Theoret. Chim. Acta (Berl.) 45, 249–256 (1977)

THEORETICA CHIMICA ACTA

On the Affinity of Cytosine towards Electrophiles*

Alberte Pullman and Anne Marie Armbruster

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S., 13, rue P. et M. Curie, F-75005 Paris, France

Ab initio SCF computations on the intrinsic preferences of the H^+, CH_3^+ and $C_2H_5^+$ cations towards the two principal sites of protonation or alkylation on cytosine, N_3 or O_2, show that this preference undergoes a continuous modification with the increase in size and complexity of the cation. N_3 is the preferred site of fixation of H^+, O_2 the preferred site of $C_2H_5^+$, while CH_3^+ has no marked preference. The exchange repulsion term of the binding energy appears responsible for the preference of $C_2H_5^+$ for O_2.

Key words: Cytosine, alkylation of ~

1. Introduction

Protonation and electrophilic attacks of the nucleic acid bases represent important biological reactions. The reactions with alkylating agents have drawn particular attention because of their involvement in carcinogenesis and mutagenesis.

Among important recent developments in this field is the demonstration that different classes of electrophilic agents differ in their affinity for target sites on the nucleic acid bases. Thus from the large amount of experimental data available on the protonation and alkylation of the purines and pyrimidines of the nucleic acids, their nucleosides and nucleotides, either free or in the nucleic acids, it is known that while protonation occurs practically always at the ring nitrogens, alkylation may occur at various positions, according to the reagent and to the conditions. Thus e.g., for guanine, while the N_7 atom is the preferred site of *protonation* and of *methylation* with a number of reagents under neutral conditions, O_6 becomes the preferred site of

* This work was supported by the Ligue Francaise contre le Cancer and the National Foundation for Cancer Research (USA)

ethylation [3] with certain ethylating agents, such as ethylnitrosourea (although not with others, e.g. ethyl-methanesulfonate or diethylsulfate). Similarly, while cytidine protonates and reacts preferentially with a variety of methylating agents at N_3, a recent report has indicated that O_2-ethylcytidine is the major product of neutral, aqueous reaction of cytidine with ethylnitrosourea [4].

From the theoretical point of view an important step towards the elucidation of the factors governing protonation and alkylation reactions was achieved through the use of the electrostatic molecular potentials [5]. The procedure consists of characterizing a molecule by the global potential created in the surrounding space by the nuclear charges and electron distribution. The potential represents then what a positive reagent "feels" when approaching the substrate molecule. Precisely, if a point charge q is placed in the potential V, its electrostatic energy of interaction with the unperturbed molecule is qV. Clearly, the potential represents only a part, the electrostatic (or Coulomb) component, of the total interaction energy. It is the sole part appreciable when the approaching charge is at large distance of the molecule. Upon closer approach, polarization of the molecule intervenes, giving rise to a polarization component in the interaction energy. Finally when the approach is still closer, a charge-transfer component appears, due to the fact that the electrons of the attacked molecule are displaced towards the approaching charge. When the approaching entity is a reagent other than a bare proton, a similar analysis of the interaction energy holds except that, at short distances, a repulsive component appears, due to the overlap of the electron clouds of the two entities in interaction [6, 7].

It was explicitly shown that the electrostatic molecular potentials give a very satisfactory account of the proton affinities of the nucleic acid bases [8, 9].

In view of the differences observed in the site of fixation of the alkylating agents, the question may be raised concerning the possible variation of their affinity towards different types of site as a function of the exact nature of these agents. Although at large distances from the target their approach will be governed essentially by the electrostatic potential, it is not excluded that, depending on the nature of the alkylating agent, the polarization and charge transfer or the exchange repulsion components of the interaction may impose a different path of approach, and finally, at equilibrium distance, a different site of fixation.

