Theoret. chim. Acta (Berl.) 24, 51—60 (1972)

The Molecular Electrostatic Potentials for the Nucleic Acid Bases: Adenine, Thymine, and Cytosine

R. BONACCORSI

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

A. PULLMAN

Institut de Biologie Physico-Chimique, 13, rue P. et M. Curie, Paris Vè, France

E. SCROCCO and J. TOMASI

Istituto di Chimica Fisica dell' Università, 56100 Pisa, Italy

Received April 29, 1971

The electrostatic potentials arising from *ab initio* MO LCAO GTO SCF wave functions for adenine, thymine and cytosine are given and discussed.

Well defined characteristic regions of immediate chemical significance are found. The analysis of such results aims at comparing different protonation sites in the same molecule as well as in different ones. Differences among the proton affinities of the nitrogen atoms (pyridine-like, amine and imine) are evidenced, as well as the distinction between oxygen and nitrogen atoms.

Les potentiels electrostatiques moléculaires calculés à partir de fonctions d'onde *ab initio* MO SCF LCAO GTO sont donnés et discutés pour l'adénine, la thymine et la cytosine.

Des régions bien définies sont obtenues dont la signification chimique apparait clairement. L'analyse de l'ensemble des résultats permet la comparaison des differents sites de protonation dans une même molécule ainsi que dans differentes molécules, ainsi que la distinction entre oxygene et azote d'une part, azotes de differents types d'autre part.

Die mit Hilfe von *ab initio* MO LCAO GTO SCF Wellenfunktionen berechneten elektrostatischen Potentiale von Adenin, Thymin und Cytosin werden angegeben und diskutiert.

Man findet genau bestimmte charakteristische Zonen von klarer chemischer Bedeutung. Durch die Analyse solcher Resultate können verschiedene Protonierungsplätze sowohl in ein und demselben Molekül als auch in verschiedenen Molekülen verglichen werden. Deutlich zeigen sich Unterschiede der Protonenaffinitäten der unterschiedlich chemisch gebundenen Stickstoffatome (pyridinartiger, Amin- und Iminstickstoff) und der Unterschied zwischen Stickstoff- und Sauerstoffatomen.

Introduction

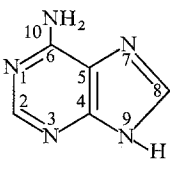
It has been suggested recently [1, 2], that the information enclosed in a molecular wave function can be used in the form of the electrostatic potential created in the neighbouring space by the nuclear charges and the electronic distribution. For a given wave function with the corresponding first order density function ϱ (1) the average value of such a potential $V(r_i)$ at a given point i of space is [2]:

$$V(r_i) = -\int \frac{\varrho(1)}{r_{1i}} \mathrm{d}\tau_1 + \sum_{\alpha}^{\mathrm{nucl}} \frac{Z_\alpha}{r_{\alpha i}}, \tag{1}$$

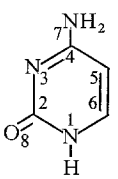
where Z_α is the nuclear charge of nucleus α.

This quantity, like the electron density distribution, is directly obtainable from the wave function and thus does not suffer from the drawbacks inherent to the classical population analysis. Moreover, its very definition seems to make it an appropriate index for the study of chemical reactivity at least in the early phase of approach of a given reagent for reactions with an ionic mechanism. In particular, the interaction energy between the molecular distribution (considered unperturbed) and an external point charge q placed at point i is $qV(r_i)$. It may be worth pointing out that $qV(r_i)$ is rigorously the first order perturbation energy of the molecule in the field of the charge q.

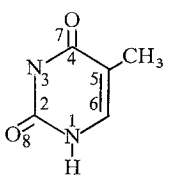
Maps of electrostatic interaction energies with a unit positive charge have already been drawn for the molecules of cyclopropane, aziridine, oxirane, thiirane, diazirine and cyclopropene [2], for oxaziridine and diaziridine [3] as well as for N_2, NOF, NH_3, H_2O and H_2S [4] using *ab initio* SCF molecular wave functions of "Slater minimal basis set" accuracy. In the present paper we report on similar curves drawn for adenine, I, cytosine, II, and thymine, III, using the *ab initio* SCF molecular wave functions obtained previously [5] in a small gaussian basis set. We have not treated guanine because its wave function was not available in the same basis set.



