Theoret. chim. Acta (Berl.) 5, 53—68 (1966)

Theoretical Study of the Electronic Properties of the Purine and Pyrimidine Components of the Nucleic Acids

I. A Semi-Empirical Self-Consistent-Field Calculation*

HÉLÈNE BERTHOD, CLAUDE GIESSNER-PRETTRE and ALBERTE PULLMAN

Université de Paris, Institut de Biologie Physico-Chimique

Received February 28, 1966

A semi-empirical self-consistent field calculation has been carried out for the purine and pyrimidine bases of the nucleic acids after an appropriate optimization of the integral values using reference compounds. Special emphasis has been laid upon results concerned with the ionization potentials, electron affinities, spectral transitions, triplet states, dipole moments and basicities of the compounds. They enable in particular a detailed interpretation of both the individual ultra-violet absorption spectra of the bases and of their interrelations.

Un calcul self-consistent semi-empirique a été effectué pour les bases puriques et pyrimidiques des acides nucléiques, à l'aide d'un jeu d'intégrales déterminées de facon optimum pour une série de molécules de référence. L'accent a été mis en particulier sur les résultats relatifs aux potentiels d'ionisation, affinités électroniques, transitions $\pi \rightarrow \pi^*$, états triplets, moments dipolaires et basicités. Il a été possible d'interpréter tant les spectres individuels des bases que leurs interrelations.

Ein semiempirisches SCF-Verfahren wurde zur Berechnung von Purin- und Pyrimidinbasen von Nucleinsäuren herangezogen, wobei besonderes Gewicht auf Ionisierungspotentiale, Elektronenaffinitäten, spektrale Übergänge, Triplettzustände, Dipolmomente und Basizitäten gelegt wurde. Insbesondere ist eine eingehende Interpretation der UV-Absorptionsspektren möglich, und zwar sowohl hinsichtlich der einzelnen Basen als auch deren Beziehungen untereinander.

Introduction

The first theoretical calculations on the fundamental purine and pyrimidine bases of the nucleic acids have been carried out in the Hückel approximation of the method of molecular orbitals [*63*]. They yielded a great mass of information concerning the structure and properties of these molecules. Similar calculations have also been carried out for many related compounds [*66*]. On account of the well-known shortcomings of the Hückel approximation and in spite of many remarkable confirmations of the theoretical results by experiment, it was nevertheless highly desirable that more refined approximations be used for the study of such important compounds. A first calculation using the Pariser-Parr approximations in a self-consistent procedure has been carried out by VEILLARD and PULLMAN [*81*], with emphasis on the electronic indices characterizing the reactivities of the different positions, the relative basicities of the nitrogens and the relative order of the molecular ionization potentials and electron affinities. No real

---

* This work was supported by grant No. GM 12289-01 of the United States Public Health Service (National Institute of General Medical Sciences).

attempt was made at this time to predict the absolute values of the ionization potentials which are known to be overestimated by the usual approximation of the core integrals [*60, 64, 74*]. A preliminary evaluation of the appropriate correction has been made since, and values of the n and π ionization potentials were predicted on this basis for the four bases of the nucleic acids [*64*]. Since, however, the modification of the core integrals influences not only the values of the orbital energies but also the distribution of the electrons in heterocyclic molecules [*74*], we felt necessary to undertake a new study of the whole problem, starting with a careful scrutinization of the choice of integrals.

In the meantime, a study of the hypochromicity in the base-pairs of DNA has appeared [*53*], which used the Pariser-Parr approximation with partial configuration mixing, and during the preparation of this manuscript a Pariser-Parr self-consistent field calculation on the nucleic acid bases by a group of Japanese authors [*50*] has also been published. We shall discuss some aspects of these publications in the forthcoming pages.

The Method

In the present state of development of the methods of quantum chemistry, there is not yet a satisfactory procedure for treating simultaneously all the electrons of complex conjugated molecules*. Thus we have adopted the usual hypothesis of σ-π separation and dealt essentially with the system of π electrons in the field of the σ-core. However, for the calculation of dipole moments, we have utilized the approximate representation of the σ-electrons provided by the procedure of del Re [*10*], adapted for conjugated systems [*3*].

The molecular orbitals of the π electrons are written as a linear combination of atomic orbitals:

$$\varphi_i = \sum_r c_{ri}\, \lambda_r \tag{1}$$

and the best orbitals of this form compatible with a single Slater determinant representing the closed-shell ground state can be found by the classical SCF Roothaan equations [*69*]. If a set of orthogonal atomic orbitals λ obtained from the usual AO's by a Löwdin transformation [*39*] is used to build the MO's, a first-order approximation to the transformed Roothaan equations is given by**

$$\sum_q (F_{pq} - \delta_{pq}\, \varepsilon_i)\, c_{qi} \tag{2}$$

with

$$F_{pq} = I_{pq} + G_{pq}\,, \tag{3}$$

$$I_{pq} = \int \lambda_p\, H^{\text{core}}\, \lambda_q\, d\tau \tag{4}$$

$$G_{pp} = \sum_j \{c_{jp}^2\, (pp \mid pp) + \sum_{r \neq p} 2\, c_{jr}^2\, (pp \mid rr)\} \tag{5}$$

$$G_{pq} = -\sum_j c_{jp}\, c_{jq}\, (pp \mid qq)\,. \tag{6}$$

This set of equations is the same as the Pariser-Parr-Pople SCF equations

---

* A promising attempt in this direction together with a discussion of the non-satisfactory existing procedures has just appeared [*61*].

** Detailed discussion of this matter has been given in recent reviews: see for instance references [*58, 41, 62*].

obtained [*60*] upon inserting into the Roothaan equations the zero-differential-overlap hypothesis of PARISER and PARR [*56*].

In principle, all the integrals should be calculated using the orthogonal basis λ; in practice, however, on account of the "built-in" approximations [*62*] of the method (correlation error, frozen σ-core, valence-state σ core, truncated basis set of AO's, poor approximation of atoms-in-molecules AO's), satisfactory results can be obtained only if a series of empirical corrections are introduced, namely:

a) a strong reduction of the one-center coulomb integrals $(pp \mid pp)$;

b) a reduction of the two-center $(pp \mid qq)$ integrals, at least for nearest-neighbors;

c) an empirical choice of I_{pq} values so as to fit spectral data for reference compounds.

