Theoret. chim. Acta (Berl.) 4, 132—144 (1966)

Quantum Chemistry Group for Research in Atomic, Molecular and Solid-State Theory
Uppsala University, Uppsala, Sweden

Pariser-Parr-Pople Calculations on Different DNA Constituents*

By

J. Ladik** and K. Appel

A semiempirical SCF LCAO MO-CI calculation has been performed for the nucleotide bases. According to the results we obtain a better agreement with the experimental singlet excitation energies, if we use the SCF eigenvectors of the Pople-type matrix in the CI calculation instead of the eigenvectors obtained after the first iteration step. On the basis of some parameter variation a set of integrals has been determined which yields as largest deviation between the experimental and theoretical excitation energies for the first two intensive excitations of the five bases (A, T, G, C and U) 1 eV. The possibilities of further improvements are discussed.

In the second part of the calculations the oscillator strength values (f) of the G-C base pair and of the GpG, GpC, CpG and CpC dinucleotides have been determined using the first 16 singlet excited configurations of these composite systems for the CI calculation. The comparison of the results obtained with the appropriate f values of the constituent single bases shows a hypochromicity of the first absorption band system in the case of all the four dinucleotides and a slight hyperchromicity in the case of the G-C base pair.

Für fünf Nukleotidbasen wurden semi-empirische SCF MO-CI Rechnungen angestellt, wobei sich nach Variation bestimmter Parameter ein Satz von Integralen ergab, der für die ersten beiden starken Banden aller untersuchten Basen eine Abweichung von höchstens 1 eV liefert. Ferner wurden die Oszillatorenstärken des G-C-Basenpaares sowie der GpG, GpC, CpG und CpC Dinukleotide bestimmt. Dabei wurden die ersten 16 einfach angeregten Zustände in die Konfigurationswechselwirkung eingeschlossen. Ein Vergleich dieser f-Werte mit denen der einfachen Basen zeigt hypochromes Verhalten der ersten Absorptionsbanden aller vier Dinukleotide, während im Falle des G-C-Basenpaares ein schwacher hyperchromer Effekt auftritt.

Un calcul semi-empirique SCF LCAO MO-CI a été effectué pour cinq nucléotides. Par variation de certains paramètres on obtient un jeu d'intégrales qui conduit a des énergies d'excitations pour les deux premières bandes intenses de chacune de ces bases (A, T, G, C et U) ne déviant plus qu' 1 eV. De plus, les forces oscillatrices de la paire de bases G-C et des dinucléotides GpG, GpC, CpG et CpC sont calculées, les 16 premièrs états singulets excités étant compris dans le calcul CI. Le comparaison de ces f avec ceux des bases constituantes montre un effect hypochrome du premier système de bandes pour les quatre dinucléotides et un faible effet hyperchrome pour la paire G-C.

Introduction

In a previous paper [*8*] we have reported some Pariser-Parr-Pople type calculations [*13*, *14*] on uracil (U). In the case of U a good agreement with the

---

* Sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture Knut and Alice Wallenberg's Foundation, and in part by the U.S. Public Health Service Research Grant No. CA 06850-01 to Uppsala University.

** Permanent Address: Central Research Institute for Chemistry of Hungarian Academy of Sciences, Budapest.

first two experimental excitation energies has been achieved by performing a limited CI between the singlet excited states using the wave functions which we have obtained as the eigenvectors of the Pople-type matrix after the first iteration step. This was, however, not the case, if we performed the limited CI with the SCF eigenvectors. To be able to decide, whether this is the general situation with pyrimidine and purine type molecules, or whether it is only accidentally so for U, it seems to be interesting to extend these calculations also to the other nucleotide bases: thymine (T), cytosine (C), adenine (A) and guanine (G). Further we intend to investigate the effect of the variation of the offdiagonal elements of the core Hamiltonian ($\beta_{i,j}$ integrals) on the results. The present calculations have been performed with three sets of values for the $\beta_{i,j}$ integrals.

It should be mentioned that Veillard and Pullman [*18*] previously have carried out a semiempirical SCF LCAO MO calculation for the ground states of the nucleotide bases and some related compounds. Since they have approximated the different integrals occuring in the Pariser-Parr-Pople method in a slightly different way than we have done, it seems interesting to compare their resulting charge distributions with ours. Further, Nesbet [*12*] has recently published a semiempirical SCF LCAO MO calculation with limited CI for the excited states of the nucleotide bases. It is also interesting to compare his results with ours.

In the second part of the calculations we have performed alimited CI calculation between the singlet excited states of the G-C base pair and the GpG, CpC, GpC and CpG dinucleotides*. In these calculations we have stopped the iteration procedure of the Pople-type matrix after the first iteration step and we have performed the CI with the eigenvectors obtained in this way. From the results we have calculated not only the excitation energies, but also the oscillator strength values (*f*) of the excitations of these systems. Comparing the latter with the appropriate *f*-values of the single bases, we could try to interpret the hypochromicity of the dinucleotides also on the basis of calculations, which assume only overlap type (and no exciton type [*17*]) interactions between the superimposed bases of DNA. Thus this part of our work is a refinement of a previous calculation, in which an attempt has been made to attack the problem of hypochromicity of polynucleotides on the basis of Hückel calculations performed on dinucleotides [*6*].