I



II



III

The particular interest of this type of molecules for this kind of study lies in the fact that each of them presents more than one potential site for protonation or electrophilic attack, thus providing a good test as to the possibilities of isopotential curves to distinguish among different positions of attack, or among different molecules. The basicity of such heterocycles has been essentially dealt with, in the past, in the π-electron approximation where it was shown that the sole consideration of the local charge of the atoms involved was not sufficient to distinguish among the different sites [6]. We shall come back in the discussion to this point and to the comparison of the isopotential results with other indices.

Computation Details

Within the LCAO MO framework, the potential $V(r_i)$ can be expressed in terms of the P_{rs} elements of the charge-bond order matrix relating to the atomic orbitals χ_r and χ_s as:

$$V(r_i) = -\sum_{r,s} P_{rs} \int \frac{\chi_r(1)\,\chi_s(1)}{r_{i1}} \mathrm{d}\tau_1 + \sum_{\alpha} \frac{Z_\alpha}{r_{\alpha i}}. \tag{2}$$

A new version of the program for calculating $V(r_i)$, which had been previously written [2] for Slater orbitals, has been coded for Gaussian basis functions. This

new version uses only the MO coefficients as input and relies on the calculation of nuclear attraction type integrals. The interaction energy with a hypothetical positive charge was calculated at regular intervals in the outer molecular space for the three molecules and iso-energy curves have been drawn accordingly.

Results and Discussion

Isopotential curves drawn in the molecular plane for the three molecules considered are given in Figs. 1, 2, and 3. Isopotential maps in selected planes perpendicular to the molecular one are reported in Figs. 4 and 5 for adenine, Fig. 6 for cytosine and Fig. 7 for thymine.

Adenine: In the molecular plane three regions of attraction appear towards the three pyridine-like nitrogens with distinct minima of -78.3 (N_3), -77.0 (N_1), and -69.4 (N_7) kcal/mole respectively. The regions around N_1 and N_3 are very similar in shape and depth, whereas the N_7-region is narrower and less deep. These regions clearly correspond to the lone pairs of the nitrogens which were apparent in the isodensity diagrams obtained from the molecular wave function [7]. They are similar to those found for the nitrogens of diazirine [2] (the numerical values are not to be strictly compared on account of the differences in the atomic basis sets).

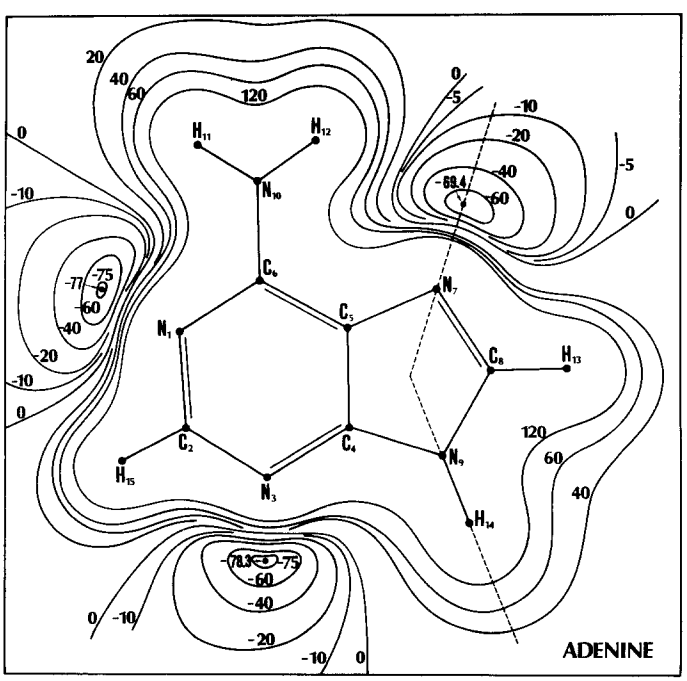


Fig. 1. Electrostatic molecular potential-energy map for adenine in the ring plane. The energies are expressed in kcal/mole

