of the ODM.

From results quoted above it may be deduced that an analysis executed in terms of localized orbitals, as well as that conducted in terms of delocalized orbitals and populations, can provide useful informations on the electronic structure of molecular systems. Such a proceeding gives not only a chemically meaningful picture of a molecule, but also, since the total molecular wave function is invariant with regard to an unitary transformation, it allows for an analysis in chemically meaningful orbitalic terms of some observables, which, such as the dipole moments, are important for understanding electronic structure. When such a proceeding is applied to a semiempirical molecular wave function, the results may represent a useful test of the goodness of the semiempirical method used; in our case the results obtained were encouraging, being in reasonable agreement with the chemical intuition and with the results of *a priori* calculations on other simpler molecules.

It is our intention to extend this study to other amino acids in order to analyse more deeply the different electronic properties of sulphur and selenium molecules as well as the transferability of the localized orbitals from one molecule to another.

Acknowledgements. The authors wish to thank Prof. Eolo Scrocco and Prof. Jacopo Tomasi for useful discussions.

References

1. Del Re, G., Pullmann, B., Yonezawa, T.: Biochim. biophysica Acta **75**, 153 (1963).
2. – J. chem. Soc. (London) **1968**, 4031.
3. Hoffmann, R., Imamura, A.: Biopolymers **7**, 207 (1969).
4. Kier, L. R., George, J. M.: Theoret. chim. Acta (Berl.) **14**, 258 (1969).
5. Hoffmann, R.: J. chem. Physics **39**, 1397 (1963).
6. Carroll, D. G., Armstrong, A. T., McGlynn, S. P.: J. chem. Physics **44**, 1865 (1966).
7. Basch, H., Viste, A., Gray, H. B.: Theoret. chim. Acta (Berl.) **3**, 458 (1965).
8. Clementi, E., Raimondi, D. L.: J. chem. Physics **38**, 2686 (1963).
9. Oughton, B. M., Harrison, P. M.: Acta crystallogr. **12**, 396 (1959).
10. Harding, M. M., Long, H. A.: Acta crystallogr. B **24**, 1096 (1968).
11. Wyckoff, R. W. G.: Crystal Structures, Vol. 5, p. 682, London: Interscience Pub. 1966.
12. Jatkar, S. K. K., Iyengar, B. R. Y.: J. Indian Inst. Sci. **31** A, 15 (1949).
13. Greenstein, J. P., Winitz, P.: Chemistry of the Amino Acids, p. 456–472, New York: John Wiley and Sons Inc. 1961.
14. Carroll, D. G., Vanquickenborne, L. G., McGlynn, S. P.: J. chem. Physics **45**, 2777 (1966).
15. Allen, L. C., Russel, J. D.: J. chem. Physics **46**, 1029 (1967).
16. Boer, F. P., Lipscomb, W. N.: J. chem. Physics **50**, 989 (1969).
17. Bonaccorsi, R., Scrocco, E., Tomasi, J.: J. chem. Physics **52**, 5270 (1970).
18. Foster, J. M., Boys, S. F.: Rev. mod. Physics **32**, 300 (1960).
19. Pritchard, R. H., Kern, C. W.: J. Amer. chem. Soc. **91**, 1631 (1969).
20. Bonaccorsi, R., Petrongolo, C., Scrocco, E., Tomasi, J.: J. chem. Physics **48**, 1500 (1968).

Dr. C. Petrongolo
Laboratorio di Chimica Quantistica
ed Energetica Moleculare del C. N. R.
Via Risorgimento 35,
I-56100 Pisa, Italia