These are the main features of the Pariser-Parr approximation [*57*]. More often than not, when this procedure is adopted, the I_{pp} integrals are calculated as:

$$\alpha_p = I_p - \sum_A (A \mid pp) - \sum_{q \neq p} (qq \mid pp) \tag{7}$$

with or without neglect of the second term and using as I_p the valence state ionization potential of atom p. Although such a simplification may be of little importance in hydrocarbons, where it seems to influence merely the absolute values of the molecular ionization potentials and where the atomic π electron-populations are close to unity, this is no more the case in heteromolecules, as mentioned in the Introduction. There, an empirical choice of U_p values in place of the first two terms of Eq. (7) is strongly required. Yet, no systematic study of this parametrization was available when we started this work, so that one of our main concerns was to determine an appropriate set of empirical U_p values together with the remaining empirical integrals. Our first objective was to obtain a set of parameters which would reproduce as many observables as possible for a number of reference compounds using the Pariser-Parr-Pople self-consistent equations without configuration interaction, at least in a first stage. In order to satisfy this purpose, we proceeded as follows:

a) The one-center coulomb integrals for carbon, nitrogen and oxygen were given the values

$$\gamma_p^0 = (pp \mid pp) = I_p - A_p \tag{8}$$

according to the Pariser relation [*54*], where I_p and A_p are the valence state ionization potential and electron affinity of atom p. We used the values tabulated by J. HINZE and H. H. JAFFÉ [*23*] for the appropriate valence states. The resulting one-center integrals are listed in Tab. 1.

For π lone-pairs like those of the pyrollic nitrogen, we have adopted also relation (8) according to JULG's argument [*28*].

b) As concerns the two-center coulomb integrals $\gamma_{pq} = (pp \mid qq)$, we tentatively adopted the theoretical values calculated with the usual Slater orbitals for all pairs p, q of non-directly bonded atoms, since this does not seem to strongly affect the results [*40*]. As concerns the coulomb integrals for pairs of bonded atoms, it has been emphasized [*40, 48, 70*] that they affect the results essentially through their difference with respect to the one-center value adopted. The best fit for this difference in carbon compounds [*40, 55, 75*] yields a value of 6.9 eV for γ_{CC}

(1.39 Å) when γ_C^0 is put equal to the value of Tab. 1. This value would be obtained through the theoretical formulas in using Slater orbitals with an effective quantum number ζ' equal to 0.83 (instead of the Slater value 1.625). For adjacent atoms, all the γ_{CC} values have been calculated with this value of ζ'. For the heteroatoms we proceeded as follows: Calling ζ_0 the fictitious effective quantum number corresponding to the empirical values of the one-center integrals listed in Tab. 1, we have admitted that for each atom the ratio ζ'/ζ_0 is the same as for the carbon-carbon integrals and have calculated the γ_{XX} values theoretically by the Roothaan expressions using the ζ' values so obtained.

Moreover we have adopted the approximation

$$\gamma_{CN} = \tfrac{1}{2}\,(\gamma_{CC} + \gamma_{NN})\,. \tag{9}$$

It is to be noted that this choice of the Coulomb integrals does not correspond to a smooth interpolation curve from the γ_{13} values to the empirical γ^0 values, as in the original Pariser-Parr quadratic expression [*57*]. Although this discontinuous character of the reduced Coulomb integrals seems to be a necessity for the obtention of good transitions in carbon compounds [*16*, *55*, *75*], we felt compelled to assess its effect in a series of heterocycles. This has been done by using the original Pariser-Parr interpolation formula [*57*]. The transitions that we obtained with a smooth curve were 0.1 to 0.6 eV higher than the ones given in Tab. 2. In addition the agreement of the calculated dipole moments with the experimental values is much better when the broken γ_{pq}-curve is used instead of the smooth curve. On the other hand, the proton magnetic resonance spectrum of pyrrole [*71*] shows that the chemical shift (τ) of the ortho hydrogen is smaller than the one of the meta hydrogen. If one accepts the semiempirical calculations of τ developed by several authors [*17*, *49*], the π-electron density on the carbon bearing the hydrogen atom considered should increase along with the τ of the proton. Thus it suggests that the charge on the ortho carbon should be smaller than the one on the meta carbon atom; this order of the charges has been obtained only with the "broken curve" approximation for the γ_{pq}.

Table 1. *Integral values*

p	γ_p^0 (eV)	ζ'	K_{Cp}	U_p (eV)
C	11.13	0.83	17.238	− 9.5
N	12.34	0.92	14.913	−11.8
NH	12.34	0.92	15.195	− 9.9
NH_2	12.34	0.92	11.579	− 8.9
O	15.23	1.135	8.573	−14.5

It may be added that the integrals calculated with the orthogonalized basis λ [*39*] lead to a larger difference $\gamma_{11} - \gamma_{12}$ than the integrals calculated with the usual Slater atomic orbital basis, whereas the difference $\gamma_{11} - \gamma_{1p}$ (for $p \neq 2$) is practically the same in the two basis. Moreover, the exchange integrals are indeed very small in the λ basis, with the exception of those involving nearest neighbours [*44*, *59*]. Since all the exchange integrals are neglected in the method, the fitting of the γ_{pq} must make up for the error so introduced.

For all these reasons we did not adopt the smooth curve for our final integral choice.

c) The remaining parameters concern the one-center and two-center core integrals α_p and β_{pq}. We have adopted as a starting point the β_{CC} value, −2.39 eV, determined for benzene (C–C = 1.39 Å) and the β_{CN} value of −2.576 eV adjusted for *s*-triazine [*57*] (C–N = 1.34 Å) and we have evaluated β_{CO}, β_{C-NH} and β_{C-NH_2} by trial and error on formaldehyde, pyrrole and aniline. The R dependence of β_{pq} adopted has been $\beta = -K/R^6$, following a suggestion by Kon [*32*].

