Theoret. chim. Acta (Berl.) 9, 279—287 (1968)

Commentationes

All Valence-Electrons Calculations of the Biological Purines and Pyrimidines

I. CNDO Calculation*

C. Giessner-Prettre and A. Pullman

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Quantique, associé au CNRS,
13, rue Pierre Curie, Paris 5^e

Received November 16, 1967

A calculation of the electronic structure of the bases of DNA has been performed by an SCF procedure including simultaneously all the valence electrons. The results are analyzed and compared to those of previous calculations made in the π electron approximation.

Die Elektronenstrukturen der DNA-Basen werden mit einer SCF-Methode unter Einschluß aller Valenzelektronen untersucht. Die Ergebnisse werden mit denen reiner π-Rechnungen verglichen.

Un calcul de la structure électronique des bases de l'acide déoxyribonucléique a été effectué par un procédé self-consistent traitant tous les électrons de valence simultanément. Les résultats sont analysés et comparés à ceux de calculs antérieurs faits dans l'approximation de la séparation σ–π.

Introduction

The electronic structure of the base components of the nucleic acids has been thoroughly investigated by different refinements of the π-electron approximation [1] since the first Hückel calculation in 1956 [2]. Recently, the σ-framework has been explicitely introduced [3, 4] in a semi-empirical fashion so as to permit the predictions of dipole moments and the calculation of intermolecular interactions of the Van der Waals-London type. In all these works, however, the π and σ electrons were treated separately so that their mutual influence could only be taken into account empirically by a careful fitting of the parameters on reference compounds.

The development of high speed computers and of new approximations within the molecular orbital formalism [5—9], nowadays allows to treat at the same time all the valence electrons of large conjugated heterocycles.

The use of such techniques appears very attractive for studying the role of the terms neglected or empirically introduced in previous calculations. We have undertaken, from that viewpoint, a systematic study of the purines and pyrimidines of the nucleic acids. In the work reported in this paper we have chosen the CNDO procedure [7] which is an extension to all-valence electrons of the well-known Pariser-Parr-Pople self-consistent formalism. Its main advantage over the

---

* This work was supported by grant no. 67-00-532 of the Délégation Générale à la Recherche Scientifique et Technique (Comité de Biologie Moléculaire).

other procedures [5, 6] is that it introduces explicitly the interactions between all valence electrons.

We do not describe the procedure again since we have followed with no modification the CNDO/2 method described by Pople and Segal [9]. The geometries of the bases [10] required as input are the same as those used in our previous self-consistent π-electron calculations [4].

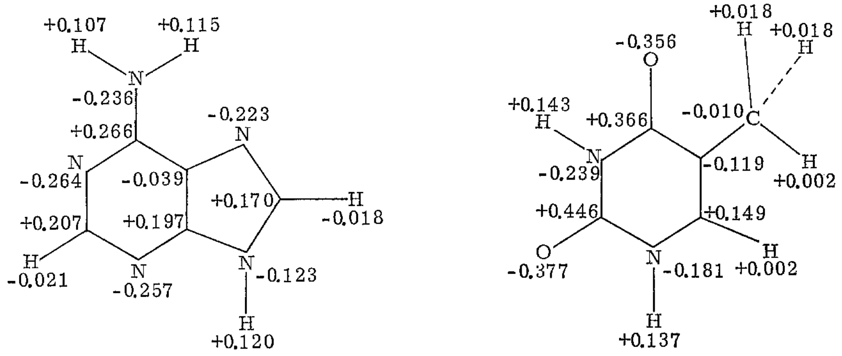
The present paper will be mainly concerned with adenine, guanine, cytosine and thymine (A, G, C, T). Similar data on uracil (U) have been in part already reported in a previous publication [11].

Electron Distribution and Dipole Moments

Fig. 1a, b, c give the distribution of the net atomic charges and of their σ- and π-components, respectively. The numerotation in purines is the chemical numbering. In pyrimidines N_1 is the nitrogen linked to ribose in ribosides.