Method

First we have solved the eigenvalue problems of the Hückel matrices of the single nucleotide bases and of the GpG, CpC, GpC and CpG dinucleotides (for the SCF LCAO MO calculation of the G-C base pair see a previous publication [*15*]). The used α_i and $\beta_{i,j}$ values for the single bases are given in ref. [7] with the exception of $\beta_{\mathrm{C,N}}$ and $\beta_{\mathrm{C,O}}$. For $\beta_{\mathrm{C,N}}$ we have used the values $\beta_{\mathrm{C,N}} = 0.80\,\beta, 1.25\,\beta$, and for $\beta_{\mathrm{C,O}}$ the values $\beta_{\mathrm{C,O}} = 1.30\,\beta, 2.00\,\beta, 1.50\,\beta$, respectively. For G-C and for the dinucleotides we have used only the values $\beta_{\mathrm{C,N}} = 0.90\,\beta$ and $\beta_{\mathrm{C,O}} = 2.00\,\beta$. In the case of the dinucleotides, we used those values of $\beta_{i,j}$ integrals between atoms belonging to different bases, which have previously been determined on the basis of the appropriate overlap integrals (for the details see [7]).

Substituting the elements of the Hückel charge-bond order matrices $\mathscr{P}^{(0)}$ into the elements

---

* By this notation we refer to dinucleotide systems, which contain two superimposed bases in the same relative steric position as they occupy in a single helix of the Watson-Crick model of DNA.

of the appropriate Pople-type matrices, the matrices $\mathscr{F}$ with elements

$$F_{ii}^{(1)} = -I_i + \tfrac{1}{2}\, p_{ii}^{(0)} (I_i - E_i) + \sum_{j \neq i} (p_{jj}^{(0)} - Z_j)\, \gamma_{ij} \tag{1}$$

$$F_{ij}^{(1)} = \beta_{ij} - \tfrac{1}{2}\, p_{ij}^{(0)}\, \gamma_{ij}\,, \tag{2}$$

were constructed and their eigenvalue problems solved. (The notation here is the same as used by Pople [*14*]).

The Coulomb integrals γ_{ij} have been approximated by the expression [*11*]

$$\gamma_{ij} = \frac{e^2}{R_{i,j} + a_{i,j}}\,. \tag{3}$$

Here the $R_{i,j}$ internuclear distances have been computed in the case of the single bases from the geometry given by Spencer [*16*] and in the case of the dinucleotides from the data given by Langridge et al [*9*]. The constants $a_{i,j}$ were obtained by using the relation

$$e^2/a_{i,j} = \tfrac{1}{2}\,(I_i - E_i + I_j - E_j)\,, \tag{4}$$

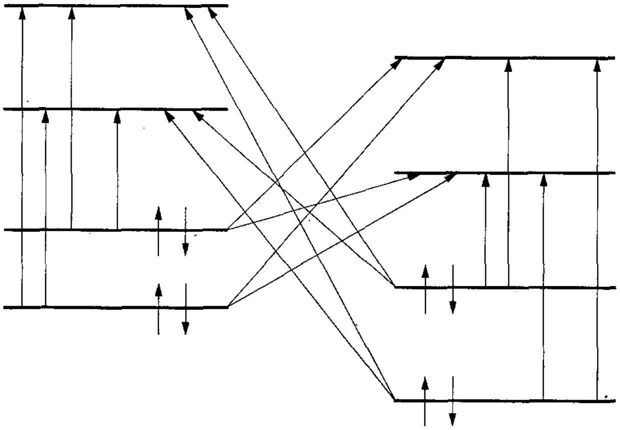

Fig. 1. The first 16 excited configurations of a composite system containing two parts

where I_i and E_i are the ionization potential and electron affinity respectively, of the i'th atom in its appropriate valence state (see Hince and Jaffé [*5*]). In the case of thymine we have used for I_i and E_i of the C-atom of the methyl group values $-4.473\,\beta$ and $-0.870\,\beta$ (values for an aliphatic C-atom), respectively, while for a C-atom in the ring they are $-4.478\,\beta$ and $-0.243\,\beta$, respectively. For $\beta_{i,j}$ the same values have been used in units of β as in the Hückel calculation ($\beta_{\mathrm{C,C}} = 1\,\beta = -2.39$ eV according to Pariser and Parr [*13*]).

Self consistency (constancy of the elements of the charge-bond order matrix up to three decimals) has been reached in five or six iteration steps.

With the aid of the eigenvectors of the matrix $\mathscr{F}$ a limited CI calculation has been performed for the first singlet excited states. For this we have used the eigenvectors of the matrix $\mathscr{F}$ obtained after the first iteration step in all cases. Further we have repeated the calculation for the nucleotide bases also with the SCF eigenvectors. For the single bases we have taken into account only the first four excited configurations $m \to m+1$, $m \to m+2$, $m-1 \to m+1$ and $m-1 \to m+2$ (m denotes the quantum number of the highest filled MO). In the case of G-C and of the GpG, CpC, GpC and CpG dinucleotides the 16 excited configurations, which arise from the abovementioned first four excited configurations of the constituent single bases, have been included in the CI calculation (see Fig. 1).

The elements of the interconfigurational matrix ${}^1\mathscr{G}$ for singlet-singlet transitions have the form [*8*]

$${}^1\mathscr{G}_{i\to k}^{j\to l} = \langle {}^1\Phi_{i\to k} \mid \mathscr{H} \mid {}^1\Phi_{j\to l} \rangle = \sum_{p=1}^{n} \sum_{r=1}^{n} (2\, c_{kp}\, c_{jp}\, c_{kr}\, c_{ir} - c_{ip}\, c_{jp}\, c_{kr}\, c_{lr})\, \gamma_{pr} + (\varepsilon_k - \varepsilon_i)\, \delta_{ij}\, \delta_{kl}\,. \tag{5}$$

Fig. 2. The SCF charge densities of the nucleotide bases: a) $\beta_{\mathrm{C},\underset{\diagup|\diagdown}{\mathrm{N}}} = 0.80\,\beta$, $\beta_{\mathrm{C,O}} = 1.30\,\beta$; b) $\beta_{\mathrm{C},\underset{\diagup|\diagdown}{\mathrm{N}}} = 0.90\,\beta$, $\beta_{\mathrm{C,O}} = 2.00\,\beta$