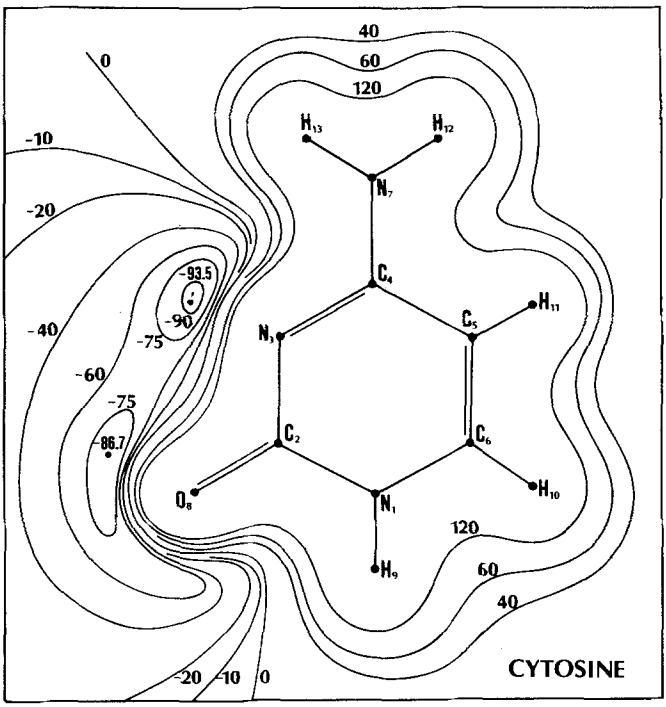


Fig. 2. Electrostatic molecular potential-energy map for cytosine in the ring plane

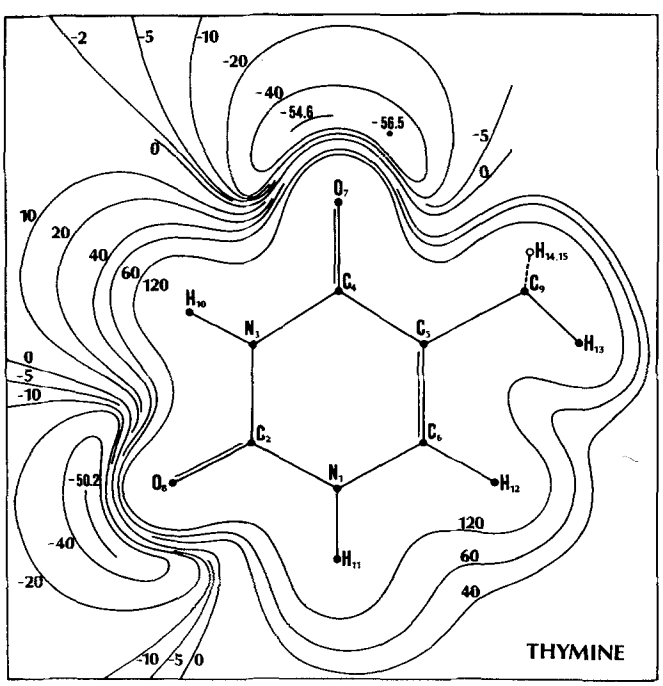


Fig. 3. Electrostatic molecular potential-energy map for thymine in the ring plane

The NH and CH regions as well as the neighbourhood of the amino group appear repulsive. Fig. 4 is instructive in this connection: it shows the potential energy above the plane of the molecule in a perpendicular section through N_{10}, C_6, and N_3. A secondary minimum of -24.8 kcal/mole is located above the amino nitrogen. Worth noting is the fact that a wide attractive area covers the whole $N_{10}C_6N_3$ region until a rather small distance. The shape of this attractive region and the relative depths of the minima around N_{10} and N_3 indicate clearly that the basicity (or nucleophilic character) of a pyridine-like nitrogen is larger than that of an NH_2-group, when the two are present, a conclusion in very good agreement with well-known experimental facts [8, 9]. Fig. 5 refers to a section