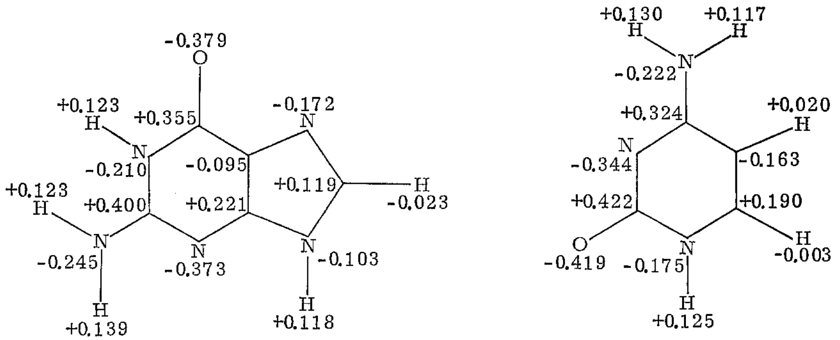
a) Total Net Charges

As would be expected, the most electronegative atoms bear the largest negative charges: all the carbonyl oxygens carry charges of the order of 0.4 e; the pyridine-type nitrogens attract 0.2 to 0.3 electron units whereas the pyrrole-like



Adenine

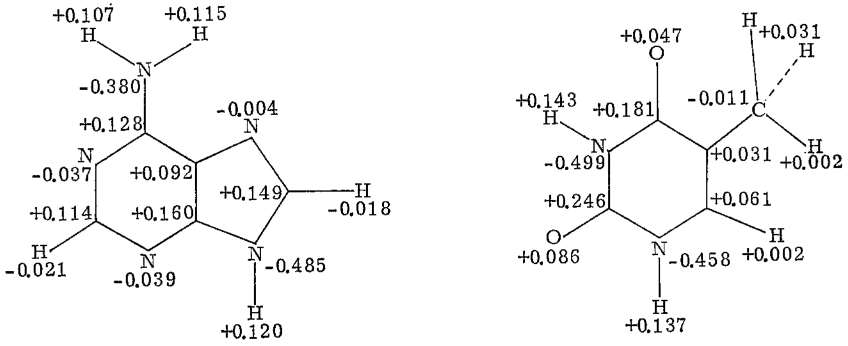
Thymine



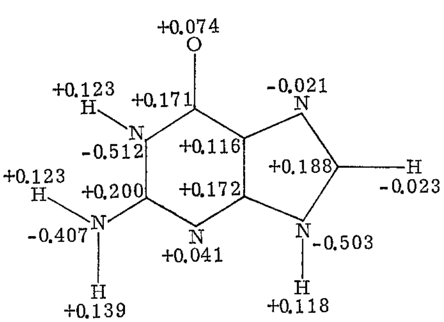
Guanine

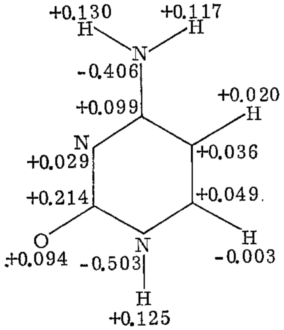
Cytosine

Fig. 1a. Net total charges



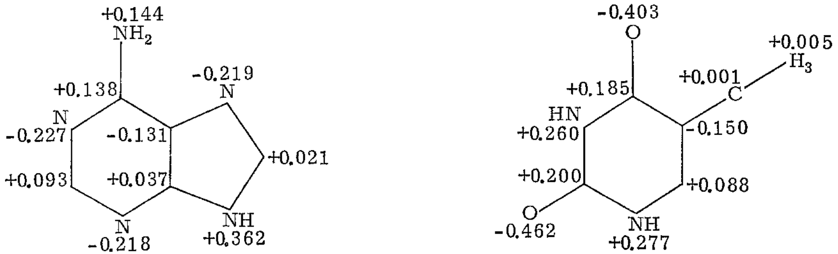
Adenine Thymine



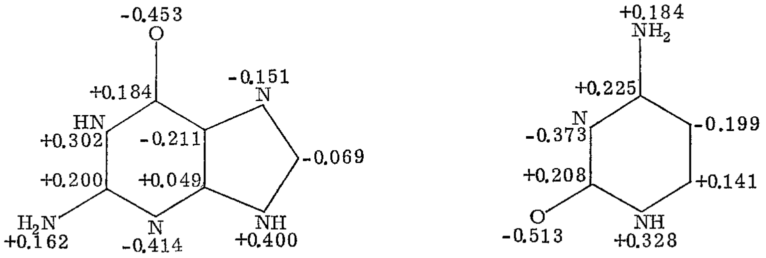


Guanine Cytosine

Fig. 1b. Net σ charges



Adenine Thymine(*)