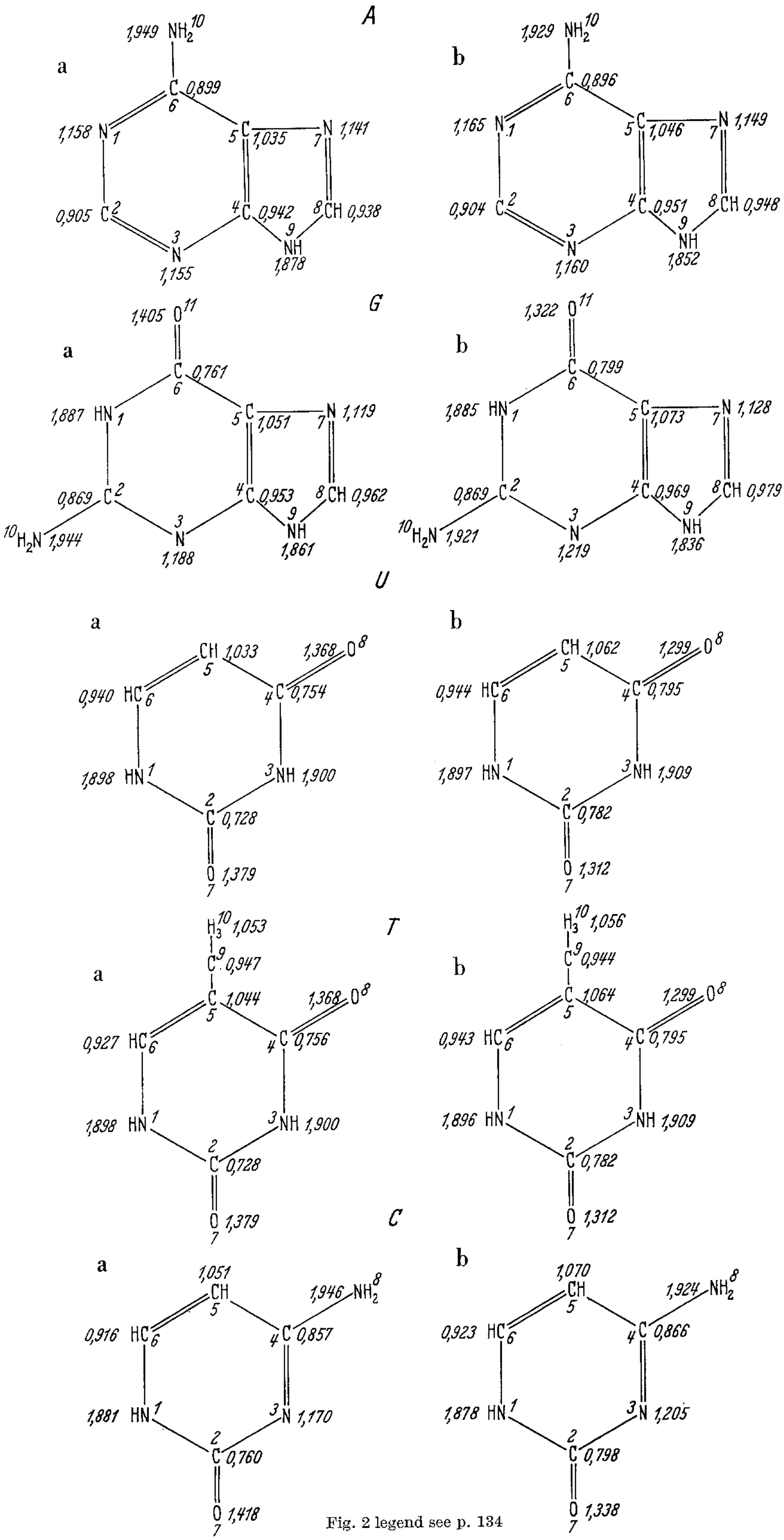

Fig. 2 legend see p. 134

Here n is the number of centers, c_{lp} denotes the p'th component of the l'th eigenvector of $\mathcal{F}$, ε_k and ε_i are the k'th and i'th eigenvalues of $\mathcal{F}$, and δ_{ij} and δ_{kl} are Kronecker δ's. In the case of a singlet-triplet transition we have an expression which differs from (5) only by the absence of the first (exchange) term in the parenthesis.

Using the eigenvectors of $^1\mathcal{G}$, b_v, we have also calculated the transition moment vectors of the different excited states by the expression

$$\boldsymbol{R}_v = \sum_{i\to k} b_{v,\, i\to k}\, \boldsymbol{R}_{i\to k}\,. \tag{6}$$

Here i and k run over the quantum numbers of the filled and unfilled states, respectively, in those excitations $i \to k$, which were taken into account in the CI calculation. The different transition moment vectors $\boldsymbol{R}_{i\to k}$ can be calculated most easily from the simplified expression [*3*]

$$\boldsymbol{R}_{i\to k} = \sqrt{2}\sum_{p=1}^{n} c_{kp}\, c_{ip}\, \boldsymbol{r}_p\,, \tag{7}$$

where $\boldsymbol{r}_p$ denotes the position vector of the p'th atomic nucleus.

Finally with the aid of the $\boldsymbol{R}_v$ vectors and the different E_v eigenvalues of $^1\mathcal{G}$ we have obtained the different oscillator strength values of the excited states:

$$f_v = 1.085 \times 10^{-5}\, \nu_v\, R_v^2;\ \nu_v = E_v/hc\,. \tag{8}$$

Results

In Fig. 2 we give the SCF charge densities for the five nucleotide bases a) with $\beta_{\mathrm{C},\,\mathrm{N}\leqslant} = 0.80\beta$ and $\beta_{\mathrm{C},\,\mathrm{O}} = 1.30\beta$ and b) with $\beta_{\mathrm{O},\,\mathrm{N}\leqslant} = 0.90\beta$, $\beta_{\mathrm{C},\,\mathrm{O}} = 2.00\beta$.

