Z. phys. Chemie, Leipzig **250** (1972) 1/2, S. 17-22

Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de La Plata (Argentina)

Influence of the Hydrogen Bonds and the Methyl Group on the π Electrons for the Guanine-Cytosine Pair

By **O. M. Sorarrain, J. A. Fornés, L. M. Boggia** and **R. Veronesi**[1]

With 1 Figure and 5 Tables (Received 26th April 1971)

Abstract

The influence of the hydrogen bonds in the $G - C$ pair is analysed. We have used the same procedures already reported for the $A - T$ pair.

The interaction of the methyl group with π electrons in the cytosine base is also analysed by means of the theory of separated electronic groups. A simplified model for that interaction is discussed. Possible imperfections and improvements in the model are considered.

1. Introduction

An analysis similar to that previously described for the $A - T$ complementary bases has been performed for the $A - G$ pair of bases. As the same procedures reported in the previous paper have been used, further consideration of them has been omitted here.

2. Model for the Methyl Group Interaction

It is, however, necessary to add the effect produced by the methyl group fixed in carbon 5 of cytosine (Fig. 1). The interaction of the methyl group with the rest of the

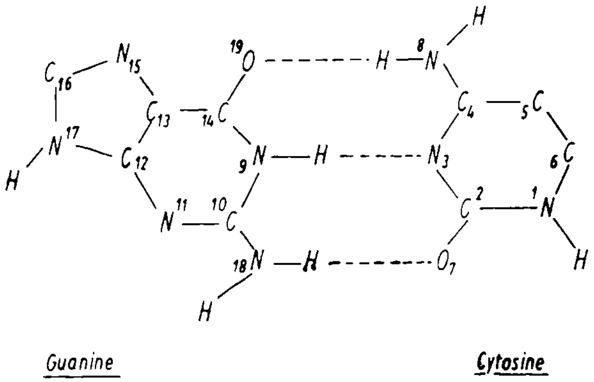

Fig. 1.

---

[1] Dr. OSCAR MANUEL SORARRAIN, JOSÉ ANTONIO FORNÉS, LUIS MARIO BOGGIA and RUBEN VERONESI, Departamento de Fisica, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata (Argentina), Calle 115 Y 49 C.-C. Vr. 67.

molecule can be calculated if the electronic interaction is assumed to have the form:

$$\begin{aligned} G_1 = {} & 3A_1^2G_\alpha + A_2^2G_{\alpha'} + A_1^2G_{\alpha''} + 2A_2^2G_\alpha + 2A_2^2G_{\alpha''} - (3A_1^2 + 2A_2^2)\,G_\alpha \\ & + (A_1^2 + A_2^2)\,G_{\alpha''} + A_2^2G_{\alpha'} \end{aligned} \tag{1}$$

The α orbitals characterise the bonds of the three hydrogens with the C in the methyl, which are identical for reasons of symmetry. $G_{\alpha''}$, denotes the σ bond between carbon 5 and the carbon in the methyl; $G_{\alpha'}$, defines the term of interaction between cytosine and the semi-localised π electrons which appear in the hyperconjugation processes.

The wave functions corresponding to the methyl group are assumed to have two different configurations: The first, whose coefficient will be called A_1, corresponds to the fundamental configuration formed with orbitals α and α'', and the second one, formed with two bonds formed by orbitals α and the pair of semi-localised orbitals α'. The weight coefficient A_2 of the last configuration is in general remarkably lower than A_1.

By varying A_1 and A_2, as will be calculated, we consider that we can approximate the expected hyperconjugation effects, although the model is not a rigorous one.

Formula (1) does not take into account the spin states, but it will be sufficient to remember that Coulomb operators produce the same results with both spin orientations; thus it is only necessary to multiply by 2 in the corresponding interaction formulas, which, insofar as the exchange operators are concerned, will not occur. The coefficients A_1 and A_2 are unknown, but we can vary them within reasonable limits.

Formula similar to those already given (1) for the hydrogen bonds can be used if it is taken into account that orbitals ϕ_α, $\phi_{\alpha'}$, $\phi_{\alpha''}$, have the following form:

$$\begin{aligned} \phi_\alpha &= (d_1\psi'_\sigma + d_2\psi'_{\pi'} + d_3\psi'_\pi + d_4\psi'_{\mathrm{H}} + d_5\psi'_5) \\ \phi_{\alpha'} &= (f_1\psi_{\mathrm{H}} + f_2\psi'^2_{\mathrm{H}} + f_3\psi'^3_{\mathrm{H}} + f_4\psi'_\sigma + f_5\psi'_\pi + f_6\psi'_{\pi'}) \\ \phi_{\alpha''} &= (c_1\psi'_\sigma + c_2\psi'_{\pi'} + c_3\psi'_\pi + c_4\psi'_{\mathrm{H}} + c_5\psi'_5) \end{aligned} \tag{2}$$

The atomic orbitals ψ'_σ, ψ'_{S}, ψ'_π, $\psi'_{\pi'}$ are located in the C of the methyl group; the orbitals ψ'_{H} belong to the hydrogens. The notation for the orbitals localised in the C of the methyl is the same as used by Kotani, already quoted [2].