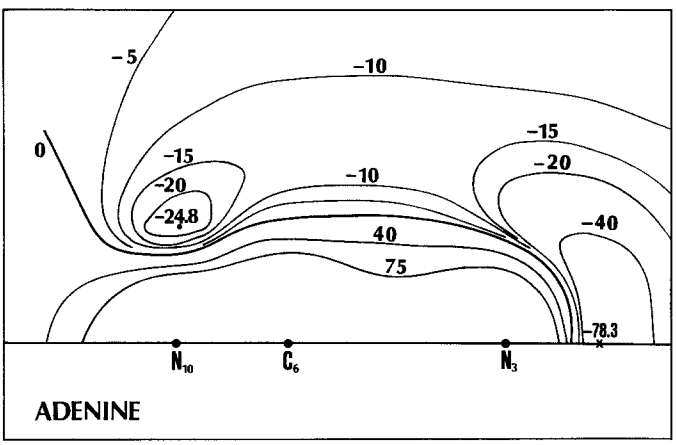


Fig. 4. Potential-energy map for adenine in the plane perpendicular to the ring plane and passing through N_{10}, C_6, and N_3 atoms

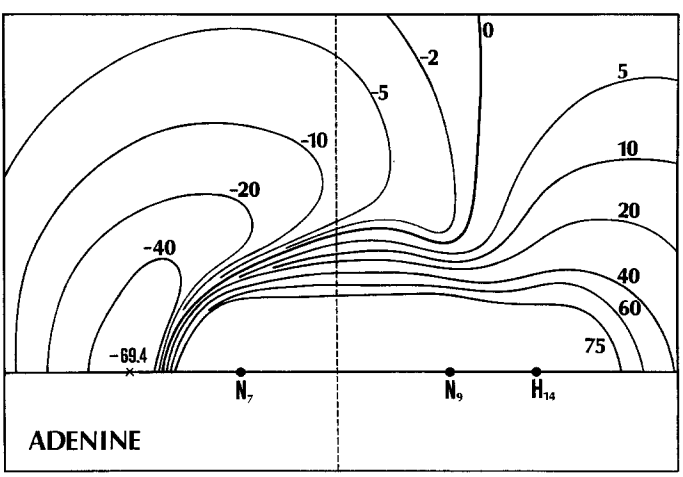


Fig. 5. Potential-energy map for adenine in a section perpendicular to the ring plane, containing N_7, N_9, and H_{14} atoms, and having as intersection the dotted line shown in Fig. 1

perpendicular to the molecular plane, containing atoms H_{14}, N_9, and N_7, and having as intersection with the molecular plane the dotted line shown in Fig. 1. A wide negative area which reaches the N_7 atom probably spreads out from the N_7 lone pair region (as in the case of N_3 in Fig. 4), but it reaches only the vertical of N_9, clearly showing the difference between amino and imino nitrogens.

As to the comparison of the present theoretical results with experimental data concerning electrophilic attack on adenine itself, one may mention that protonation, as well as alkylation, involves the ring nitrogens. Among them, the preferred regions of attack are definitely N_1 and N_3, the distinction among the two positions being difficult: results on the base itself indicate N_1 as the first protonation site [10, 11] but N_3 was found more reactive than N_1 towards alkylating agents [12].

In nucleotides, nucleosides, and in the nucleic acids, N_1 seems to be the first protonation site [13], alkylation occurring at both N_1 and N_3 [14]. In RNA, N_1 is the principal minor site of alkylation (the major site is N_7 of guanine) while this site shifts to N_3 in DNA (where N_1 is involved in hydrogen-bonding) [15].