In this paper we propose to explore, as a preliminary to more complete and complex studies, the *intrinsic* preferences of the CH_3^+ and $C_2H_5^+$ cations towards the nitrogen and oxygen sites on a nucleic acid base and to compare them to the intrinsic preference manifested by H^+. The base chosen is cytosine, which possesses both N and O sites and is smaller and thus easier to deal with theoretically than guanine. We underline the fact that this introductory study involves only the base and the cation and does not take into account the nature of the alkylating compound, nor its aptitude to generate a cationic species. It is devoted to the determination of the *intrinsic tendencies of the alkyl cations*, as a function of their size and structure. The study involves also, naturally, the protonation reaction.

2. Method

We have used the *ab initio* LCAO MO SCF procedure [10]. For this unravelling study we have adopted an STO 3G basis set [11]. We used the supermolecule approach, where the system cytosine-cation is treated as a single entity considering all the electrons in the field of all the nuclei placed in the appropriate relative positions, and the results compared to those of the isolated subunits computed separately in the same fashion. The interaction energy for each relative position of the two entities in interaction is thus:

$$\Delta E = E(\text{cytosine-cation}) - E(\text{cytosine}) - E(\text{cation}) \tag{1}$$

The geometry used for the cytosine molecule (Fig. 1a) is taken from its crystal structure [12]. It was kept unchanged in all the computations. For the methyl and

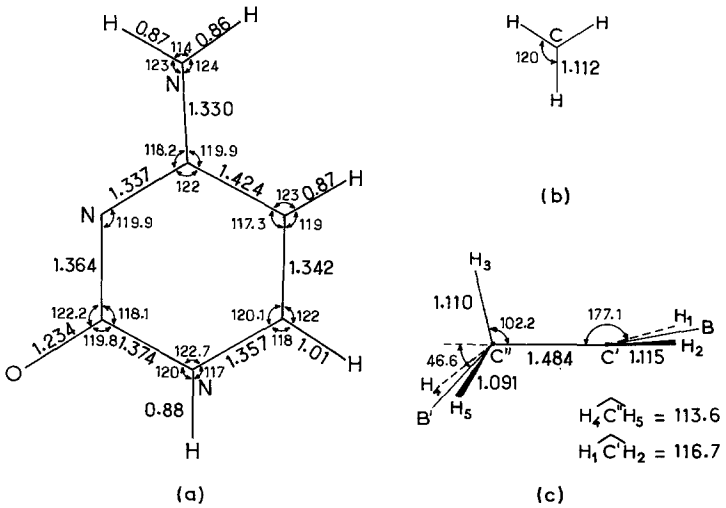

Fig. 1. Geometries used as starting data for computations. (a) Cytosine. (b) Methyl cation. (c) Ethyl cation. *B* and *B'* are the bissectrices of the $\widehat{H_1C'H_2}$ and $\widehat{H_4C''H_5}$ angles respectively

ethyl cations the computations were started with the most stable geometries obtained by W. A. Lathan *et al.* [13] after full optimization in the STO 3G basis (Fig. 1b and c). Partial relaxation of the geometrical constraints on the cations was later allowed to permit their deformation in the course of binding to the molecule.

The computations of the molecular potential of cytosine was done using our modified version of the program Magasen [14]. In different positions, a computation of the various components of the binding energy upon protonation and alkylation was performed as explained in Refs. [6] and [7], respectively.

The total energy of interaction at a given distance of approach is the sum

$$\Delta E = \Delta E_C + \Delta E_{DEL} + \Delta E_{EX} \tag{2}$$

where ΔE_C is the pure electrostatic (or Coulomb) energy of interaction of the two entities supposed unperturbed, ΔE_{DEL} is the sum of the polarization and charge transfer energies, and ΔE_{EX} is the exchange repulsion term, absent in the case of a bare proton.

3. Results and Discussion

3.1. Protonation

As a starting element for the forthcoming discussion we give here the molecular electrostatic potential of cytosine computed with the STO 3G basis set (Fig. 2). The map shows the existence of two deep potential wells facing the N_3 and O_2 atoms, respectively. The deepest well (-104.5 kcal/mole) is associated with N_3 which represents therefore the primary center of electrostatic attraction towards a point positive charge and thus towards the protons. It may be observed that this well is about 20 kcal/mole deeper than that associated with O_2.

