Theoret. chim. Acta (Berl.) 13, 278—287 (1969)

# *Ab initio* Calculations on Cytosine, Thymine and Adenine*

B. MELY and A. PULLMAN

Institut de Biologie Physico-Chimique
Laboratoire de Biochimie Quantique associé au C.N.R.S., Paris

Received January 13, 1969

An *ab initio* LCAO SCF calculation has been performed in a small contracted GTO basis set on three bases of ADN. A check of the representativity of the basis set is reported for formamide and pyrrole.

Es wurde eine *ab initio* LCAO-SCF-Rechnung mit einem kleinen, kontrahierten GTO-Basissatz für drei ADN-Basen durchgeführt. Eine Prüfung des Einflusses dieses Basissatzes wird für Formamid und Pyrrol beschrieben.

Un calcul *ab initio* dans une petite base de gaussiennes contractée a été fait pour trois bases de l'ADN. Un calcul-test sur la representativité de la base atomique est donné sur les exemples de la formamide et du pyrrole.

The structure of the fundamental bases of the nucleic acids has been studied thoroughly by the method of molecular orbitals in the $\pi$-electron approximation [1] and more recently by different procedures including all the valence electrons [2, 3, 4, 5]. The comparison of the all-valence electrons calculations between themselves and with the previous results has shown [6] that a relative constancy in the most outstanding molecular features emerges from all calculations, particularly as concerns the large values of the dipole moments predicted for guanine and cytosine as compared to those of adenine and thymine, the general directions of these moments, the order of the $\pi$ ionization potentials, the characteristics of the oxygen lone-pair and the hybridization ratios on the different atoms [7, 8]. Some differences appear, however, which cannot all be traced back to the differences in the fundamental frameworks of the semi-empirical methods [8], so that less empirical calculations seem desirable. We report here on the results of all-electrons, non-empirical, SCF calculations on the two pyrimidines, cytosine and thymine, and on adenine.

## Method

The procedure employed is the LCAO MO SCF [9] method treating all electrons in the field of the nuclear framework. The molecular orbitals were written as combinations of *Gaussian* orbitals and the IBMOL [10] program was utilized. The large number of electrons (58 in cytosine, 66 in thymine and 70 in adenine) and the low symmetry of the molecules precluding the use of a large

---

* This work was supported by grant 67-00-532 of the Délégation Générale à la Recherche Scientifique et Technique (Comité de Biologie Moléculaire).

Table 1. *Final exponents in the* GTO *basis*[a]

| GTO | C | N | O | H |
|---|---|---|---|---|
| $s_1$ | 56.2694 | 79.2170 | 106.268 | 5.40044 |
| $s_2$ | 8.79230 | 12.3790 | 16.6050 | 0.817530 |
| $s_3$ | 1.69090 | 2.38000 | 3.19300 | 0.181650 |
| $s_4$ | 0.225400 | 0.317000 | 0.425700 | |
| $x_1, y_1, z_1$ | 0.739300 | 1.06100 | 1.44700 | |
| $x_2, y_2, z_2$ | 0.172590 | 0.249000 | 0.339200 | |

[a] Notation.The uncontracted Gaussian function, $\chi$ on a center $A$ is defined as:

$$\chi(A, \alpha, l, m, n) = N(x - A_x)^l (y - A_y)^m (z - A_z)^n \exp(-\alpha r_A^2)$$
$$= N x_A^l y_A^m z_A^n \exp(-\alpha r_A^2).$$

$N$ is a normalisation factor.

GTO basis set, a preliminary study was made on small molecules ($H_2$, $N_2$, CH, NH, OH) in the hope to obtain a "molecule-calibrated" basis of a reasonable size. This has shown [11] that a quite satisfactory representation could be reached with a carefully choosen small basis, even though the total energy is still far from the Hartree-Fock limit. The final exponents for the $4^s - 2^p$ heavy atom set and the $3^s$ hydrogen basis are given in Table 1. The results obtained in the hydride series suggested the possibility of further contracting this basis set with the coefficients indicated in Table 2. A check on the validity of the contracted *versus* the uncontracted set has been performed on the formamide molecule which contains essentially all the atoms of the nucleic acid components in similar environments. Moreover another testing has been made on the pyrrole molecule for which a GTO MO SCF calculation with a more extended basis set is available for comparison [12]. The results are briefly summarized below.