In Tab. 1 we give the first four singlet excitation energies of the nucleotide bases obtained by a limited CI together with experimentally found first two

Table 1. *The first singlet excitation energies of the nucleotide bases in eV*

	U		T		A		G		C	
	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF
	a) $\beta_{\mathrm{C},\,\mathrm{N}\leqslant} = 0.80\,\beta$, $\beta_{\mathrm{C},\,\mathrm{O}} = 1.30\,\beta$ ($\beta = -2.39$ eV)									
ΔE_1	4.80	5.65	4.85	5.64	4.20	4.57	2.94	3.96	2.80	4.40
ΔE_2	5.58	6.49	5.61	6.46	4.73	5.01	4.49	5.22	5.05	5.65
ΔE_3	5.83	6.94	5.85	6.96	5.35	5.83	5.33	5.92	5.76	6.79
ΔE_4	7.34	8.30	7.33	8.28	5.92	6.45	6.88	7.60	8.14	8.24
					(4.63 shoulder)					
$\Delta E_{1,\,\mathrm{exp}}$ [*16*]	4.81		4.67		4.75		4.49		4.61	
$\Delta E_{2,\,\mathrm{exp}}$ [*17*]	6.11		5.94		5.99		5.03		6.26	
	b) $\beta_{\mathrm{C},\,\mathrm{N}\leqslant} = 0.90\,\beta$, $\beta_{\mathrm{C},\,\mathrm{O}} = 2.00\,\beta$ ($\beta = -2.39$ eV)									
ΔE_1	5.78	6.19	5.25	6.19	4.21	4.75	2.99	4.11	2.99	4.68
ΔE_2	7.48	7.86	6.69	7.89	4.73	5.02	4.66	5.52	5.01	5.86
ΔE_3	7.84	8.32	7.31	8.32	5.62	6.17	5.19	6.33	6.26	7.10
ΔE_4	10.21	10.47	9.54	10.47	6.09	6.74	7.21	8.41	8.46	8.39
					(4.63 shoulder)					
$\Delta E_{1,\,\mathrm{exp}}$ [*16*]	4.81		4.67		4.75		4.49		4.61	
$\Delta E_{2,\,\mathrm{exp}}$ [*17*]	6.11		5.94		5.99		5.03		6.26	

Table 2*. *The oscillator strength values and transition moment vectors of the first singlet excitation energies of the nucleotide bases*

	U		T		A		G		C	
	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF
a) $\beta_{\mathrm{C,\,N}\le} = 0.80\,\beta$, $\beta_{\mathrm{C,\,O}} = 1.30\,\beta$ ($\beta = -2.39$ eV)										
f_1	0.761	0.803	0.767	0.804	0.039	0.011	0.381	0.560	0.201	0.448
f_2	0.013	0.068	0.013	0.066	0.433	0.498	0.088	0.451	0.393	0.011
f_3	0.115	0.155	0.121	0.148	0.772	0.777	0.854	0.266	0.400	1.172
f_4	0.560	0.513	0.565	0.513	0.527	0.699	0.027	0.003	0.028	0.050
R_1 (Å)	1.353	1.280	0.353	1.281	0.327	0.169	1.222	1.277	0.908	1.084
α_1°	—123.0	—125.0	172.2	54.8	67.1	151.6	117.8	120.7	— 27.2	— 23.1
R_2 (Å)	0.163	0.347	1.401	0.344	1.028	1.070	0.475	0.998	0.948	0.153
α_2°	—142.1	147.1	—125.2	— 31.4	—127.4	—119.4	— 88.8	—142.1	117.7	143.7
R_3 (Å)	0.476	0.508	0.939	0.495	1.290	1.240	1.359	0.721	0.895	1.412
α_3°	—164.3	45.2	117.0	—133.6	128.1	135.6	—150.9	—142.5	58.3	79.6
R_4 (Å)	0.938	0.844	0.091	0.845	1.013	1.117	0.213	0.064	0.200	0.265
α_4°	116.8	114.1	— 68.9	— 66.2	—131.2	53.7	172.9	56.9	26.3	124.2
$f_{1,\,\mathrm{exp}}$ [17]	0.18		0.18		0.28		~ 0.20		0.13	
$f_{2,\,\mathrm{exp}}$ [17]	0.30		0.37		0.54		0.27		0.72	
b) $\beta_{\mathrm{C,\,N}\le} = 0.90\,\beta$, $\beta_{\mathrm{C,\,O}} = 2.00\,\beta$ ($\beta = -2.39$ eV)										
f_1	0.786	0.765	0.768	0.765	0.139	0.014	0.435	0.623	0.216	0.417
f_2	0.078	0.108	0.008	0.079	0.383	0.544	0.098	0.133	0.112	0.007
f_3	0.075	0.079	0.141	0.108	0.778	0.855	0.069	0.106	0.984	1.149
f_4	0.649	0.765	0.695	0.646	0.529	0.709	0.244	0.227	0.149	0.065
R_1 (Å)	1.251	1.193	1.297	1.194	0.619	0.187	1.296	1.323	0.914	1.013
α_1°	59.5**	58.8	239.4	239.0	70.0	— 65.4	— 72.6	— 71.2	165.0	170.2
R_2 (Å)	0.347	0.340	0.114	0.340	0.967	1.118	0.494	0.527	0.507	0.116
α_2°	156.5	— 30.7	50.3	— 31.1	65.0	67.9	222.4	37.5	— 37.8	207.7
R_3 (Å)	0.332	0.386	0.471	0.387	1.264	1.265	0.392	0.440	1.347	1.366
α_3°	22.8	217.4	171.9	36.6	— 29.9	152.0	103.1	218.4	—110.2	265.3
R_4 (Å)	0.857	0.842	0.917	0.843	1.001	1.101	0.624	0.558	0.451	0.299
α_4°	— 74.1	— 73.8	110.0	— 73.9	71.8	68.8	7.9	7.2	— 25.4	— 38.4
$f_{1,\,\mathrm{exp}}$ [17]	0.18		0.18		0.28		~ 0.20		0.13	
$f_{2,\,\mathrm{exp}}$ [17]	0.30		0.37		0.54		0.27		0.72	

---

* The positive sign of the α angles means the anti-clockwise direction (see Fig. 3).