Cytosine: The in-plane diagram of cytosine (Fig. 2) presents a new situation due to the presence of a carbonyl group adjacent to a pyridine-like nitrogen: a wide attractive region for electrophilic agents appears on this side of the molecule, with two deep minima, one located in the direction of the molecule, with two deep minima, one located in the direction of the nitrogen lone-pair, the other at an angle of 55° from the C=O bond. This last location is quite interesting in view of the fact that the *global* isodensity curves show no directional properties around the carbonyls [7]. In fact there is a slight anisotropy of the electron density hidden inside the roughly spherical cloud of oxygen, the SCF canonical orbitals comprising one nearly pure oxygen *p* atomic orbital perpendicular to the CO bond, and another (roughly *sp*) hybrid along the CO bond, each occupied by a pair of electrons [5]. If one transforms this canonical set into a set of localized orbitals according to the criterion of Foster and Boys [16], one obtains two oxygen lone-pair orbitals pointing approximately at 120° from each other on each side of the CO bond [17]. In the molecules previously studied [1, 2, 4], the minima of the potentials were located approximately in the direction of the localized lone-pair orbitals: in the present case the lone-pair directionality around the carbonyl oxygen is clearly overwhelmed by the effect of the neighbouring groups, the strong attractive character of the neighbour nitrogen enhancing the oxygen minimum on one side, the repulsive character of the NH group being strongly felt on the other side. The strong effect of the environment appears also in the diagram of thymine (*vide infra*).

An intermediate situation may be found in the CH_2NHO molecule [3] where two minima on the oxygen site are still present but noticeably influenced by the attractive character of the nitrogen lone-pair and by the repulsive effect of the NH bond.

A rather remarkable feature concerning cytosine is the fact that the potential well is deeper for N_3 than for O_8. Note also that the minimum for N_3 is deeper than for either nitrogen in adenine.

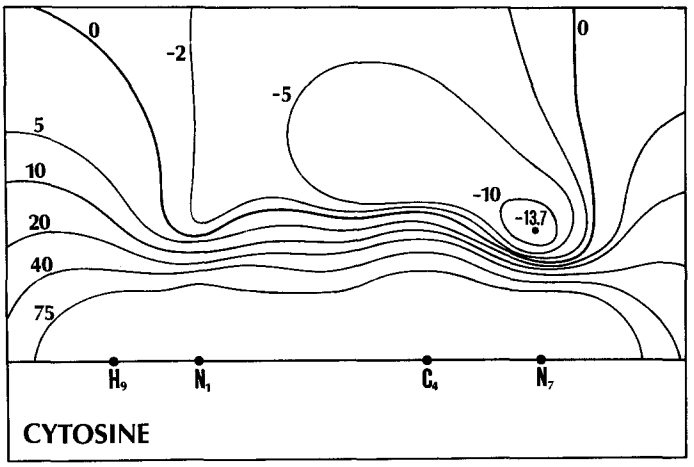


Fig. 6. Potential-energy map for cytosine in the plane perpendicular to the ring plane and passing through H_9, N_1, C_4, and N_7

The amino nitrogen N_7 has two equivalent minima of -13.7 kcal/mole above and below the molecular plane while the electrostatic potential around the imino nitrogen N_1 is very similar to the one found in the corresponding N_9H zone of adenine (see Fig. 6).

From the experimental point of view, protonation of cytosine, its nucleoside and nucleotide, occurs on N_3 [10, 18], and so does alkylation [19, 15], both reactions being easier in cytosine than in adenine. The basic pK's are 4.6 and 4.2 respectively [9, 18, 11]. In DNA, where the N_3 position is involved in hydrogen-bonding the reactivity of cytosine towards methylation is much reduced or even suppressed depending on the reagent [15]. A protonation of the oxygen of cytosine in DNA with no disruption of hydrogen bonding has been reported [20].

All these experimental facts are in very satisfactory agreement with the conclusions that can be drawn from the potential energy curves.

Thymine: The in-plane isopotential curves of thymine are given in Fig. 3. This molecule has no-pyridine-like nitrogens, but has two carbonyl oxygens one of which is adjacent to one NH group on each side, the other has one NH neighbour and one CCH_3 on the other side. The shape of the potentials shows two regions of attraction towards the two oxygen atoms, the rest of the molecule being repulsive. As noted in the case of cytosine, the directionality of the attraction is strongly influenced by the environment: the oxygen O_8 which has two NH neighbours presents one symmetrical potential well, whereas two minima appear near O_7, one being clearly pushed away from the NH region.