One of the tests put on the molecular wave functions has been their ability to give reasonnable values of the molecular dipole moments (in the limits of accuracy which may be reached without configuration mixing [13, 14]). These have been computed using the dipole-length operator and the exact values of the integrals in the gaussian basis set[1].

Table 2. *Contraction coefficients*

| | C | N | O | H |
|---|---|---|---|---|
| $s_1$ | 0.2085 | 0.2037 | 0.1998 | 1.0000 |
| $s_2$ | 0.7035 | 0.7028 | 0.7019 | 0.4500 |
| $s_3$ | 0.2526 | 0.2579 | 0.2623 | 0.5940 |
| $s_4$ | 1.0000 | 1.0000 | 1.0000 | |
| $x_1, y_1$ | 0.3827 | 0.4059 | 0.4250 | |
| $x_2, y_2$ | 0.3069 | 0.3581 | 0.3938 | |
| $z_1$ | 0.5285 | 0.5419 | 0.5249 | |
| $z_2$ | 0.6090 | 0.5953 | 0.6119 | |

---

[1] An appropriate addition to the IBMOL program has been written by Ph. Millié et D. J. David, whose help is gratefully acknowledged.

## Formamide and Pyrrole

Three calculations have been performed for formamide: the first two are based on the planar geometry of Kurland and Wilson [15] (KW and KWC designate the corresponding calculations with uncontracted and contracted set respectively). The third (uncontracted) is based on the non-planar structure of Costain and Dowling [16] (CD). The main results of the KW and KWC computations are given in Table 3 where recent data [17] corresponding in accuracy to a best atom double zeta Slater set (BADZ) are also reproduced for comparison.

Table 3. *Results for formamide in different basis sets*

| | KW | KWC | BADZ [17] |
|---|---|---|---|
| Energies of occupied orbitals[a] (a.u.) (reversed sign) | 20.7200 | 20.8166 | 20.532 |
| | 15.6873 | 15.7870 | 15.594 |
| | 11.3797 | 11.5483 | 11.387 |
| | 1.4427 | 1.4405 | 1.394 |
| | 1.2695 | 1.2752 | 1.239 |
| | 0.86917 | 0.89485 | 0.877 |
| | 0.75579 | 0.78736 | 0.767 |
| | 0.65808 | 0.68975 | 0.672 |
| | 0.58410 | 0.60316 | 0.601 |
| | $\pi$ 0.56342 | $\pi$ 0.60169 | $\pi$ 0.577 |
| | 0.40100 | 0.43349 | 0.436 |
| | $\pi$ 0.38618 | $\pi$ 0.41901 | $\pi$ 0.416 |
| Virtual orbitals[a] (a.u.) | $\pi^*$ 0.21881 | $\pi^*$ 0.19745 | |
| | 0.25094 | 0.37571 | |
| | 0.28902 | 0.45562 | |
| | 0.35194 | 0.53061 | |
| Dipole moment[b] (debye units) | 3.97 | 4.03 | 4.95 |
| angle with NC direction[b] | 37°.4 | 38°.3 | 42° |

[a] Unlabeled values are $\sigma$ orbitals.

[b] Experimental value: $\theta = 39.60$ [17], $\mu = 3.7 \pm 0.06\ D$ [17].

The CD results are not enough different from the KW-ones to require tabulation, with the exception of the dipole moment which becomes 3.51 $D$, $\theta = 38°7'$ (a similar decrease accompanies the change in geometry in BADZ [17]). The similarity of the results obtained in the contracted set with the more elaborate BADZ calculation are clearly encouraging[2]. The results of a Mulliken population analysis [19] of the KWC wave function is given in Fig. 1. Clearly enough, the general image of the amino and carbonyl groupings is in complete agreement with that obtained with larger basis sets [17, 18]. Even the *numbers* are very similar. The strong discharge of the hydrogen atoms which appears in the present calculation (as well as with the uncontracted basis) is also present in other GTO SCF results [17, 18, 12, 20]. As a consequence, the aminonitrogen carries a very large excess of $\sigma$ electrons. The oxygen atom is both $\sigma$ negative and $\pi$ negative and the carbon atom is $\sigma$ and $\pi$ positive. Overlap considerations would indicate CH bonds stronger than NH bonds, a reasonable conclusion.