** See footnote on page 141.

transition energies $\Delta E_{1,\,\mathrm{exp}}$ and $\Delta E_{2,\,\mathrm{exp}}$. The significance of the underlined values will be explained in the discussion.

Tab. 2 contains the oscillator strength values and the transition moment vectors. The directions of the latter are given by the angle between the vectors and the line pointing from atom 1 to atom 2 (for the numbering of the molecules see Fig. 2).

In Tab. 3 we give the first triplet excitation energy of the nucleotide bases calculated without CI.

In Tab. 4 the 16 first singlet excitation energies of the G-C base pair are given which we have obtained using a limited CI and only parameters b).

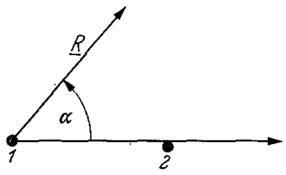

Fig. 3. The specification of the direction of the transition moment vectors

For the purpose of comparison in Tab. 5 we give the first four triplet excitation energies of the G-C base pair again calculated with the parameters b) and withoutCI.

In Tab. 6 we give the f-values of G-C and of the GpG, CpC, GpC and CpG dinucleotides obtained with CI using parameters b). For comparison we have included in the table also the first four f-values of the G and C single bases computed after the first iteration step with the same $\beta_{i,\,j}$ parameter values.

Finally in Tab. 7 we compare the sum of the f-values of the single bases and of the composite systems. In column a we give the sum of the f-values of the composite systems while column b gives the sum of the f-values of the constituent bases. The underlined values in the table indicate the sums of f-values belonging to the first two intensive absorption regions.

Table 3. *The first triplet excitation energy of the nucleotide bases (in eV)*

	U		T		A		G		C	
	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF	1. it.	SCF
a)	1.38	2.24	1.43	2.25	2.51	2.83	1.10	1.95	0.82	1.81
b)	1.88	2.35	1.51	2.35	2.60	2.96	1.11	2.06	0.64	1.85

Table 4. *The first 16 singlet excitation energies of the G-C base pair (in eV)*
$\beta_{C,N} = 0.90\,\beta$, $\beta_{C,O} = 2.00\,\beta$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. it.	2.97	3.12	3.21	4.70	4.98	5.23	5.53	5.78	5.85	6.17	6.47	7.44	7.74	8.11	8.59	9.53
SCF	4.10	4.52	4.72	5.51	6.27	6.30	6.45	6.99	7.22	7.25	7.56	8.27	8.36	9.24	9.28	9.70

Table 5. *The first four triplet excitation energies of the G-C base pair in eV*

	1	2	3	4
1. it.	1.07	1.20	3.12	3.48
SCF	1.99	2.11	4.54	4.55

Table 6. *Oscillator strength values of the* G-C *base pair and the* GpG, CpC, GpC and CpG *dinucleotides* (with excitation energies, in eV, in parentheses)

G	GpG	GpC	G-C	CpG	CpC	C
0.435 (2.99)	0.062 (2.66)	0.011 (2.73)	0.425 (2.97)	0.001 (2.74)	0.050 (2.78)	0.216 (2.99)
	0.116 (2.73)	0.108 (2.95)	0.094 (3.12)	0.121 (2.94)	0.193 (2.88)	
	0.617 (2.94)	0.506 (3.12)	0.148 (3.21)	0.506 (3.11)	0.135 (2.98)	
	0.008 (3.57)	0.001 (4.07)	0.007 (4.70)	0.007 (4.07)	0.001 (3.64)	
0.098 (4.66)	0.103 (4.24)	0.012 (4.73)	0.003 (4.98)	0.009 (4.72)	0.101 (4.92)	0.112 (5.01)
	0.132 (4.46)	0.006 (5.10)	0.178 (5.23)	0.002 (5.09)	0.156 (5.02)	
	0.011 (4.81)	0.160 (5.22)	0.000 (5.53)	0.178 (5.21)	0.004 (5.92)	0.984 (6.26)
0.069 (5.19)	0.074 (5.05)	0.294 (5.39)	0.537 (5.78)	0.604 (5.52)	0.112 (6.15)	
	0.120 (5.18)	0.001 (5.62)	0.141 (5.85)	0.002 (5.63)	0.158 (6.17)	
	0.002 (5.27)	0.364 (5.85)	0.040 (6.17)	0.051 (5.90)	1.569 (6.29)	
	0.000 (5.61)	0.235 (5.98)	0.001 (6.47)	0.171 (5.93)	0.023 (6.41)	
0.244 (7.21)	0.019 (6.29)	0.000 (6.46)	0.268 (7.44)	0.001 (6.46)	0.058 (7.06)	
	0.238 (6.78)	0.262 (7.46)	0.003 (7.74)	0.260 (7.46)	0.121 (8.51)	0.149 (8.46)
	0.277 (7.00)	0.000 (7.96)	0.051 (8.11)	0.004 (7.95)	0.237 (8.55)	
	0.004 (7.26)	0.352 (8.17)	0.001 (8.59)	0.075 (8.13)	0.000 (9.20)	
	0.002 (8.15)	0.023 (8.62)	0.001 (9.53)	0.001 (8.82)	0.005 (9.80)	
$\Sigma f = 0.846$	$\Sigma f = 1.785$	$\Sigma f = 2.335$	$\Sigma f = 1.898$	$\Sigma f = 1.987$	$\Sigma f = 2.923$	$\Sigma f = 1.461$
	1.692*	2.307*	2.307*	2.307*	2.922*	

---

* For comparison, the sums of the oscillator strength values for the constituent bases are also given.