Guanine Cytosine

Fig. 1c. Net π charges

---

* The charge of the hydrogen of the methyl group is calculated as the sum of the contributions of these hydrogens to the π orbitals

nitrogens are clearly less negative. NH_2 groups show an interesting constancy in global appearance. Qualitatively all these findings are in fair agreement with the image of the total net populations obtained previously by the semi-empirical approach [3, 4, 11]*, in which an approximate treatment of the σ bonds was added either to a Hückel or a Pariser-Parr-Pople representation of the π-electron cloud. The sole serious difference concerns the hydrogens bound to carbon atoms, which appear less discharged presently than in the semi-empirical approaches, and are even sometimes negatively charged, a feature which seems to be rather characteristic of the CNDO procedure [11, 12, 13].

b) Net σ- and π-Charges

Even though the overall picture of the total charges is similar to that obtained by simpler procedures, a separate examination of the σ and π contributions reveals some interesting differences mainly due to the behavior of the atoms carrying non-bonding lone-pairs: in the CNDO results these σ lone-pairs are strongly delocalized. This is especially illustrated by the slightly positive σ charges of N_3 of cytosine and N_3 of guanine, as shown in Fig. 1b. On the contrary in the localized-bond approximation which does not touch the non-bonding electrons, these nitrogen atoms are strongly negative [1]. The same phenomenon is observed for oxygen, but since Berthod and Pullman's parametrization for planar molecules leads to C=O σ bonds having a low polarity, the effect observed in the CNDO results is not so drastically different. It seems that the CNDO procedure through either its parametrization or the approximations used in the evaluation of the integrals, overemphasizes the delocalization of the lone pair, especially in the case of nitrogen. For example, the σ charge of the heteroatom in pyridine is -0.081 e according to CNDO against -0.309 e according to Berthod-Pullman [3] and -0.216 e in a recent non-empirical all-electron computation [14]. For the other atoms of the molecules considered here the localized-bond approach and CNDO give similar σ charges. Both type of calculations agree for the σ-polarity of C–NH and N–H bonds as well as for the small σ-charge displacement along the C=O bond.

In spite of the differences observed, there is an interesting feature of the CNDO results which gives support to the fundamental hypothesis of Del Re's procedure for the study of σ systems: as already observed in uracil and 5-fluorouracil [11] the σ atomic charges are determined only by the nature of the atom and by its nearest neighbors, and not by long-range interactions.

As to the π electron displacements, the present results (Fig. 1c) confirm our previous observation [11] that the π-charges are very similar to those calculated by the Pariser-Parr-Pople approximation, in spite of an entirely different parametrization. For thymine** we find only a very small contribution of the hydrogens to the π cloud of the ring (0.005 e).

Thus, on the whole, with the exception of the delocalization of the non-bonding electrons which will be investigated further, the charges computed by an

---

* Diagrams of the σ and π distributions which were not detailed in the original publications can be found in the review paper [1].

** The configuration adopted for the hydrogen atoms of the methyl group is the one suggested by Hoogsteen for 1-methylthymine [15].

approximation including all the valence electrons in an SCF formalism are in qualitative agreement with the results from separate treatment of σ and π electrons.

c) *Dipole Moments*

The dipole moments have been calculated from the electron distribution of Fig. 1 by the expression given by Pople and Gordon [12], so as to take into account the non-sphericity of the atomic charges [16]. We have given in Table 1 the details of the different contributions to the moments of the nucleic bases (pyrrole and pyridine are added for comparison). The inclusion of the contribution of the atomic dipoles is clearly quite important and is in fact necessary for a good agreement with the experimental values of the moments (the data concerning pyridine and pyrrole are particularly instructive in this connection).

Table 1. *Dipole moments (the first number is the moment in debye units, the second is its angle in degrees with* $N_1 C_4$ *in* C, U, T, *with* $N_3 C_6$ *in* G *and* A. *Sign as in Ref.* [3]*)*