---

[2] For a more detailed analysis and further comparison with other basis sets [18], see Ref. [11].

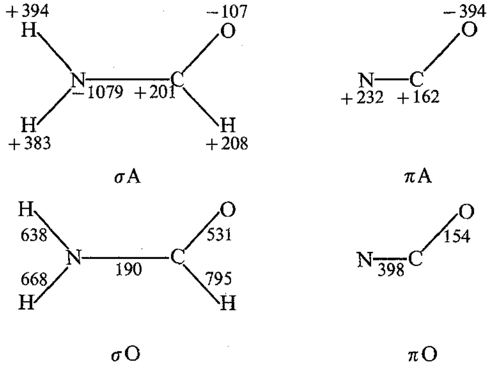


Fig. 1. σA, πA: Gross atomic populations in millielectron units (excess: negative, defect: positive). σO, πO: Gross overlap populations [19] (same unit)

Proceeding further, we give in Table 4 the main results obtained with our contracted basis for the pyrrole molecule. The orbital energy values are very similar to those calculated by Clementi for the σ orbitals, the π-orbitals beeing systematically higher, which results in a π-ionization potential, calculated by Koopmans' theorem, of 9.73 eV, in reasonable agreement with the 8.9 eV experimental value [21]. Table 5 gives the calculated dipole moment and its decomposition into the σ and π components. The direction and size of the moment appear as commanded by the π component from which a small σ-component substracts, results very similar to those of more empirical calculations [22, 23, 3].

Table 4. *Energies of the molecular orbitals in pyrrole (eV; reversed sign)*

| Clementi | | | | Present work | | | |
|---|---|---|---|---|---|---|---|
| $A_1(\sigma)$ | $B_2(\sigma)$ | $B_1(\pi)$ | $A_2(\pi)$ | $A_1(\sigma)$ | $B_2(\sigma)$ | $B_1(\pi)$ | $A_2(\pi)$ |
| 427.47 | 310.88 | 17.18 | 10.56 | 430.95 | 311.01 | 16.50 | 9.73 |
| 310.88 | 309.61 | 11.57 | | 311.00 | 309.60 | 10.71 | |
| 309.63 | 28.15 | | | 309.62 | 28.14 | | |
| 36.02 | 21.69 | | | 36.72 | 21.74 | | |
| 29.81 | 16.99 | | | 30.18 | 16.96 | | |
| 22.45 | 16.39 | | | 22.63 | 16.26 | | |
| 21.17 | | | | 21.41 | | | |
| 17.62 | | | | 17.65 | | | |
| 15.69 | | | | 15.49 | | | |

Table 5. *Pyrrole dipole moment (debye units)*

| | [a] | [b] | [c] |
|---|---|---|---|
| $\mu_{\text{total}}$ | 2.10 | 2.16 | 1.14 |
| $\mu_\sigma$ | −0.49 | −0.72 | −1.72 |
| $\mu_\pi$ | 2.59 | 2.88 | 2.87 |

[a] Calculated with Clementi's wave function.

[b] Calculated with our wave function.

[c] Point-charge approximation.

The calculated value, although reasonable, is larger than the experimental value of 1.71 $D$[3]. It is practically the same as the corresponding value which we have calculated from Clementi's wave function with a very analogous $\sigma-\pi$ decomposition.

The Mulliken population analysis gives gross $\sigma$ and $\pi$ charges which are listed in the last column of Table 6 giving also Clementi's values and other results obtained in various $\pi$ or all-valence electrons treatments. Our numbers are again very similar to those obtained with the more extended basis, the $\sigma$-displacements beeing generally larger here. The strong discharge of the hydrogens is again observed. On the whole, the qualitative image of the pyrrole polarity is very constant: in particular $C_\beta$ is more negative than $C_\alpha$ both in $\pi$ and total charges (total only in CNDO and IEHT) like in many other molecular orbital calculations [22, 26][4]. The charge displacements appear very large in the population analysis of the GTO wave functions but it must be kept in mind that not too great an importance should be attached to the values, insofar as boundary surfaces are not unequivocally defined [13]. At any rate different basis sets, and for that matter, different procedures, *should* give different gross populations.