Table 7. *Sums of f-values within spectral regions for base compounds (a) as compared with those for the constituent bases (b); negative values of (a — b)/b indicate hypochromicity, positive values hyperchromicity*

Compound	Region (eV)	$a = \Sigma f^{\mathrm{comp}}$	$b = \Sigma f^{\mathrm{bases}}$	$\frac{a-b}{b}$
GpG	2.6 — 3.6	0.803	0.870	— 0.077
	4.2 — 4.8	$\overline{0.246}$	$\overline{0.196}$	0.255
	5.0 — 5.6	$\overline{0.196}$	$\overline{0.138}$	0.420
	6.3 — 8.2	0.540	0.488	0.107
GpC	2.7 — 3.2	0.625	0.651	— 0.040
	4.0 — 5.3	$\overline{0.179}$	$\overline{0.279}$	— 0.359
	5.3 — 6.5	0.894	0.984	— 0.091
	7.2 — 7.8	$\overline{0.262}$	$\overline{0.244}$	0.074
	7.9 — 9.6	0.375	0.149	1.517
G-C	2.7 — 3.2	0.667	0.651	0.025
	4.0 — 5.3	$\overline{0.188}$	$\overline{0.279}$	— 0.326
	5.3 — 6.5	0.719	0.984	— 0.269
	7.2 — 7.8	$\overline{0.271}$	$\overline{0.244}$	0.111
	7.9 — 9.6	0.053	0.149	— 0.064
CpG	2.7 — 3.2	0.628	0.651	— 0.035
	4.0 — 5.3	$\overline{0.190}$	$\overline{0.279}$	— 0.319
	5.3 — 6.5	0.829	0.984	— 0.158
	7.2 — 7.8	$\overline{0.260}$	$\overline{0.244}$	0.066
	7.9 — 9.6	0.080	0.149	— 0.463
CpC	2.7 — 3.7	0.379	0.432	— 0.123
	4.9 — 5.1	$\overline{0.257}$	$\overline{0.224}$	0.147
	5.9 — 7.1	2.036	1.968	0.035
	8.5 — 9.8	$\overline{0.363}$	$\overline{0.298}$	0.218

Discussion

Comparing the obtained charge densities of the single bases in cases a) and b) (see Fig. 2) with those obtained by Veillard and Pullman [*18*] we can see that on the one hand there are only slight differences in the charge distributions between the cases a) and b). On the other hand the results of Veillard and Pullman show a little larger deviation, especially in the cases of nitrogen atoms in the sp^2 hybrid states, which contribute two electrons to the delocalized π electron system. The general picture is that the $\beta_{i,\,j}$ integral values, used by these authors, allow a larger delocalization for the $2\,\pi$ electrons of the sp^2-N atoms, than the $\beta_{i,\,j}$ values used by us.

Turning now to the problem of the excitation energies of the single bases we have to mention first of all that our attempt to obtain realistic singlet excitation energies with the parameter values $\beta_{\mathrm{C,\,N}} = 1.25\,\beta$ and $\beta_{\mathrm{C,\,O}} = 1.50\,\beta$ ($\beta =$ — 2.39 eV), which proved previously [*8*] successful in the case of U, failed completely in the case of C. Therefore we have not continued the calculation of the other nucleotide bases with these parameter values.

In Tab. 1 which gives the results obtained from the performed semiempirical limited CI calculation of the singlet excitation energies of the bases in cases a) and

b) the underlined values are those excitation energies which belong to the first two transitions having large oscillator strength values (underlined f-values in Tab. 2) as compared to the f values of the other excitations. Since the more intensive absorption bands usually cover the weaker ones, which occur in their neighbourhood, we have to compare only the two underlined excitation energies with the known two experimental values.

In connection with this selection of "measurable excitations" we have to make some comments. A detailed analysis of the numerical data obtained has shown that since the interconfiguration matrix elements between singlet excited states are rather small (the interconfigurational matrix elements between triplet excited states are somewhat larger), the obtained excitation energies differ only slightly from those, which we should obtain without CI. Therefore we can expect that a more detailed CI calculation, taking into account a large number of configurations, would influence the values of the first excitation energies and the corresponding f-values only slightly. This has the consequence that the choice of the first two intensive excitations would probably remain the same. On the other hand we can see from Tab. 2 that changes in the used $\beta_{i,j}$ integral values can influence the resulting f values more seriously. This can be demonstrated also if we compare our f values with those given in NESBET's [12] paper. The change of the f values with the $\beta_{i,j}$ values will only in the case of U influence our choice of the first two intensive excitations. Here as second intensive band should be chosen the third excitation with parameters a) and the second one with parameters b)*. It should be mentioned that with NESBET's [12] results the choice of the first two intensive excitations is completely the same as with our parameter values (NESBET does not treat U and T separately). It is also interesting to compare in Tab. 2 the f values obtained with the first iterated and SCF eigenvalues of the $\mathcal{F}$ matrix. We can see that in some cases (for instance in case a) f_2 and f_3 of G and C, or in case b) f_1 of A and f_2 of C, (see Tab. 2) there are drastic changes, which may change also the choice of the first two intensive excitations. For this reason the use of the SCF f values and excitation energies seems more founded. From all these we can conclude that although the selection of the first two intensive excited states of the nucleotide bases is of course somewhat arbitrary, yet it seems probable that the experimentally measured first two absorption maxima are due to the excitations, whose excitation energies we have underlined in Tab. 1.