The Coulomb interactions have the form:

$$\begin{aligned} & c_1^2(3A_1^2 + 2A_2^2)\,(d_1D_{\pi\sigma\pi\sigma} + d_1d_4S_{\sigma\mathrm{h}}D_{\pi\sigma\pi\sigma} + d_2^2D_{\pi\pi'\pi\pi'} + d_3^2D_{\pi\pi\pi\pi} + d_4^2M_{\pi\pi} \\ & + d_4d_5S_{\sigma\mathrm{h}}D_{\pi\mathrm{s}\pi\mathrm{s}} + d_5^2D_{\pi\mathrm{s}\pi\mathrm{s}}) + c_1^2(A_1^2 + A_2^2)\,(c_1D_{\pi\sigma\pi\sigma} + c_1c_4S_{\sigma\mathrm{h}}D_{\pi\sigma\pi\sigma}S_{\sigma\mathrm{h}} \\ & + c_2^2D_{\pi\pi'\pi\pi'} + c_3^2D_{\pi\pi\pi\pi} + c_4^2M_{\pi\pi} + c_4c_5S_{\mathrm{sh}}D_{\pi\mathrm{s}\pi\mathrm{s}} + c_5^2D_{\pi\mathrm{s}\pi\mathrm{s}}) + (f_1f_2f_3)\,M_{\pi\pi}A_2^2 \end{aligned} \tag{3}$$

The integrals D have been taken for $R = 1{\cdot}5$ Å, while $S_{\sigma\mathrm{h}}$, S_{sh}, $N_{\pi\pi}$ have been taken for $R = C - H$ in methyl; and $M_{\pi\pi}$ is taken for an R equal to the distance between the carbon 5 (Fig. 1) and the hydrogens in the methyl group.

The exchange integrals have the form:

$$c_1^2(3A_1^2 + 2A_2^2)(d_1C_{\pi\sigma\sigma\pi} + 2d_1d_5C_{\pi\sigma s\sigma} + d_2^2C_{\pi\pi'\pi\pi'} + d_3^2C_{\pi\pi\pi\pi} + d_4^2N_{\pi\pi} + d_5^2C_{\pi ss\pi})$$

$$+ c_1^2(A_1^2 + A_2^2)(c_1^2C_{\pi\sigma\sigma\pi} + 2c_1c_5C_{\pi\sigma s\pi} + c_2^2C_{\pi\pi'\pi\pi'} + c_3^2C_{\pi\pi\pi\pi} + c_4^2N_{\pi\pi} + c_5^2C_{\pi ss\pi})$$

$$+ (f_1 + f_2 + f_3)N_{\pi\pi}A_2^2 \tag{4}$$

where the integrals C have been taken for $R = 1{\cdot}5$ Å, while $S_{\sigma h}$, S_{sh}, $N_{\pi\pi}$, correspond to the distances previously mentioned. To correct the wave functions, we can apply the theory of perturbations, similar considerations to those already quoted being valid [2].

The choice of the coefficients corresponding to the atomic orbitals supplied by the C and the hydrogens in the methyl group constitutes an additional problem, similar to that already dealt with in respect of the hydrogen bonds [2]. Although the above mentioned values are not known, they can be estimated on the basis of the following criteria:

First, the C in methyl must supply hybrids which constitute tetrahedral bonds. If direction $C - C$ is taken as σ, the result is that the bond in question will give us terms of interaction between orbitals π supplied by carbon 5, (C_5) and the orbitals given by the methyl C, (C_m), which can be estimated as being $a\psi_s' + b\psi_\sigma'$, the ψ' orbitals being the $(2s)\,(2p)$ for the carbon.

The coefficients a and b being unknown, we will choose as approximation a certain value which we will call "typical". For the bond between C_5 and C_m, we assume an orbital:

$$\phi = N(\psi_s'^{C_5} + \psi_\sigma'^{C_5} + \psi_s'^{C_m} + \psi_\sigma'^{C_m}) \quad (\sigma \text{ is taken along the } C_5 - C_m \text{ direction})$$

Then, for the "typical orbital", we have:

$$\phi_{\lambda''} = N(\psi_s'^{C_m} + \psi_\sigma'^{C_m})$$

The interaction terms between the three remaining σ bonds formed by the carbon and the hydrogens belonging to the methyl group constitute a similar problem. As regards these bonds, we know that they must be described by mutually orthogonal orbitals. They must also be orthogonal with the $C - C$ bond already mentioned. Nevertheless, we will estimate a first approximation for these orbitals on the condition that the hybrids supplied by the C of the methyl have their highest bond values in the $C - H$ directions; taking into account, moreover, that for reasons of symmetry it will be sufficient to analyse one out of the three mentioned, where the corresponding molecular orbitals have the form:

$$\phi_x = a'\psi_s + b'\psi_\sigma + c'\psi_\delta + d'\psi_\gamma + e'\psi_H$$

In this way ϕ_α is assumed to be an "average orbital" for the rotation of the hydrogens along the $C_5 - C_m$ direction. This simplified model permits us to choose directions δ and γ normal to $C - C$ direction, in such a way that the H lies on the (σ, δ) plane.