Table 6. *Gross atomic populations in pyrrole by different methods (millielectron units)*

| | | [a] | EHT[b] | IEHT[b] | CNDO[b] | Clementi | present work |
|---|---|---|---|---|---|---|---|
| $\sigma$ | H(N) | +193 | +304 | +219 | +103 | +339 | +408 |
| | H($\alpha$) | +60 | +107 | +57 | -1 | +204 | +250 |
| | H($\beta$) | +54 | +108 | +29 | -1 | +192 | +246 |
| | N | -459 | -749 | -739 | -443 | -749 | -962 |
| | $C_\alpha$ | +62 | +117 | +160 | +133 | -30 | -13 |
| | $C_\beta$ | -43 | -110 | +13 | +40 | -160 | -206 |
| $\pi$ | N | +307 | +413 | +643 | +365 | +341 | +375 |
| | $C_\alpha$ | -65 | -65 | -210 | -94 | -75 | -82 |
| | $C_\beta$ | -89 | -142 | -112 | -89 | -95 | -106 |

[a] $\sigma$ from a modified Del Re procedure [22], $\pi$ from a Pariser-Parr-Pople calculation [23].

[b] Extended Hückel, iterative extended Hückel, CNDO/2, like in Ref. [8].

Moreover these populations should *not* be used as point-charges for evaluating the dipole moments [27] as may be seen in the last column of Table 5. (The corresponding values obtained with Clementi's populations are -1.39, 2.59 and 1.20 debyes respectively).

## Cytosine, Thymine and Adenine

Each molecule (Fig. 2) has been treated as an all-electrons problem by the procedure described using the same geometries as in Ref. [3].

The energy quantities resulting from the SCF calculation are given in Table 7; the dipole moment data are summarized in Table 8 and the gross atomic and overlap populations in Fig. 3.

---

[3] In benzene, recent redetermination of E. D. Bergmann and Feichenfeld (private communication).

[4] See Ref. [11] for a further discussion on this point.

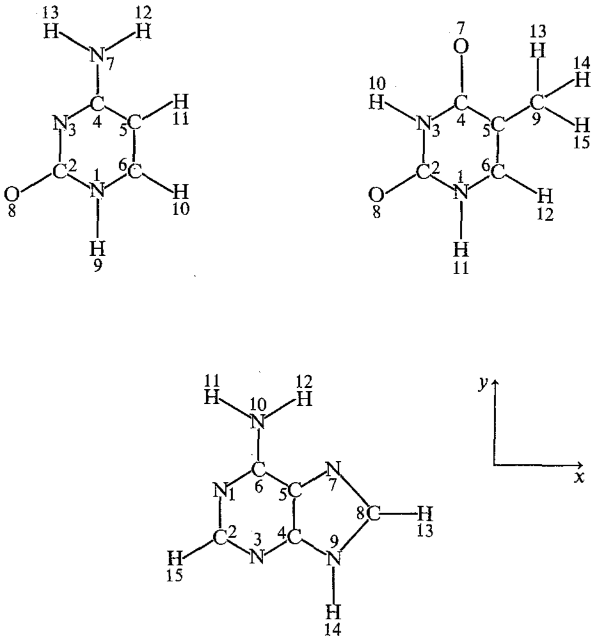


Fig. 2. Atom numbering and axes in the nucleic bases

### *a) Energy Levels*

In all molecules, the deepest energy levels are the 1*s* orbitals of the heavy atoms C, N, O, nearly undistorted by molecular formation. They start at about 40 eV the other $\sigma$ MO's which extend as high as 11–12 eV, the $\pi$ levels starting around 19–20 up to about 10 eV.

In the three molecules the *homo* is always a $\pi$ orbital, the next (or second next in $T$) being $\sigma$ with a separation however of at least 1.6 eV. Koopmans' theorem would thus predict the first ionization potential as being $\pi$ in the three bases with the increasing order:

$$A < C < T$$

and the spacing of the first two orbitals seems to preclude a reversal of the ordering by correlation effects.