There remains the choice of the U_p values. As a starting point we adopted $U_C = -9.5$ eV which is known to give satisfactory values of the ionization potentials in hydrocarbons [*60*, *74*] and determined U_O, U_N, U_{NH} and U_{NH_2} so as to reproduce the ionization potentials of formaldehyde, pyridine, pyrimidine, pyrrole and aniline. Moreover in this trial-and-error procedure, we have attempted to reproduce at the same time the values of the dipole moments of the reference compounds, obtained by adding to the calculated π component, the σ component obtained in the way mentioned above.

The set of integral values which gave the best over-all agreement with as many properties as possible in the reference compounds considered is listed in Tab. 1, and the results concerning the reference compounds are given in Tab. 2. It is seen that as far as ground-state properties are concerned, the agreement with experiment is quite satisfactory. The agreement obtained for spectroscopic transitions is rather gratifying also.

Table 2. *Reference compounds*

	ionization pot. (eV)		dipole moment (D)		Singlet transitions (eV)		first triplet (eV)	
	th.	obs.	th.	obs.	th.	obs.	th.	obs.
Pyridine[a]	9.5	9.3 [*83*]	2.2	2.2 [*9*]	5.4 – 6.1 – 6.3	4.9 – 6.4 – 7.0 [*31*]	3.8	3.7 [*18*]
Pyrimidine	9.9	9.9 [*80*]	2.3	2.4 [*72*]	5.8 – 6.6 –	5.1 – 6.5 [*7*]	4.2	
Pyrrole[a]	9.0	8.9 [*80*]	1.8	1.8 [*4*]	6.0 – 6.3 – 7.0	5.9 – 6.8 – 7.2 [*27*]	3.8	
Aniline[a]	8.1	8.2 [*80*]	1.8	1.6 [*42*]	4.9 – 5.5 – 5.9	4.4 – 5.4 – 6.4 [*30*]	3.7	
Formaldehyde[a]	11.8	11.8 [*74*]	2.4	2.3 [*73*]	7.8	8.0 [*74*]	3.0	
Formamide	10.0	10.2 [*83*]	3.6	3.7 [*35*]	7.1 – 9.8 –	7.2 – 9.2 [*25*]	4.5	
2-pyridone[b]	8.3		3.8	~4 [*34*]	4.5 – 5.2 –	4.2 – 5.5 [*29*]	2.8	

[a] Molecules used for adjusting the integral values.

[b] The observed values correspond to N-methyl 2-pyridone.

Results and Dicussion

A. Ionization potentials and electron-donor properties

According to Koopmans' theorem [*33*] the negative of the energy of the highest filled molecular orbital (HOMO) obtained in the SCF calculation is an approximation to the first molecular ionization potential. The values of the HOMO for the four nucleic bases are given in the first column of Tab. 3.

It is seen that the numerical values obtained are definitely smaller than the previous values of Veillard-Pullman [*81*], of Nesbet [*53*], or of the Japanese workers [*50*]. This was to be expected as a result of our more appropriate choice of U_p values. As concerns the order of magnitude obtained for the ionization potentials, it seems to be now in the reasonnable region [*22*].

On the other hand, the order of decreasing ionization potentials shown by our present calculations is

$$U > C > A > G\,.$$

This order is the same as that previously obtained by VEILLARD and PULLMAN and by the other authors, and in fact is also the same as that predicted initially by the Hückel approximation [*63*]. In particular, in conformity with early predictions by PULLMAN and PULLMAN [*63, 65*] guanine should be the best π-electron donor among the four nucleic bases.

A confirmation of these results seems to be provided by recent measurements of the oxidation potentials (anodic waves) at the stationnary graphite electrode [*15*], which show that the two pyrimidines are not oxidizable, whereas the two purines are, and indicate moreover that guanine is oxidized more easily than adenine. Therefore as predicted by the theoretical calculations, guanine appears as the best electron-donor among the four bases. These results are also confirmed by studies on the charge-transfer complexes between the nucleosides and nucleotides of the bases or the bases themselves and chloranil [*14*].

Table 3. *Energies of the highest filled and lowest empty orbitals*

	HOMO (eV)	LEMO (eV)
Guanine	−7.59	+1.47
Adenine	−7.92	+1.52
Cytosine	−8.16	+0.87
Uracil	−9.15	+0.38

A comment is appropriate at this point concerning the values of the ionization potentials obtained by NAGATA et al. [*50*]. Not only are their ionization potentials quite generally appreciably higher than ours but, moreover, their calculations predict an ionization potential of thymine which should exceed that of uracil by some 1.6 eV. This is, of course, an unacceptable result as it is well-known that the methyl group has a much smaller effect on the ionization potentials (a few tenth of an eV), effect which moreover tends generally to decrease the potential of the parent compound. In fact the discrepancy in the results of the Japanese authors comes from a trivial error made in the chemical formula of uracil. They have considered as "uracil" throughout their calculations, a rare tautomeric lactim form of this molecule. This error leads them to difficulties in other fields also, so that at one occasion they study separately a molecule which they call "demethylated thymine" without apparently realizing that this is in fact the appropriate form of uracil that they should have considered.

B. Electron affinities and electron-acceptor properties

The values of the energies of the lowest empty molecular orbital (LEMO) give a measure of the electron affinities of the bases or at least of their electron acceptor capacities.

It can be seen from the energies of the molecular orbitals listed in Tab. 3 that the LEMO's appear, in the four nucleic bases, with a positive sign in our calculation as opposed to the negative sign obtained by other authors [*50, 53, 81*]. This is due to the general rising of the orbital energy values due to the U_p adjustement.

The relative order of the LEMO energies indicates that the pyrimidines should be better electron-acceptor than the purines, in agreement with VEILLARD-PULLMAN and the other SCF calculations, but in contradiction with the previous Hückel prediction according to which adenine and cytosine should be better acceptors than guanine and uracil [*65*].

It is practically impossible to have a direct check of the exactness of the absolute values of these energies since the measurements of molecular electron affinities are practically inexistent and moreover their exact relation to the LEMO energies is not entirely clear [*15*, *79*]. However, some informations on the relative electron acceptor properties of these molecules are available, through the measurements of the polarographic reduction potentials at the dropping mercury electrode. For the compounds considered here, it is found [*79*] that adenine and cytosine are reducible whereas guanine and uracil are not.