Molecule	$\mu_{Q\sigma}$ [a]		μ_{sp} [b]		μ_{σ} [c]		μ_{π}		$\mu_{Q\sigma} + \mu_{\pi}$		$\mu_{\sigma} + \mu_{\pi}$		exp. values [d]
Pyrrole	1.31	180	0.78	0	0.53	180	2.52	0	1.21	0	1.99	0	1.8
Pyridine	0.22	180	1.60	180	1.83	180	0.33	180	0.55	180	2.15	180	2.2
Adenine	1.11	−111	1.59	53	0.71	27	2.34	75	1.24	80	2.86	64	3.0
Thymine	1.20	−173	1.45	32	0.64	86	3.94	33	2.92	43	4.35	39	
Uracile	1.21	−170	1.53	36	0.70	87	4.20	29	3.08	36	4.61	36	3.9
Guanine	2.23	153	2.24	−20	0.28	63	7.27	−29	5.04	−30	7.26	−27	
Cytosine	1.78	− 89	3.17	84	1.41	75	6.39	108	4.70	115	7.61	102	

[a] Contribution obtained from the net charges.
[b] Contribution obtained from the atomic dipoles.
[c] Sum of the first two components.
[d] As in Ref. [4].

But the most striking feature of the data is that the total dipole moments of the nucleic bases obtained by the CNDO procedure follow the same trend as that displayed by the values calculated previously by Berthod, Giessner and Pullman [4] and already present in the early simpler evaluations [3, 17]: *cytosine and guanine are predicted to have dipole moments of the same order of magnitude, much larger than uracil and thymine, themselves with a larger moment than adenine.*

The absolute values themselves are very similar to the values calculated previously. The agreement with experiment is satisfactory for the known cases of adenine, thymine and uracil*.

Energy Levels and Transition Energies

a) *Molecular Orbitals and Ionization Potentials*

Table 2 gives the individual orbital energies for A, G, C, U and T. Although such data must certainly be taken with caution, it is no doubt permissible to look

---

* In the case of uracil the calculated value given in Ref. [11] did not include the atomic dipoles.

Table 2. *Occupied energy levels in the ground states of the nucleic bases (eV) (sign reversed)*[a]

U	T	C	G	A
52.87	53.03	52.36	55.49	55.33
45.09	45.87	44.91	51.27	50.39
43.64	43.96	41.69	48.23	44.13
40.22	41.59	38.85	42.73	38.86
36.88	39.61	36.61	39.50	37.34
32.25	32.88	31.66	38.98	33.94
27.41	31.98	27.75	32.08	30.41
π 26.38	27.38	25.74	30.57	29.94
26.36	π 26.90	π 25.53	28.50	27.73
25.71	26.75	25.45	27.82	π 27.07
21.39	24.74	20.96	π 27.60	25.64
π 20.07	π 22.21	20.01	26.48	22.62
19.83	21.35	π 19.91	π 22.82	π 21.15
19.12	20.00	18.40	22.46	20.51
π 18.65	π 19.66	17.91	21.61	19.82
17.72	19.22	π 17.20	20.60	18.07
16.42	18.58	15.56	π 19.66	π 17.89
13.83	π 17.56	13.68	19.55	17.31
π 13.18	17.42	π 13.21	17.38	15.76
12.90	15.88	11.81	16.75	15.11
π 11.88	13.74	π 10.78	15.65	π 14.14
	π 13.10		π 15.34	13.11
	12.85		15.08	π 12.57
	π 11.37		π 14.21	11.46
			π 13.11	π 10.08
			12.77	
			11.55	
			π 9.06	

[a] π orbitals are indicated as such. All other orbitals are σ.

for trends and evolutions in such a family of compounds. The highest occupied orbital ("homo"), the energy of which is an indication of the value of the ionization potential of the molecule and whose symmetry is an indication of the symmetry of the ionized species, is found to be π for all four DNA bases as well as for uracil. The order of decreasing energy (in absolute value) of the "homo" is the same as the one obtained by Pariser-Parr-Pople π calculations [4], that is:

$$U > T > C > A > G .$$

In particular CNDO confirms the very low value calculated for guanine for which no experimental data are available. Although the numerical values are too large and too widely spread, there is no contradiction with the order of the experimental data which indicate for the ionization potentials [18].

$$U > T > A = C .$$

Whereas the highest occupied and lowest empty orbitals are π-orbitals for the four molecules, the second highest in energy is a σ-orbital. CNDO/2 certainly overestimates the energy of the σ orbitals compared to the π orbitals since this method indicates a σ homo for pyridine and ethylene [13, 19]. However the intermingling of the σ and π orbitals calculated is strong enough to suggest that part of the σ electrons have an energy of the same order of magnitude as the π except the ones located on the highest orbitals. Similar results have been obtained in non-empirical all-electron computations on other conjugated molecules [14, 20, 21].