It thus seems that non-empirical all-electron calculations indicate that the first ionization potentials in these molecules is of $\pi$-character as was predicted earlier on the basis of much simpler calculations [28] and recently confirmed by experimental measurements [29]. Moreover, the relative order of the values is also the same as found previously and experimentally [6].

The lowest empty orbitals appear as $\pi^*$ in character with a rather large gap before the first $\sigma^*$ level. The bearing of this feature on the possibilities of excited states representation will be dealt with separately as well as detailed comparison with the results of all-valence electrons calculations.

Table 7. *Energies of the molecular orbitals in cytosine, thymine and adenine (eV). Unlabeled orbitals are $\sigma$. Upper part contains the occupied orbitals (sign changed), lower part contains the first virtual orbitals*

| C | T | A |
|---|---|---|
| 565.42 | 567.00 | 432.55 |
| 429.28 | 566.78 | 429.70 |
| 428.79 | 431.87 | 429.66 |
| 428.35 | 431.07 | 429.21 |
| 315.09 | 316.69 | 429.14 |
| 314.55 | 315.07 | 314.43 |
| 313.73 | 313.28 | 313.91 |
| 311.55 | 310.87 | 313.37 |
| 39.46 | 309.81 | 313.00 |
| 36.84 | 41.37 | 311.91 |
| 35.48 | 40.18 | 40.60 |
| 32.63 | 37.80 | 38.33 |
| 31.24 | 36.00 | 36.10 |
| 26.55 | 31.77 | 34.79 |
| 25.71 | 28.11 | 33.76 |
| 23.37 | 26.91 | 30.44 |
| 22.58 | 25.95 | 26.57 |
| 21.78 | 23.02 | 25.54 |
| 20.31 | 22.58 | 25.13 |
| 19.09 | 20.99 | 23.22 |
| $\pi$ 18.73 | $\pi$ 19.57 | 21.79 |
| 18.51 | 19.13 | 20.64 |
| $\pi$ 16.27 | 17.72 | 20.36 |
| 16.17 | $\pi$ 17.55 | $\pi$ 19.34 |
| $\pi$ 14.44 | 17.52 | 18.85 |
| 11.89 | 16.80 | 18.30 |
| $\pi$ 11.45 | $\pi$ 16.52 | 17.83 |
| 11.40 | 16.28 | $\pi$ 16.99 |
| $\pi$ 9.79 | 15.09 | $\pi$ 14.93 |
|  | 13.30 | 14.11 |
|  | 12.11 | 12.98 |
|  | $\pi$ 11.91 | $\pi$ 12.79 |
|  | $\pi$ 10.16 | 11.67 |
|  |  | 11.27 |
|  |  | $\pi$ 9.41 |
|  |  |  |
| $\pi^*$ 2.20 | $\pi^*$ 2.29 | $\pi^*$ 2.69 |
| $\pi^*$ 4.60 | $\pi^*$ 4.71 | $\pi^*$ 3.73 |
| $\pi^*$ 9.77 | 8.52 | $\pi^*$ 5.61 |
| 10.26 | $\pi^*$ 9.09 | 9.11 |
| 10.80 | 10.67 | 9.93 |
| 12.35 | 10.75 | 10.75 |
|  | 12.87 | $\pi^*$ 10.97 |

### *b) Dipole Moments*

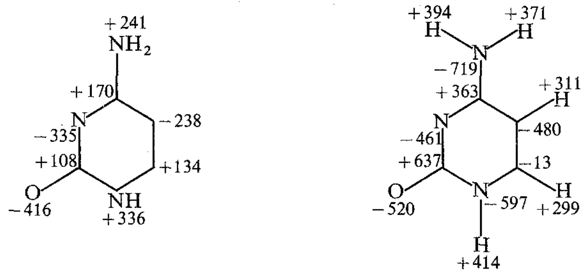
The calculated values of the dipole moments confirm entirely the earlier predictions of a high dipole for cytosine as compared to adenine and thymine [30, 22, 23]. Obviously this is a very constant feature of all reasonable calculations [6]. The directions of the moments are very similar to those obtained previously (adenine beeing the most different). It is seen here again that the $\pi$-component is