It appears thus that in this respect the results of the Hückel calculation are more in agreement with the experimental data than the self-consistent results. It is, however, difficult to draw definite conclusions on this point, on account of the highly complex character of the reactions which may be determinant in the reduction processes.

C. Spectroscopic properties

1. The positions and intensities of the bands. The near ultra-violet absorption spectra of the nucleic acid bases and of their derivatives have been examined rather carefully by different authors [*1*, *12*, *46*]. The spectra in aqueous solution have been extended down to 185 mμ by VOET et al. [*82*] and a similar study has been performed [*7*] by CLARK and TINOCO in an attempt to classify and systematize the transitions. This has been very recently completed by a careful examination of the vapor spectra of some of the bases [*6*].

For the sake of the comparison with the theoretical computations, the most useful results are those of CLARK and TINOCO which establish correlations among the spectra of the bases. Thus, it is advantageous for a theoretical investigation, and in particular for one carried out in the framework of the "refined" approximations, to be able not only to reproduce through calculations the characteristics of the spectra of isolated molecules, but to put into evidence or to confirm general correlations involving such characteristics among series of related compounds.

In fact, in the present case, the comparison of the experimental results with the theoretical ones is quite satisfactory from both points of view. Thus, CLARK and TINOCO classify the observed transitions into three groups that they name B_{2u}, B_{1u} and E_{1u} by analogy to the corresponding nomenclature of the bands in benzene. For the sake of convenience, we shall retain this nomenclature although the following discussion implies nothing concerning the postulated relationship to the benzene spectrum. Now, a careful examination of the positions of these bands in the four nucleic bases shows the possibility of a further systematization:

a) In all four molecules the B_{2u} band corresponding to the longest wave-length of absorption is situated between 260 and 275 mμ (4.76 to 4.51 eV), with guanine and cytosine absorbing towards the longer wave-length limit and adenine and uracil absorbing towards the shorter wave-length limit. In the vapour phase [*6*] this distinction between G, C and A, U is still emphasized since the first absorption maxima of guanine and cytosine are shifted to 293 and 290 mμ respectively

(4.2 – 4.3 eV) while the corresponding bands of uracil and adenine are shifted to 244 – 252 mμ (5.1 – 4.9 eV).

b) On the other end of the observed spectrum (the E_{1u} region) the four molecules exhibit two absorption maxima, one at 202 – 208 mμ (6eV) and one around 180 – 188 mμ (6.8 – 6.6 eV).

c) In between these extremes, there is the B_{1u} band which is not observed, however, in all molecules; thus it exists as a shoulder in the spectrum of cytosine at 237 mμ (5.2 eV) and of guanine at 250 mμ (5 eV) while no comparable peak appears neither in adenine nor in uracil. It must be added however that experiments in optical rotatory dispersion suggest the existence of a weak hidden band at about 230 mμ (5.4 eV) in uracil [*7*]. Finally a suggestion must also be mentionned, according to which the first band in adenine is in fact a composite one [*46*] consisting of two superimposed transitions, a hypothesis supported by measurements of the dichroism in the crystal spectrum of 9-methyladenine [*77*] and more recently by studies of polarized fluorescence [*5*]:

The examination of our theoretical findings shows a far-reaching agreement with these general results and classifications (Tab. 4):

a) As concerns the transition of longest wavelength, the calculations distinguish clearly two groups among the four bases: guanine and cytosine, absorbing toward longer wavelength at 4.4 and 4.2 eV respectively, and adenine and uracil absorbing towards somewhat shorter wavelength, at 5.2 – 5.1 eV. Not only is this division in agreement with the general features of this absorption but the calculated values of the transitions agree very satisfactorily with the ones observed in the most recent, vapour phase measurements.

b) Leaving aside for the moment the next calculated band, we find then for the four bases a transition in the range of 5.9 to 6.1 eV, clearly corresponding to the first band in the E_{1u} region. Then, comes a group of rather closely packed calculated transitions which may reasonnably be correlated with the second E_{1u} absorption for which they suggest a composite structure: in adenine, the 6.3 eV component would be too weak to be observable but the 6.4 and 6.9 eV transitions are reasonably intense; in cytosine the 6.5 and 7.2 eV components are very weak so that the observed band could be essentially the 6.8 eV transition; a similar situation occurs in guanine whereas uracil shows a strong calculated band at 6.3 eV and a weak one at 7.2 eV.

c) As to the intermediate region, we find a second transition in the range of 5.3 to 5.5 eV for all four compounds. This transition corresponds clearly to the B_{1u} maximum observed in guanine and cytosine. Interestingly enough, the calculated intensity of this theoretical band is exceptionally very low in uracil in accordance with the absence of an observable maximum in this region for this compound.

As to the case of adenine, which is also devoid of an observable band in this region, it seems extremely tempting to consider the calculated transition as representing the second component of the first, composite band system, although its calculated intensity may seem rather high.

In general the calculated values of the oscillator strength in the four compounds considered here are in fair agreement with the empirical values with the possible exception of the first band of uracil which appears too intense theoretically. A

Table 4. *Transitions in the nucleic bases* (from ground state)