The isopotential curves in the perpendicular plane passing through H_{11}, N_1, C_4, and O_7 (Fig. 7) show that the imino group is surrounded in this molecule by a repulsive area, although immediately above the N atom a decrease of the $V(r_i)$ values may be evidenced.

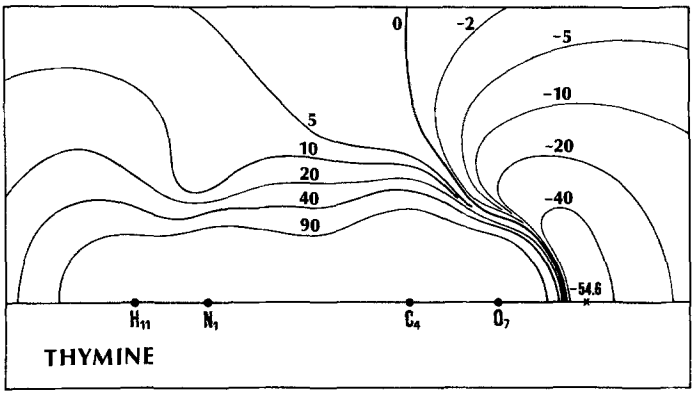


Fig. 7. Potential-energy map for thymine in the plane perpendicular to the ring plane and passing through H_{11}, N_1, C_4, and O_7

The depth of the minima in thymine is much smaller than that of the corresponding wells in cytosine, in complete agreement with the fact that thymine and thymidine are much less basic than cytosine and cytidine [15] and that neutral thymine does not undergo alkylation in the conditions under which the other bases react [11, 12]; the same is true in the nucleic acids [15, 17].

General Remarks and Conclusion

The present extension of the study of isopotential curves to large heterocycles clearly confirms the interest and usefulness of this technique for extracting from a wave function the information it encloses in a form which appears suitable for the investigation of problems of chemical reactivity.

Clear-cut differences appear between the different zones of complex molecules and the positions and depths of the potential wells are undoubtedly connected with the ease of electrophilic attack. Even though the electrostatic energy considered here is but a fraction of the total energy of interaction of the molecule with an approaching proton, it clearly appears as a determining element. Although polarization and charge transfer effects are neglected in this representation, it seems – in the case studied – as if they had their maxima in the same directions as those given by the consideration of the electrostatic energy alone.

Quite interesting is the fact that the present technique allows a neat discrimination between the different possible sites of protonation on a same molecular skeleton, an achievement which is not possible on simple inspection of the isodensity contours. The reason for this is that potential energies are very sensitive to small differences in electron densities, differences which are not perceptible in the global density contours generally utilized[1].

[1] It is perhaps not superfluous to remark that a variation in density of $6 \cdot 10^{-3}$ electron per $(a_0)^3$ is sufficient to produce a variation of 1 kcal/mole in the interaction energy with a unit positive charge situated at a distance of 4 atomic units.

Another advantage of the isopotential curves is the possibility of comparing potential protonation sites of different kinds (oxygen with nitrogen, pyridine-like nitrogen with amino nitrogens, etc. ...) and also sites on different molecules. A noticeable result of the present study is the fact that doubly-bonded oxygen atoms appear less basic than doubly-bonded nitrogens. This is probably related to the greater directionality of the charge cloud around the nitrogens.

If we consider the nitrogen atoms, we can remark that

a) pyridine-like atoms have the greatest proton affinity; the region of negative potential extends above and below the molecular plane but the minimum is in the plane.

b) The amino nitrogens have a remarkable proton affinity in the regions outside the molecular plane (corresponding to potential minima of -24 kcal/mole for adenine and -13.7 kcal/mole for cytosine).

c) Finally, the imino atoms seem to have the lowest proton affinity; the isopotential maps do not show minima directly related to this type of atoms, although from Figs. 5, 6, and 7 one is allowed to infer a small negative contribution due to their π electrons.