The examination of the coefficients of the atomic orbitals in the *second* homo shows that in all carbonyl compounds, this orbital contains an important contribution of the oxygen atomic orbitals, the purest ones occuring in uracil and thymine. This second homo gives thus a rough measure of the relative value of the oxygen "lone-pair" ionization potentials: it is interesting that uracil appears as the poorest $n(\mathrm{O})$ donor and guanine as the best $n(\mathrm{O})$ donor among these oxygenated bases in conformity to an earlier prediction of PULLMAN and ROSSI [22] which also indicated that in a given compound the oxygen lone-pairs should be more ionizable than the nitrogen lone-pairs, a conclusion supported by the present data. It is not really possible, however, to speak of nitrogen lone-pairs in the present calculation since they appear as strongly mixed with other atomic orbitals in the molecular orbitals. Thus "energy-wise" there are no lone-pairs on nitrogen in the sense that they do not appear as separately ionizable insofar as the energies of the molecular orbitals can be taken as meaningful from the point of view of ionization. However "population-wise" the *total* σ-population of each pyridinic nitrogen shows very little loss or gain of electrons with respect to the classical image of three sp^2 orbitals occupied by four electrons. In fact, strangely enough, at the end of the calculation a typical σ-population of the pyridinic nitrogens is $s^{1.4}\,(xy)^{2.6}$, wery close to $s^{4/3}\,(xy)^{8/3}$. Whether this is the real image of the electron distribution remains to be elucidated and will be discussed further elsewhere.

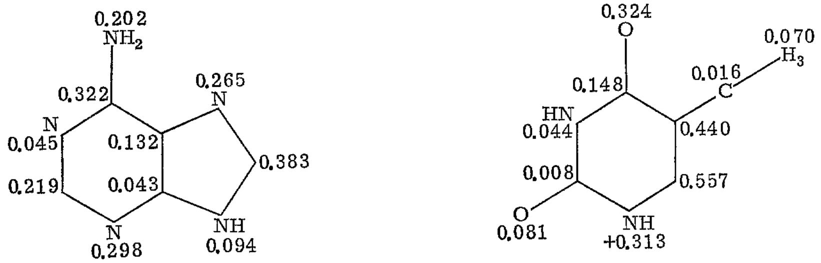
b) Excited States

The transition energies calculated from the ground state orbitals are given in Table 3. As to be expected from the parametrization of CNDO/2 [19] the calculated values are too large by an order of magnitude. Nevertheless they indicate that guanine and cytosine absorb at longer wavelength than adenine and thymine, in agreement with experiment [23] and with previous π calculations [24—26]. In addition, the bathochromic shift observed from uracil to thymine is very well reproduced. The inspection of the orbital energies in thymine and uracil shows that this bathochromic shift is only due to a rising of the highest occupied orbital; the two compounds have their lowest empty orbital at the same energy. As concerns the triplet states, thymine is the base having the lowest first triplet. This result is in accord with the experimental result which attributes to this molecule the phosphorescence of DNA [27]. Further calculations are in progress by improved CNDO procedures with a particular emphasis on the improvment of the calculated energy values and will be reported later.

Table 3. *Calculated values of the first transition energies (eV)*

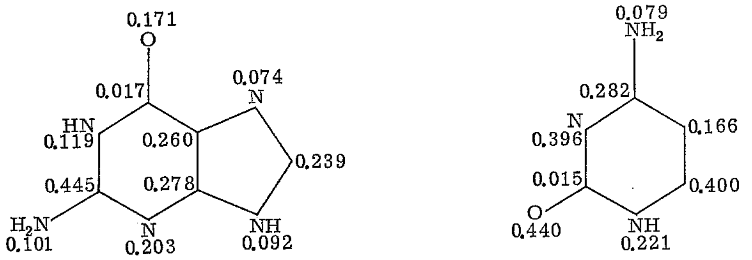
Base	Singlet	Triplet
A	9.2	6.5
U	9.3	5.9
T	8.9	5.6
G	7.6	6.0
C	8.3	6.6

It seems that the numerical discrepancy of the calculated energetic quantities with the experimental data does not affect appreciably the wave function obtained. This is suggested by the comparison of the distribution of the unpaired electrons



Adenine

Thymine



Guanine

Cytosine

Fig. 2. Unpaired electron distribution in the first SCF excited states

in the first excited state (Fig. 2) and the corresponding result obtained in the Pariser-Parr π-electron approximation [25]. In addition to the large fraction of unpaired electrons located on the C_5–C_6 bond of thymine, CNDO/2 confirms that a net decrease of the C_5C_6 π bond order accompanies the excitation of this molecule (0.842 in the ground state and 0.348 in the first excited state), features which appeared in *all* the different refinements of the π-electron approximation [28].