		experimental		theoretical				
		ΔE^a (eV)	f^b	$i \rightarrow j^c$	ΔE(eV)	f	$\theta^{\circ d}$	triplets (eV)
Guanine Experimental values (except vapour) from 9 ethyl guanine	B_{2u}	4.5 (4.2 – 4.4)	0.1	7 → 8	4.4	0.4	+104	3.2
	B_{1u}	4.9	0.3	7 → 9	5.4	0.6	−166	3.7
	E_{1u}	6.0	...	7 → 10	6.0	0.1	− 59	5.0
				6 → 8	6.3	0.1	−101	5.3 (6 → 9)
		6.6	1.1^e	5 → 8	6.6	0.2	+179	5.8
				6 → 9	6.8	0.1	−140	6.0 (4 → 8)
Cytosine	B_{2u}	4.5 (4.3)	0.2	5 → 6	4.2	0.1	+ 85	3.1
	B_{1u}	5.2	0.2	4 → 6	5.3	0.4	+164	3.9
	E_{1u}	6.1	0.6	5 → 7	6.0	0.5	+133	4.1
				4 → 7	6.5	0.05	− 90	5.7 (5 → 8)
		6.7	...	3 → 6	6.8	0.3	−147	5.8
				5 → 8	7.2	0.03	− 71	5.9 (4 → 7)
Adenine	B_{2u}	4.8 (4.9)	0.3	6 → 7	5.2	0.4	−144	3.7
				6 → 8	5.5	0.3	− 58	4.5
	B_{1u}	...	...	...	...	...	...	
	E_{1u}	6.0 (6.0)	0.4	5 → 7	6.1	0.5	− 78	5.1
				6 → 9	6.3	0.01	−153	5.4 (5 → 8)
		6.7	...	4 → 7	6.4	0.2	− 35	5.6
				5 → 8	6.9	0.2	+ 10	6.7 (4 → 8)
Uracil	B_{2u}	4.8 (5.1)	0.2^f	5 → 6	5.1	0.7	+174	2.7
	B_{1u}	(see text)	...	4 → 6	5.5	0.02	−154	4.4 (4 → 7)
	E_{1u}	6.1 (6.0)	0.2	5 → 7	5.9	0.2	−115	4.8 (4 → 6)
				4 → 7	6.3	0.5	+135	5.0 (5 → 7)
		6.8 (6.6)	...	5 → 8	7.2	0.01	−166	6.0

---

[a] Classification and values from reference [*7*]. In parenthesis, vapor phase values from reference [*6*].

[b] From reference [*12*].

[c] Orbital jump.

[d] Measured counterclockwise from the line C_4–C_5 in purines and N_1–C_4 in pyrimidines.

[e] Value obtained from the area of the envelope of the two E_{1u} bands [*12*].

[f] Evaluated from the spectrum of reference [*7*].

more complete discussion of the *f* values will be presented in a forthcoming paper relating the results of the inclusion of configuration interaction.

Before leaving this subject, we would like to comment on the corresponding calculation by NAGATA et al. [*50*]. These authors have given the first three calculated singlets for adenine, cytosine, guanine, uracil and thymine. Their agreement with experiment is satisfactory for the first bands of adenine and cytosine only. No detailed systematic comparison with the other bands has been carried out, but the values reported in the table seem less satisfactory than those of the present calculation. On the other hand the value obtained for the first transition in uracil is far too low in energy (4.2 eV). As already mentioned, this is not surprising since the calculations have been made for a rare tautomeric form of uracil. In fact, the calculated value which should be taken for uracil is the value 6.4 eV

(198 mμ), reported by the authors for what they call "demethylated thymine". It is seen that this is not satisfactory either.

Another point must be stressed in connection with the problem of spectroscopic transitions. In their paper the Japanese authors compare their SCF results for the longest wavelength of absorption of the bases with those of a Hückel calculation by Ladik and Hoffman [*36*] and conclude that the SCF treatment improves "strikingly" the results. The choice of the Ladik-Hoffman values is, however, particularly infortunate for such a comparison since they have been obtained with use of an erronous value of the Hückel resonance integral β, a fact recognized by Ladik and Hoffman themselves [*24*]. When a reasonable value of β, say 3.26 eV [*26*], is utilized, the order of magnitude of the Hückel longest wavelengths of absorption of the bases falls well within the range of the observed values, around 4.5 eV.

It is perfectly understood that the Hückel approximation is generally not very suitable for calculating spectroscopic transitions, particularly in absolute values. The misuse of a parameter should not, however be confused with the shortcomings of the method itself.

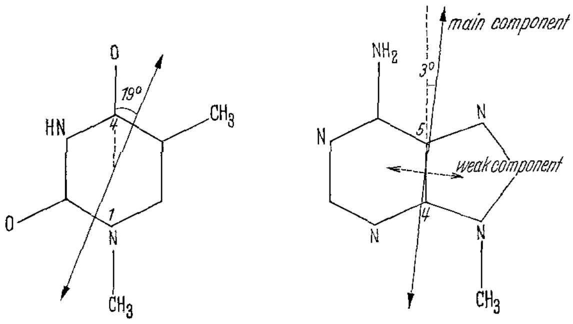


Fig. 1. Polarization of the first transitions in 1-methylthymine and 9-methyladenine (after Stewart et al. [*77*, *78*])

2. *The directions of polarization.* The direction of polarization of the transitions have been studied experimentally in some details for 9-methyladenine and 1-methylthymine [*77*, *78*]. The results, supplemented by a study of the A–T dimer indicate that the first band of 1-methylthymine is polarized at 19° degrees with respect to the N_1–C_4 axis, (Fig. 1) the polarization of the second absorption below 240 mμ being qualified as "reversed". For 9-methyladenine, the major component of the first long wavelength transition is at 3° with respect to the C_4–C_5 line (transverse axis) and the second weaker component is suggested to be polarized at right angle to the first one (Fig. 1).

Moreover, the absorption band below 240 mμ for 9-methyladenine is assigned a longitudinal polarization. This results for 9-methyladenine are in partial disagreement with the previous assignements of Mason [*46*] who advocated a strong longitudinal component in the first band of adenine with a small contribution of a transversely polarized band. As to the other compounds, the only available experimental result indicates that the first two bands (at 275 and 230 mμ) in 9-ethylguanine are polarized at right angles to each other [*5*].

The calculated directions of polarization are given in Tab. 4 and are drawn in Fig. 2 for the first three transitions in each base. Insofar as the experiments on methylthymine are an indication on the situation in uracil, the theoretical values are in satisfactory agreement with experiment. As concerns adenine, the theoretical transitions that we have assigned to the two components of the 1st band are indeed mutually perpendicularly polarized with the first component closer to the transverse axis than to the longitudinal one. The second band is long-axis polarized.