Comparing now the underlined two excitation energies of the single bases in Tab. 1 with the first two experimental excitation energies we can find the following regularities.

1. For the two excitations of U and T and for the first excitation of A, the excitation energies calculated with the first iterated eigenvectors show a better agreement with the experimental ones in both cases a) and b), than the theoretical values calculated with the SCF eigenvectors. At the same time, however, for the first excitation of G and C in both cases a) and b) they give a very bad agreement (deviations of 1.5 — 1.6 eV). Therefore and for the reasons mentioned above it is better to restrict ourselves to the SCF results**.

2. Comparing the SCF-CI excitation energies in cases a) and b) we can see that using the $\beta_{i,j}$ values of case b) we obtain for the first excitation energies of G and

---

* It should be mentioned that in the cases of U and T with both sets of parameters a) and b) our fourth excitation is more intensive than the second or third one. Since, however, this fourth excitation lies about 1.5 — 2.0 eV above the third excited state, it is not probable that it will cover the third one. Furthermore using other $\beta_{i,j}$ values NESBET [12] obtained approximately the same f values for the third and fourth excited states of T. Therefore the above choice of the second intensive excited state is probably the correct one.

** A further disadvantage of the method of limited CI after the first iteration of the $\mathcal{F}$ matrix is, as one can see from Tab. 1 and 2, that the results are more strongly depending on the choice of the $\beta_{i,j}$ parameters, than the results of the SCF-CI calculations.

C somewhat better agreement than with the parameters of case a). On the other hand, however, the SCF-CI results with the parameters b) give a considerably worse agreement for both excitations of U and T and for the second absorption maximum of C (deviations up to 2.00 eV). The use of the parameter values of case a) pointed out by G. BERTHIER [*2*] gives a tolerable compromise, in which larger deviations between the experimental and theoretical excitations than about 1.00 eV do not occur (we find the maximal deviation by the second excitation of T to be equal of 1.02 eV). It should be mentioned that NESBET [*12*] in this case only has a deviation of 0.50 eV, but his results for G and C show a worse agreement with the experimental values, than ours.

To reproduce better the spectra of the nucleotide bases it would be necessary to perform similar calculations on larger series of substituted purine and pyrimidine type compounds varying the used $\beta_{i,j}$ integrals. From these rather time consuming calculations it would be possible to determine the best possible set of $\beta_{i,j}$ integrals for the reproduction of spectra. It should be pointed out, however, that according to our opinion, since the nucleotide bases are chemically rather different compounds, this "best" series very probably will be just another compromise and will not describe exactly the experimental spectra.

Turning now to the problem of oscillator strength values we find the usual situation in that the agreement between the theoretical and experimental values in both cases a) and b) is bad (see Tab. 2). In most cases also the ratio between the two underlined theoretical f-values differs from the experimental ratio. We can further see from the data given in Tab. 2 that the directions of the different transition moment vectors (the α_i angles) are extremely sensitive to the choice of the $\beta_{i,j}$ parameter values. We find also larger changes in the α_i angles, than in the absolute values of the $\boldsymbol{R}_i$ vectors, if, instead of SCF eigenvectors, we use the eigenvectors of the $\mathscr{F}$ matrix obtained after the first iteration step. All this indicates that the simple Pariser-Parr-Pople method is not suitable for a good approximation of the intensity of such complicated systems as the nucleotide bases.

From Tab. 3, which gives the first triplet excitation energies of the nucleotide bases calculated without CI and with the sets a) and b) of $\beta_{i,j}$ integrals, we can see that the SCF values are about 2.00 eV for G and C and are higher for A, T, and U. We find that the values obtained with the eigenvectors of the $\mathscr{F}$ matrix after the first iteration step are in all cases considerably smaller than the SCF values. Comparing the SCF triplet excitation energies of cases a) and b) we can see that the change of the used $\beta_{i,j}$ parameters has smaller effect on the first triplet excitation energies of the nucleotide bases than on the singlet ones (see Tab. 1).

The values of the first 16 singlet excitations of a G-C pair given in Tab. 4 were obtained after a CI of the 16 configurations, which arise from the previously considered four configurations of the appropriate single bases (see Fig. 1), using the $\beta_{i,j}$ integrals of case b). Comparing these SCF results with those referring to the G and C single bases (see the second part of Tab. 1), we can see that some excitations preserve their identities (for example the first excitation of the G-C base pair can be classified as the first excitation of G, or the third excitation of G-C comes from the third excitation of C).

For most excitations, however, we do not find such a clear situation after the CI, and so we cannot classify the different excitations as pure G type, or pure C type, or as "mixed" excitations (see Fig. 1).

It is interesting to point out that the first two SCF triplet excitation energies of a G-C base pair are again around 2 eV (1.99 eV and 2.11 eV, respectively; see Tab. 5). This is not in too bad agreement with the experimental result of Douzou et al., who have found for the first triplet excitation of DNA the value of 2.55 eV [*4*].