Table 8. *Dipole moments (debye units) in the nucleic bases*[a]

| | Exact value | | | | | | Point-charge approximation | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total | | $\sigma$ | | $\pi$ | | total | | $\sigma$ | $\pi$ |
| A | 2.83 | 93 | 0.77 | 32 | 2.46 | 101 | 4.55 | 91 | 0.99 | 4.55 |
| C | 6.76 | 103 | 1.29 | 68 | 5.76 | 111 | 14.55 | 116 | 4.32 | 10.38 |
| T | 4.22 | 50 | | — | | — | 10.53 | 57 | — | — |

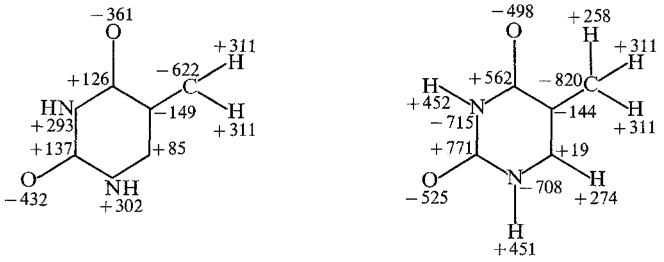
Approximate experimental moments: A: 3 *D*; T: 3.9 *D* (see Ref. [23]).

[a] The first number is the moment, the second number is the angle with respect to the *y* axis (measured counterclockwise).



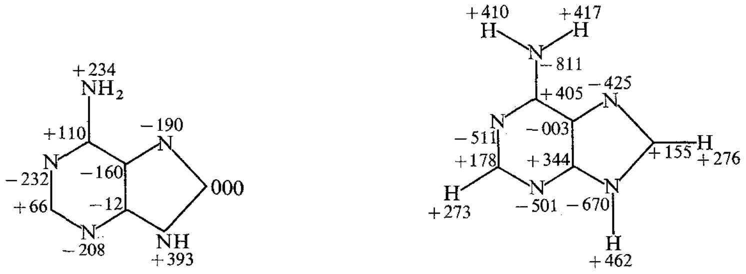
$\pi$ total

Cytosine



$\pi$ total

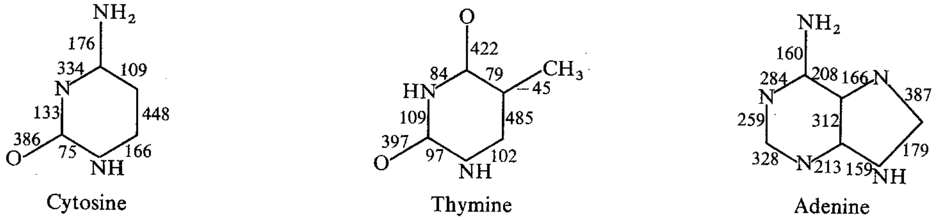
Thymine



$\pi$ total

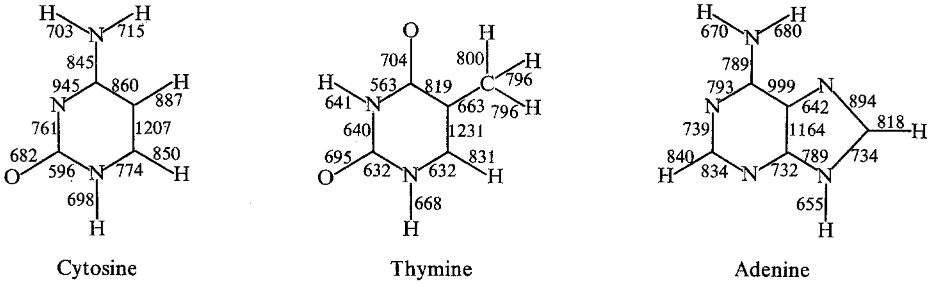
Adenine

Fig. 3a. Gross populations (net) (millielectron units)



Cytosine Thymine Adenine

Fig. 3 b. $\pi$-overlap populations (millielectron units)



Cytosine Thymine Adenine

Fig. 3c. Gross overlap populations

the most important part of the moment both in determining size and direction. Finally the second part of Table 8 illustrates once more the inadequacy of the point-charge approximation insofar as numerical values are required, although here it seems to give both the directions and the relative lengths of the moments.