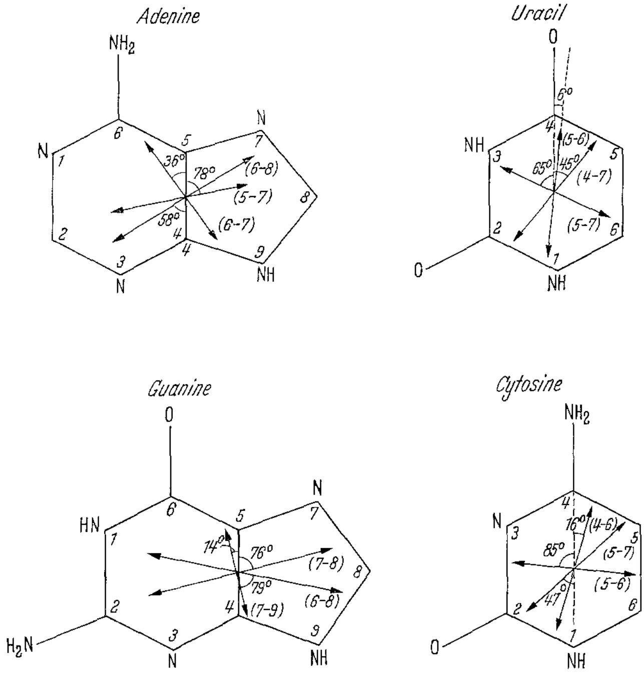


Fig. 2. Direction of polarization (theoretical) of the first transitions in the bases

Guanine and cytosine may be grouped together with a behaviour opposed to that of the other two bases: their first transition is perpendicular to the axis C_4–C_5 (short axis) or N_1–C_4 whereas the moment of the second band is along this axis.

It is interesting to mention that the early Hückel calculations of de Voe and Tinoco [*12*] indicated the same opposed behavior of cytosine and uracil as concerns the directions of polarization of the first two bands, and also as concerns the longitudinal character of the first transition in guanine. Their result for adenine (1st band longitudinal) which fitted Mason's conception is, however, not in accord with the present calculation which seems more compatible with the more recent measurements on the crystal spectrum.

3. *The position of triplet states.* We have indicated in Tab. 4 the calculated positions of the triplets corresponding to the previously discussed singlets in the four nucleic bases. Although a more precise location of the triplet states must await the results of the inclusion of configuration superposition, we would like to point out some interesting features of the present results: a) the order of increasing

energy of the first excited triplet with respect to the fundamental state is:

$$U < C < G < A\,.$$

The triplets of the pyrimidines are thus situated lower than those of the purines, with the triplet of uracil being the lowest of all. b) the singlet-triplet splitting is the smallest in cytosine followed closely by guanine, and the largest in uracil.

From the experimental viewpoint, precise results concerning the triplet states in these molecules are still rather scarce. The phosphorescence of the purine bases, their nucleosides and nucleotides and of DNA has been studied and characterized by different authors [*2*, *13*, *38*, *68*, *76*]. The most recent and most precise measurements concern the first triplet state of adenine [*8*], the maximum of which is at 3.05 eV with a singlet triplet separation of 1.7 eV. This last result is in satisfactory agreement with the calculated value of 1.5 eV. The comparison of guanine with adenine is difficult on account of the closeness of the emitting regions [*2*]. LONGWORTH [*38*] locates the phosphorescence maximum of guanine at a lower energy than the corresponding maximum of adenine, while determinations by HELÈNE [*19*] indicate practically no difference between the two molecules from that point of view. On the other hand, phosphorescence studies [*21*] in dinucleotides containing guanine and adenine seem to indicate triplet-triplet energy transfer from guanine to adenine thus suggesting the locating of the adenine triplet below that of the guanine one, at least in the experimental context. The results of the configuration-interaction calculations will be particularly interesting in this connection.

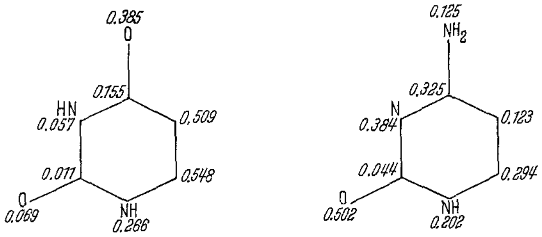
As concerns the triplets of the pyrimidine bases, their identification is rendered difficult by the very low phosphorescence of these compounds [*2*] which is observable only at relatively high concentrations and at 77 °K [*20*]. On the basis of studies of the phosphorescence of the isolated compounds [*19*] and of energy transfer in dinucleotides [*20*, *21*], DOUZOU et al. [*20*] suggest that the first triplets of the pyrimidines are lower than those of the purines and that uracil has the lowest triplet among the four bases, a result in agreement with our predictions.

D. Spin densities in the first triplet state

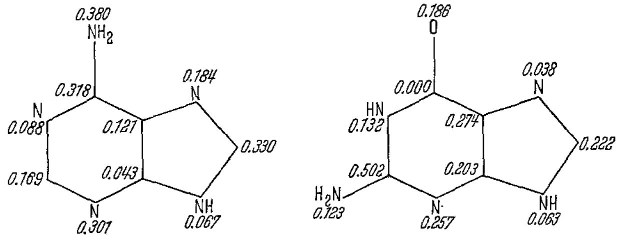
Fig. 3 gives the distribution of the unpaired electrons in the two singly occupied orbitals of the first excited state of the four bases. In fact, since we use here the virtual orbitals of the ground state calculations for the building of the excited states, we cannot distinguish between a singlet and the corresponding triplet as far as electron distribution is concerned. But it is considered that the distribution of the unpaired electrons may give a first approximation to the spin density distribution in the first triplet.

The most significant results concern the molecule of uracil for which they indicate a very high concentration of lone-electrons on the carbons C_5 and C_6, by far the highest density among all the atoms of the molecule. Moreover the bond-order of the C_5–C_6 bond which was very large in the ground state, becomes relatively very small (0.377) in the first excited state, whereas the other bonds of the molecule are little affected by the excitation. This situation indicates that the electronic excitation is localized in this molecule, to a large extent, in the C_5–C_6 bond and this result is in agreement with the similar findings of MANTIONE and PULLMAN [*45*] made in the limits of the Hückel approximation and which they have related to the ability of thymine and some of its analogues to undergo photodimerization. For cytosine we find a much smaller concentration of lone electrons at the C_5–C_6 bond. In fact, we obtain the highest concentration of lone electron on the oxygen, followed by N_3 and C_4, the C_6 carbon coming next.

Also, the bond order of the C_5–C_6 bond, very high in the ground state, is not particularly decreased in the excited state. These results are in agreement with the fact that in contrast to thymine, cytosine does not dimerize easily.