Finally, one seemingly important effect appears in the study of the electrostatic potential seen by an approaching proton, namely the repulsive role of the peripheral hydrogen atoms attached to the molecule. These hydrogens are always more or less discharged with respect to a neutral atom (especially when bound to nitrogen in an amino group or in NH), so that the effect of the nuclei overbalances the attraction of the electronic density. This creates large repulsive regions which should help directing the approaching proton into the corridors leading to the strongly attractive zones. This effect of the peripheral hydrogens has never been taken into account in studies of basicities, and might well be quite important. The effect is perhaps exaggerated in the present results where the wave functions utilized correspond to a rather strong depopulation of the hydrogen atoms. But in any event, the qualitative effect would remain (similar effects can be observed in the small molecules of Ref. [2] and [3]).

In conclusion, we think that pursuing the study of reactivity along these lines is worth further efforts. The use of *ab initio* wave functions (and if possible even of higher accuracy than the present ones) is of course of prime importance at the beginning, but recent investigations using the CNDO-method have proven extremely encouraging [21] and a close parallel study with *ab initio* results will probably lead to a satisfactory calibration of semi-empirical procedures.

References

1. Bonaccorsi, R., Petrongolo, C., Scrocco, E., Tomasi, J.: In: Quantum aspects of heterocyclic compounds in chemistry and biochemistry, Jerusalem Symposia Vol. II, 181 (1970).
2. — Scrocco, E., Tomasi, J.: J. chem. Physics **52**, 5270 (1970).
3. — — — Theoret. chim. Acta (Berl.) **21**, 17 (1971).
4. — — — Theoretical section progress report, Laboratorio di Chimica quantistica del C.N.R. e Istituto di Chimica Fisica dell'Università, p. 35. Pisa 1969–1970.
5. Mély, B., Pullman, A.: Theoret. chim. Acta (Berl.) **13**, 278 (1969).
6. Nakajima, T., Pullman, A.: J. Chim. physique **55**, 793 (1958).

7. Pullman, A., Dreyfus, M., Mély, B.: Theoret. chim. Acta (Berl.) **16**, 85 (1970).
8. — Pullman, B.: Les Théories électroniques de la Chimie Organique, Masson (1952), for numerous examples.
9. Albert, A.: In: Physical methods in heterocyclic chemistry, ed. by A. R. Katritzky. New York: Academic Press 1963.
10. Cavalieri, L. F., Rosenberg, B. H.: J. Amer. chem. Soc. **79**, 5352 (1957).
Zubay, G.: Bioph. Bioch. Acta **28**, 644 (1958).
11. Christensen, J. J., Rytting, J. H., Izatt, R. M.: Biochemistry **9**, 4907 (1970).
12. Pal, B. C.: Biochemistry **1**, 558 (1962).
13. Bryan, R. F., Tomita, K.: Acta crystallogr. **15**, 1179 (1962).
14. Brookes, P., Lawley, P. D.: J. chem. Soc. (London) **1960**, 539.
Jones, J. W., Robins, R. K.: J. Amer. chem. Soc. **85**, 193 (1963).
Lawley, P. D., Brookes, P.: Biochem. J. **92**, 19C (1964); **89**, 127 (1963).
15. — Progress in nucleic acid research **5**, 89 (1966).
16. Foster, J. M., Boys, S. F.: Rev. mod. Physics **32**, 300 (1960).
17. Pullman, A., Berthod, H.: Unpublished results.
18. Christensen, J. J., Rytting, J. H., Izatt, R. M.: J. physic. Chem. **91**, 2700 (1967).
19. Brookes, P., Lawley, P. D.: J. chem. Soc. (London) **1962**, 1348; Bioph. Bioch. Acta **26**, 450 (1957).
20. Dove, N. F., Wallace, F. A., Davidson, N.: Bioch. Bioph. Res. Comm. **1**, 312 (1959).
21. Giessner-Prettre, C., Pullman, A.: Comp. Rend. Ac. Sci. Paris **272**C, 750 (1971).

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