Since for the semiempirical CI calculation of the GpG, GpC, CpG, and CpC dinucleotides we have used only those eigenvectors of the $\mathscr{F}$ matrix, which we obtained after the first iteration step, and since we have applied only the $\beta_{i,j}$ parameters of set b), we do not try to compare the resulting excitation energies with the experimentally determined first two singlet excitation energies of the single bases. Therefore in Tab. 6 we compare only the obtained oscillator strength values of the mentioned dinucleotides with the f-values of the appropriate single bases calculated in the same approximation with the same parameter values. With this comparison we try to make a further step [*6*] for the interpretation of hypochromicity of DNA assuming only overlap type interaction between the superimposed bases. In the case of GpG and CpC on the basis of the singlet excitation energies (given in Tab. 6 in parentheses) we can divide the different excitations into four parts. The excitations within each part of the spectrum lie approximately in the same spectral region and their excitation energies do not differ much from one of the excitation energies of the single G or of the single C base. In the case of the GpC and CpG "mixed" dinucleotides and of the G-C base pair the sitution is somewhat more complicated. If we investigate simultaneously the excitation energies of G and C we can see that now we have five spectral regions: the first around 3 eV (the first excitation of G and C), the second at about 4.5 — 5 eV (the second and third excitations of G and the second one of C), the third around 6 eV (the third excitation of C), the fourth around 7 eV (the fourth excitation of G) and finally the fifth above 8 eV (the fourth excitation of C). We have indicated these spectral regions by horizontal lines in Tab. 6 and we give them explicitly in the first column of Tab. 7. In the bottom of Tab. 6 we give the sum of the oscillator strength values of the different systems and for the composite systems also the sum of Σf of their two constituents (in parantheses). We can see that with the exception of CpC the two sums do not agree.

In Tab. 7 we compare the sums of the f-values of the constituent bases and those of the composite systems. The summation is always extended to all the f-values, which belong to excitations lying in the indicated region. In the last column is given the change of total intensity of the region under consideration in percent of the original intensities (negative values: hypochromicity, positive values: hyperchromicity). The underlined two values indicate the values which belong to the first two intensive absorption regions, which we have selected in a similar manner as the first two intensive singlet excitations of the single bases (see above). From the results obtained we can see that we have for all the four investigated dinucleotides a hypochromicity in the first intensive absorption region, which is small for the "mixed" dinucleotides (— 4 % for GpC and — 3.5% for CpG, respectively), it is greater for GpG (— 7.7%) and has its largest value for CpC (— 12.3%). At the same time for the G-C base pair we find a slight hyperchromicity (+ 2.5%). In the case of the second intensive absorption region we find a rather large hypochromicity for GpC, CpG and G-C (— 9.1%, — 15.8%, and

— 26.9%, respectively), while for the "pure" dinucleotides we have a hyperchromicity (+ 25.5% for GpG and + 3.5% for CpC).

These promising results show that there is a possibility to describe the observed hypochromicity of DNA with the aid of a model which assumes only overlap interaction between the superimposed bases. Of course it would be highly desirable to continue the calculation with the SCF eigenvectors of the $\mathscr{F}$ matrix and to investigate the effect of the uncertain $\beta_{i,j}$ integrals on the hyperchromicity results. Further, in order to establish a more precise theory it would be necessary to investigate the effect of the number of configurations included in the CI calculation on the results, and to extend the calculations as first step to the other 12 dinucleotides and later to polynucleotides. Finally, a still more realistic calculation would take into account simultaneously overlap type interactions and exciton-van der Waals type interactions between the superimposed bases of DNA.

Acknowledgements. We should like to express our gratitude to Professor P.-O. LÖWDIN for many interesting discussions and for the opportunity of working in the stimulating atmosphere of the Uppsala Quantum Chemistry Group. We are greatly indebted to Professors KIMIO OHNO and J. KOUTECKÝ for their useful remarks, to Dr. G. BERTHIER for pointing out suitable $\beta_{i,j}$ integral values and to Fil. kand. F. BYSTRAND for his help in some parts of the programming work.

References

[1] BEAVEN, G. H., E. R. HOLIDAY, and E. A. JOHNSON: In The Nucleic Acids, Ed. CHARGAFF, E., and I. N. DAVIDSON, Vol. I, p. 493. New York: Academic Press 1955.

[2] BERTHIER, G.: (private communication).

[3] DAUDEL, R., R. LEFEBVRE, and C. MOSER: Quantum Chemistry. New York: Interscience Publ. Inc. 1959.

[4] DOUZOU, P., J. C. FRANCQU, M. HAUSS, and M. PTAK: J. Chim. physique 58, 926 (1961).

[5] HINZE, T., and J. JAFFÉ: J. Am. chem. Soc. 84, 540 (1962).

[6] HOFFMANN, T. A., and J. LADIK: J. theor. Biol. 6, 26 (1964).

[7] — — Advances in Chemical Physics 7, 84 (1964).

[8] LADIK, J., and K. APPEL: Preprint No. 79, Quantum Chemistry Group, University of Uppsala, 1962.

[9] LANGRIDGE, R., H. R. WILSON, C. W. HOOPER, M. H. F. WILKINS, and L. D. HAMILTON: J. mol. Biol. 2, 44 (1960).

[10] LÀNG, L., and J. LADIK: Periodica Polytechnica Techn. Univ. Budapest 8, 267 (1964).

[11] MATAGA, N., u. K. NISHIMOTO: Z. physik. Chemie 13, 140 (1957).

[12] NESBET, R. K.: Biopolymers Symposia 1, 129 (1964).

[13] PARISER, R., and R. G. PARR: J. chem. Physics 21, 466, 767 (1953).

[14] POPLE, J. A.: Trans. Faraday Soc. 49, 1375 (1953).

[15] REIN, R., and J. LADIK: J. chem. Physics 40, 2466 (1964).

[16] SPENCER, M.: Acta Cryst. 12, 59 (1959).

[17] TINOCO, I. Jr.,: J. Am. chem. Soc. 82, 4785 (1960); 83, 5047 (1961). — J. chem. Physics 34, 1067 (1961). — RHODES, W.: J. Am. chem. Soc. 83, 3609 (1961). — NESBET, R. K.: Mol. Physics 5, 63 (1962). — DEVOE, H., and I. TINOCO, Jr.: J. mol. Biol. 4, 518 (1962).

[18] VEILLARD, A., and B. PULLMAN: J. theor. Biol. 4, 37 (1963).

(Received September 14, 1965)