Uracil Cytosine



Adenine Guanine

Fig. 3. Distribution of the unpaired electrons in the two singly occupied orbitals of the first excited state of the nucleic bases

As far as adenine is concerned, the odd electrons appear to be concentrated mainly on the extracyclic nitrogen, and on carbon 8 of the imidazole ring with appreciable densities also on C_6 and N_3 of the pyrimidine ring. In guanine, the maximum concentration of the odd electrons is largely displaced towards the benzene ring with a maximum at C_2. These results are similar to those of the Hückel approximation [*67*]. There is no great concentration of the lone electrons on any particular bond, a situation which agrees, in connection with the aforementioned consideration, with the absence of photodimerization of purines.

E. Dipole moments

The calculated values of the dipole moments obtained for the four bases are given in Tab. 5. The agreement with the known experimental values is satisfactory although the value obtained for adenine is a little too low. In agreement with previous calculations [*3*, *11*] we predict high values of the moments for guanine and cytosine.

Table 5

	μ_π	μ_{total}	μ_{exp} [*11*]
A	2.0	2.3	3.0
U	3.3	4.0	3.9
G	6.1	7.2	
C	5.4	7.1	

F. Electronic charges and basicities

The distribution of the π-electronic charges may be used for the study of a number of properties of the purine and pyrimidine bases. We shall not discuss this problem in any detail because the overall distribution obtained here is similar to those obtained previously in our laboratory and so the discussion would follow closely the previous ones [*66, 81*]. We feel, however, the need of considering more closely one point.

In agreement with Pullman-Veillard [*81*] we find for the relative order of the electron densities on the nitrogens $N_7 > N_3 > N_1$, in adenine and $N_3 > N_7$ in guanine. Nagata et al. [*50*] obtain a different order for adenine, namely $N_1 > N_3 > N_7$ and they invoke, in favor of the correctness of their results, the fact that protonation occurs at the N_1 position of this molecule. The argument may, however, be easily returned against its authors. Thus, protonation of guanine is well known to occur at N_7 [*37*] in agreement with an early theoretical prediction [*52*], while the electron density of Nagata et al. is much larger at N_3 than at N_7 for this molecule.

Table 6

	N_1	N_3	N_7[a]
A	−1.68	−1.60	−1.20
G		−1.14	−1.60

[a] With the appropriate correction related to a pentacyclic nitrogen [*51, 81*].

As a matter of fact, it has been shown a number of years ago [*51*], that the basicities in heterocycles containing more than one nitrogen atom cannot be considered as being a function of the charge densities only, but vary rather like:

$$\sum_p Q_p\,(dd \mid pp)$$

where Q_p is the net charge of atom p and d is the lone-pair orbital of the nitrogen considered. The values of this quantity for adenine and guanine are given in Tab. 6. They give the order

$$N_1 \approx N_3 > N_7$$

for the basicities of the nitrogens in adenine, and

$$N_7 > N_3$$

for guanine and are thus in general agreement with experiment.

References

[*1*] Beaven, G. H., E. R. Holiday, and E. A. Johnson: The nucleic acids. Ed.: Chargaff, E., et J. N. Davidson. Vol. I, 493. New York: Academic Press 1955.
[*2*] Bersohn, R., and I. Isenberg: J. chem. Physics **40**, 3175 (1964).
[*3*] Berthod, H., and A. Pullman: J. Chim. physique **55**, 942 (1965).
[*4*] Buckingham, A. D., B. Harris, R. J. W. LeFèvre: J. chem. Soc. (Lond.) **1953**, 1626.
[*5*] Callis, P. R., E. J. Rosa and W. T. Simpson: J. Amer. chem. Soc. **86**, 2292 (1964).
[*6*] Clark, L. B., G. G. Peschel, and I. Tinoco Jr.: J. physic. Chem. **69**, 3615 (1965).
[*7*] —, and I. Tinoco Jr.: J. Amer. chem. Soc. **87**, 11 (1965).
[*8*] Cohen, B. J., and L. Goodman: J. Amer. chem. Soc. **87**, 5487 (1965).
[*9*] Cumper, C. W. N., and A. I. Vogel: J. chem. Soc. (Lond.) **1956**, 3621.
[*10*] del Re, G.: J. chem. Soc. (Lond.) **1958**, 4031.
[*11*] de Voe, H., and I. Tinoco Jr.: J. mol. Biol. **4**, 500 (1962).
[*12*] — — J. mol. Biol. **4**, 518 (1962).
[*13*] Douzou, P., J. Francq, M. Hauss, and M. Ptak: J. Chim. physique **58**, 926 (1961).