As for the orbitals in the hyperconjugation process, according to COULSON [1], they have been approximated by

$$\psi'_{H_1} - \tfrac{1}{2}(\psi'_{H_2} + \psi'_{H_3})$$

multiplied by a normalization factor. Lastly, let us say that even though the estimations made are not very rigorous, we consider them very useful, at least to enable us to know the amount of the perturbation in the cytosine base, when the interaction with the methyl group is considered. This is, actually, the desired objective.

3. Numerical Results and Discussion

An analysis similar to that previously reported for the *A* − *T* has been performed for the guanine-cytosine pair. The same conclusions arise from the results obtained:

Table 1. Electronic distribution

	Guanine-Cytosine Separted Extended Hückel Method
3	1·521
7	1·522
8	1·797
9	1·720
18	1·836
19	1·495

Table 2. Electronic distribution

		Guanine-Cytosine		
λ	λ'	0·8	0·9	0·95
3	0·8	1·604	1·672	1·613
	0·9	1·606	1·614	1·624
7	0·8	1·540	1·541	1·542
	0·9	1·540	1·544	1·547
8	0·8	1·761	1·760	1·760
	0·9	1·764	1·762	1·759
9	0·8	1·755	1·752	1·749
	0·9	1·753	1·749	1·745
18	0·8	1·798	1·792	1·790
	0·9	1·805	1·801	1·798
19	0·8	1·503	1·505	1·506
	0·9	1·502	1·508	1·510
	Extended Hückel Method			

Table 3

Guanine	G–C	Cytosine	G–C
−15·9274	−15·6437	−14·8548	−14.5932
−12·0143	−11·8991	−13·0343	−13·2655
− 7·1233	− 6·9274	− 8·2975	− 8·4174
− 5·7748	− 5·7335	− 4·1454	− 4·2123

Two highest filled and two lowest empty orbitals in both bases. (ev).

For the interaction with the methyl group, the orbitals of the methyl C have been selected in the manner already reported. These orbitals play a role in the $C - C$ and $C - H$ bonds. Thus, for instance, the $C - H$ bonds have the form:

$$\phi_\alpha = N(\phi_c + e\psi'_H)$$

where ϕ_c involves orbitals s and p of the methyl C and ψ'_H is the 1 s orbital in the hydrogen; N is a convenient normalization factor.

The e parameter has been varied within a certain range, together with the relation A_1/A_2, where $(A_2 = \sqrt{1 - A_1^2})$, between the selected configurations. The effective changes of the methyl C have been estimated by means of the expression $Z - s$, s being the corresponding value, according to the Slater rules, to the two 1 s electrons of the methyl C which have not been considered.

All possible combinations for parameters e and A_1, i.e., $e = 1{\cdot}0$, $0{\cdot}9$, $0{\cdot}85$, $0{\cdot}8$; $A_1 = 0{\cdot}8$, $0{\cdot}85$, $0{\cdot}9$, $0{\cdot}95$, have been taken. For all these combinations, the distribution of charges in the elements involved in the hydrogen bonds are affected in the same fashion. As an example, we give one possible combinations of e and A_1:

Table 4. Electronic distribution

	Cytosine Extended Hückel Method
3	1·525
7	1·518
8	1·799
	$e = 0{\cdot}9 \quad A_1 = 0{\cdot}85$

Table 5

Cytosine	5-Methyl Cytosine
−14·8548	−14·6725
−13·0343	−12·7175
− 8·2975	− 8·2615
− 4·1454	− 4.2348

Two highest filled and two lowest empty orbitals (ev).
$e - 0{\cdot}9 \quad A_1 = 0{\cdot}85$

4. Last Comments and Final Conclusions

Considerations similar to those previously presented in the former paper are valid in relation to the introduction of the methyl group. But some further difficulties arise here due to there not yet being a satisfactory model describing the hyperconjugation effects. Nonetheless, the orbitals proposed by Coulson could be used in a different manner. In any event, we feel that other more orthodox methods do not change the essential conclusions arrived at (i.e., the theory of separated electronic groups and the theory of perturbations are very useful in describing the interaction of the methyl group on the electrons in cytosine), since the perturbation matrix would not perceptibly change with any other model. Obviously, if a totally satisfactory model were found, the interaction of the methyl group with the set of π electrons could be "ab-initio" dealt with by means of the operators defined in the formula (10) of our former paper.

Regarding the interaction between guanine and cytosine, all considerations given in our former paper are clearly valid, and are therefore not repeated.

References

[1] Coulson, C. A.: Valence. Oxford: Univ. Press 1961.

[2] Fornés, J. A., O. M. Sorarrain, L. M. Boggia and R. Veronesi: Z. phys. Chemie **250** (1972) 23.