### *c) Populations*

The numbers of Fig. 3 are self-explanatory. As already observed, the gross atomic populations are very large. A large part of them, it must be recalled, are due to the overlap charge arbitrarily halved over each partner in a pair. Certainly the numerical values of the "charges" should not be taken too litterally since they depend on the atomic basis set. On a relative scale however they give a certain image of the electronic displacements upon molecular formation [19]. A comparison of this image with that of more empirical calculations [8] will be carried out elsewhere as well as the study of the "hybridization ratios".

## References

1. For a review, see Pullman, A., and B. Pullman: Advances in quantum chemistry **4**, 267 (1968).
2. Jordan, F., and B. Pullman: Theoret. chim. Acta (Berl.) **9**, 242 (1968).
3. Giessner-Prettre, C., and A. Pullman: Theoret. chim. Acta (Berl.) **9**, 279 (1968).
4. Pullman, A., E. Kochanski, M. Gilbert, and A. Denis: Theoret. chim. Acta (Berl.) **10**, 231 (1968).
5. Rein, R., N. Fukuda, G. A. Clarke et F. E. Harris: J. theoret. Biol. **21**, 88 (1968).
6. Pullman, A.: Ann. New York Acad. Sci., in press.
7. — Yale Seminar on $\sigma$ electrons, Decembre 1967, Benjamin, in press.
8. — Int. J. quant. Chemistry S **2**, 187 (1968).
9. Roothaan, C. C. J.: Rev. mod. Physics **23**, 69 (1961).
10. Clementi, E., and D. R. Davis: J. comput. Physics **1**, 223 (1966).

11. Mely, B.: Thèse 3è cycle, Paris 1969.
12. Clementi, E., H. Clementi, and D. R. Davis: J. chem. Physics **46**, 4731 (1967).
13. Mulliken, R. S.: J. chem. Physics **36**, 3428 (1962).
14. Grimaldi, F., A. Lecourt, and C. Moser: Int. J. quant. Chemistry **1S**, 153 (1967).
15. Kurland, R. J., and E. B. Wilson: J. chem. Physics **27**, 585 (1957).
16. Costain, C. C., and J. M. Dowling: J. chem. Physics **32**, 158 (1960).
17. Basch, H., M. B. Robin, and N. A. Kuebler: J. chem. Physics **47**, 1201 (1967).
18. Robb, M. A., and I. G. Csizmadia: Theoret. chim. Acta (Berl.) **10**, 269 (1968).
19. Mulliken, R. S.: J. chem. Physics **23**, 1833, 1841 (1955).
20. Praud, L., P. Millié, and G. Berthier: Theoret. chim. Acta (Berl.) **11**, 169 (1968).
21. Omura, I., H. Baba, and K. Higasi: J. Physic. Soc. Japan **28**, 147 (1955).
22. Berthod, H., and A. Pullman: J. Chim. Physique **52**, 942 (1965).
23. – C. Giessner-Prettre, and A. Pullman: Theoret. chim. Acta (Berl.) **5**, 53 (1966).
24. Julg, A., and P. Carles: Theoret. chim. Acta (Berl.) **7**, 103 (1967).
25. Pilar, F. L., and J. R. Morris: J. chem. Physics **34**, 389 (1959).
26. Adam, W., and A. Grimison: Theoret. chim. Acta (Berl.) **7**, 342 (1947).
27. Giessner-Prettre, C., and A. Pullman: Theoret. chim. Acta (Berl.) **11**, 159 (1968).
28. Pullman, A., and M. Rossi: Biochim. biophysica Acta **88**, 211 (1964).
29. Lifschitz, C., E. D. Bergmann, and B. Pullman: Tetrahedron Letters (London) **46**, 4583 (1967).
30. De Voe, H., and I. Tinoco Jr.: J. molecular Biol. **4**, 500 (1962).

Dr. A. Pullman
Institut de Biologie Physico-Chimique
Laboratoire de Biochimie Quantique
13, rue Pierre et Marie Curie, Paris V$^e$, France