[*14*] Duchesne, J., P. Machmer, and M. Read: C. R. Acad. Sci. **260**, 2081 (1965).
[*15*] Elving, P. J., W. A. Struck, and D. L. Smith: (In press).
[*16*] Forster, L. S., and K. Nishimoto: J. Amer. chem. Soc. 87, 1459 (1965).
[*17*] Fraenkel, G., R. E. Carter, A. McLachlan, and J. H. Richards: J. Amer. chem. Soc. 82, 5846 (1960).
[*18*] Goodman, L., and R. W. Harrell: J. chem. Physics **30**, 1131 (1959).
[*19*] Hélène, C.: Thèse, Paris 1966.
[*20*] —, P. Douzou et A. M. Michelson: Bioch. Bioph. Acta **109**, 261 (1965).
[*21*] — — — Proc. Natl. Acad. Sc. U.S. **55**, 376 (1966).
[*22*] —, R. Santus, and P. Douzou: Photochem. and Photobiol. **5**, 127 (1966).
[*23*] Hinze, J., and H. H. Jaffé: J. Amer. chem. Soc. 84, 540 (1962).
[*24*] Hoffman, T. A., and J. Ladik: Advances in chem. Physics VII, 84 (1964).
[*25*] Hunt, H. D., and W. T. Simpson: J. Amer. chem. Soc. **75**, 4540 (1953).
[*26*] Isenberg, I., and A. Szent-Gyorgyi: Proc. Natl. Acad. Sc. U.S. **45**, 519 (1959).
[*27*] Jaffé, H. H., and M. Orchin: Theory and application of ultraviolet spectroscopy. New York: Wiley 1963.
[*28*] Julg, A.: J. Chim. physique **55**, 413 (1958).
[*29*] Katritzky, A. R.: Physical methods in heterocyclic chemistry. Vol. II. New York and London: Academic Press 1963.
[*30*] Kimura, K., and S. Nagakura: Mol. Physics **9**, 117 (1965).
[*31*] Klevens, H. B., and J. R. Platt: Tech. Rep. Univ. of Chicago, part I p. 145 (1953—54).
[*32*] Kon, H.: Bull. chem. Soc. Jap. **28**, 275 (1955).
[*33*] Koopmans, T. A.: Physica **1**, 104 (1933).
[*34*] Krackov, M. H., C. M. Lee, and H. G. Mautner: J. Amer. chem. Soc. **84**, 892 (1965).
[*35*] Kurland, R. J., and E. B. Wilson: J. chem. Physics **27**, 585 (1957).
[*36*] Ladik, J., and T. A. Hoffman: Biopolymers, Symposia **1**, 117 (1963).
[*37*] Lawley, P. D., and P. Brookes: Biochem. J. **89**, 127 (1963).
[*38*] Longworth, J. W.: Biochem. J. **84**, 104 P (1962).
[*39*] Löwdin, P. O.: J. chem. Physics **18**, 365 (1950).
[*40*] Lykos, P. G.: J. chem. Physics **35**, 1249 (1961).
[*41*] — Adv. in quantum chemistry **1**, 171 (1964).
[*42*] McClellan, A. L.: Tables of experimental dipole moments. San Francisco and London: W. H. Freeman and Co. 1963.
[*43*] Machmer, P., et J. Duchesne: Nature **206**, 618 (1965).
[*44*] McWeeny, R.: Proc. Roy. Soc. **A 227**, 288 (1955).
[*45*] Mantione, M. J., and B. Pullman: Bioch. Bioph. Acta **91**, 387 (1964).
[*46*] Mason, S. F.: J. chem. Soc. (Lond.) **1954**, 2071.
[*47*] — J. chem. Soc. (Lond.) **1960**, 219.
[*48*] Murrell, J. N., and L. Salem: J. chem. Physics **34**, 1914 (1961).
[*49*] Musher, J.: J. chem. Physics **37**, 34 (1962).
[*50*] Nagata, C., A. Imamura, Y. Tagashira, and M. Kodama: Bull. Chem. Soc. Japan **38**, 1638 (1965).
[*51*] Nakajima, T., and A. Pullman: J. Chim. physique **1958**, 793.
[*52*] —, and B. Pullman: Bull. Soc. Chim. France **1958**, 1502.
[*53*] Nesbet, R. K.: Biopolymers, Symposia **1**, 129 (1964).
[*54*] Pariser, R.: J. chem. Physics **21**, 568 (1953).
[*55*] — J. chem. Physics **24**, 250 (1956).
[*56*] —, and R. G. Parr: J. chem. Physics **21**, 466 (1953).
[*57*] — — J. chem. Physics **21**, 767 (1953).
[*58*] Parr, R. G.: Quantum theory of molecular electronic structure. New York: Benjamin 1963.
[*59*] Peradejordi, F.: Cahiers de physique **17**, 158 (1963).
[*60*] Pople, J. A.: Trans. Far. Soc. **49**, 1375 (1953).
[*61*] —, D. P. Santry, and G. A. Segal: J. chem. Physics **43**, S 129 (1965).
[*62*] Pullman, A.: Molecular Biophysics. p. 81. New York: Acad. Press 1965.
[*63*] —, and B. Pullman: Bull. Soc. chim. France **1958**, 766; **1959**, 591.
[*64*] —, et M. Rossi: Biochim. biophysica Acta **88**, 211 (1964).

[65] PULLMAN, B., and A. PULLMAN: Proc. Natl. Acad. Sci. U.S. **44**, 1197 (1958).
[66] — — Quantum Biochemistry. New York: Wiley-Interscience 1963.
[67] — Molecular Biophysics, Proceedings of the Squaw Valley Summer School, p. 117. New York: Academic Press 1965.
[68] RAHN, R. O., J. W. LONGWORTH, J. EISINGER and R. G. SHULMAN: Proc. Natl. Ac. Sc. U.S. **51**, 1299 (1964).
[69] ROOTHAAN, C. C. J.: Rev. mod. Physics **23**, 69 (1959).
[70] RUEDENBERG, K.: J. chem. Physics **34**, 1861 (1961).
[71] SCHAEFFER, T., and W. G. SCHNEIDER: J. chem. Physics **32**, 1224 (1960).
[72] SCHNEIDER, W. C.: J. Amer. chem. Soc. **70**, 627 (1948).
[73] SHOOLERY, J. N., and A. H. SHARBAUGH: Physic. Rev. **82**, 95 (1951).
[74] SIDMAN, J. W.: J. chem. Physics **27**, 429 (1957).
[75] SIMMONS, H. E.: J. chem. Physics **40**, 3554 (1964).
[76] STEELE, R. H., and A. SZENT-GYORGYI: Proc. Natl. Acad. Sci. U.S. **43**, 477 (1957).
[77] STEWART, R. F., and N. DAVIDSON: J. chem. Physics **39**, 255 (1963); Biopolymers, Symposia 1, 465. New York: Interscience 1964.
[78] —, and L. H. JENSEN: J. chem. Physics **40**, 2071 (1964).
[79] STRUCK, W. A., and P. J. ELVING: J. Amer. chem. Soc. **86**, 1229 (1964).
[80] Tables of ionization potentials of molecules and radicals. Monograph Series of the Research Institute of Applied Electricity, Hokkaido Univ. Nr. 4 Appendix 1956.
[81] VEILLARD, A., and B. PULLMAN: J. theor. Biol. **4**, 37 (1963).
[82] VOET, D., W. B. GRATZER, R. A. COX, and P. DOTY: Biopolymers **1**, 193 (1963).
[83] WATANABE, K., T. NAKAYAMA, and J. MOTTL: J. Quant. Spectrosc. Rad. Transf. **2**, 369 (1962).

Prof. ALBERTE PULLMAN
Université de Paris
Institut de Biologie Physico-Chimique
13, rue Piere Curie, Paris 5e (France)