Conclusion

The present work confirms that the CNDO/2 procedure is appropriate for the calculation of dipole moments and that even though the numerical values of the energies calculated are not accurate, they reproduce in a reasonable qualitative fashion the most outstanding features of the experimental facts.

On the other hand, it has been shown that a simultaneous treatment of all valence electrons with no hypothesis on the valence state of the atoms involved led to an image of the electron distribution quite similar in many respects (dipole moments, π electron displacements, spin densities) to the image obtained by more empirical procedures. It seems that a localized-bond procedure for the σ electrons added to an adequately parametrized SCF procedure for the π electrons gives a global representation of the electronic structure of this kind of molecules which is quite satisfactory, probably owing essentially to the fact that the choice of the Pariser-Parr-Pople π-integrals was carefully made [4] on a series of reference compounds, taking into account the presence of the underlying σ structure [1], so that one could not speak of a completely independent treatment of the σ and π systems.

References

1. A review will be found in Pullman, A., and B. Pullman: Advances in quantum chemistry, vol. IV. (In press).
2. Pullman, A., B. Pullman et G. Berthier: C. R. Acad. Sci. Paris **243**, 380 (1956).
3. Berthod, H., et A. Pullman: J. Chim. physique **52**, 942 (1965).
4. —, C. Giessner-Prettre und A. Pullman: Theoret. chim. Acta (Berl.) **5**, 53 (1966).
5. Hoffman, R.: J. chem. Physics **39**, 1397 (1963); **40**, 2474, 2480 (1964).
6. Cusachs, L. C., and B. B. Cusachs: J. physic. Chem. **71**, 1060 (1967).
7. Pople, J. A., D. P. Santry, and G. A. Segal: J. chem. Physics **43**, S 129 (1965).
8. —, and G. A. Segal: J. chem. Physics **43**, S 135 (1965).
9. — — J. chem. Physics **44**, 3289 (1966).
10. Spencer, M.: Acta Cryst. **12**, 59 (1959), for all the bases except cytosine which has been taken from Barker, D. L., and R. E. Marsh: Acta Cryst. **17**, 1581 (1964).
11. Berthod, H., C. Giessner-Prettre und A. Pullman: Theoret. chim. Acta (Berl.) 8, 212 (1967).
12. Pople, J. A., and M. Gordon: J. Amer. chem. Soc. 89, 4253 (1967).
13. Jordan, F.: Ph. D. Thesis, Univ. of Pennsylvania 1967.
14. Clementi, E.: J. chem. Physics **46**, 4731 (1967).
15. Hoogsteen, K.: Acta Cryst. **16**, 28 (1963).
16. Coulson, C. A.: Trans. Faraday Soc. **38**, 433 (1942).
17. De Voe, H., and I. Tinoco Jr.: J. mol. Biol. **4**, 560 (1962).
18. Lifschitz, C., E. D. Bergmann, and B. Pullman: Tetrahedron Letters **1967**, 4583.
19. Giessner-Prettre, C., and A. Pullman: Unpublished data.
20. Clementi, E., H. Clementi, and D. R. Davis: J. chem. Physics **46**, 4725 (1967).
21. — J. chem. Physics **46**, 4737 (1967).
22. Pullman, A., and M. Rossi: Biochim. biophys. Acta 88, 211 (1964).
23. Clark, L. B., and I. Tinoco Jr.: J. Amer. chem. Soc. 87, 11 (1965).
24. Berthod, H., C. Giessner-Prettre et A. Pullman: C. R. Acad. Sci. Paris **262**, 2657 (1966).
25. — Int. J. Quant. Chem. **1**, 123 (1967).
26. Denis, A., and A. Pullman: Theoret. chim. Acta (Berl.) **7**, 110 (1967).
27. Helene, C.: Thèse, Paris 1965.
28. Pullman, A.: In: "The Triplet State". Cambridge University Press, 1967.

Dr. Alberte Pullman
Institut de Biologie Physico-Chimique
13 Rue Pierre Curie
Paris V, France