Table 1 gives the result of the evaluation of the total energy of protonation of cytosine at two sites: N_3 and O_2, computed in the supermolecule approach, optimizing the position of the proton in each case.

Table 1. Protonation energies of N_3 and O_2 of cytosine and their components[a]

	ΔE	E_C	E_{POL}	E_{CT}
N_3	-293.3	-105.4	-45.4	-142.5
O_2	-291.5	-85.6	-43.4	-162.5

[a] ΔE and components as defined in the text (kcal/mole).

It may be observed that although the N adduct is the most stable one, the difference in energies between this product and the one protonated at O_2 is only of about 2 kcal/mole. The table indicates also the decomposition of the total interaction energy into its three components: the electrostatic, polarization and charge transfer ones carried out as in Ref. [6]. It may be observed that while the electrostatic energies at the equilibrium positions remain very close to those corresponding to the two potential minima of Fig. 2 and thus maintain an energy difference of about 20 kcal/mole in favor of the N-protonated form, this situation is substantially modified by the effect of the remaining components. Thus, within the STO 3G computations, the polarization energy is also 2 kcal/mole in favor of the N-protonated form but the charge transfer component is on the contrary appreciably (20 kcal/mole) in favor of the O-protonated species. As a result, the stabilities of the two forms become comparable with, however, a slight advantage on behalf of N-protonation.

3.2. Methylation

The results for methylation point to the still closer equivalence of the interaction energies for N_3 or O_2-alkylation. A planar CH_3^+ ion ($HCH = \alpha = 120°$) was first

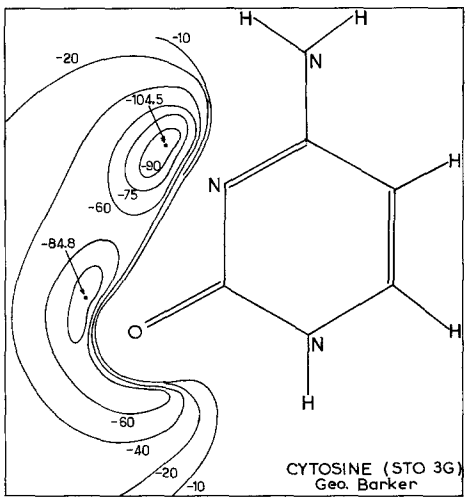

Fig. 2. Molecular electrostatic potential of cytosine computed in the STO 3G basis

allowed to approach the two possible sites in the direction defined by the potential minima of Fig. 2. The CN (or CO) distance d and the value of α were varied, as well as the angles $\theta_1 = \widehat{C_2NC}_{methyl}$ and $\theta_2 = \widehat{C_2OC}_{methyl}$. In the equilibrium positions ($d_{CN} = 1.49$ Å; $d_{CO} = 1.46$ Å; $\alpha = 109{,}5$; $\theta_1 = 120°$; $\theta_2 = 120°$), the interaction energies are equal to $-170{,}7$ and -171 kcal/mole for alkylation at N_3 and O_2 respectively. The two products have thus practically identical stabilities.

The decomposition of the binding energy into its components (Table 2) indicates that at the equilibrium distance, the global attraction is in favor of the nitrogen, this stemming essentially from the difference in the electrostatic attraction which is about 30 kcal/mole (20 in the case of the proton). But a new interesting difference

Table 2. Binding energies of CH_3^+ to N_3 and O_2 of cytosine[a]

	d	ΔE	E_C	E_{DEL}	E_{AT}	E_{EX}	$\Delta E'$
N_3	1.49	−170.7	−136.6	−199.3	−336.0	134.3	−201.7
O_2	1.46	−171.0	−105.5	−201.9	−307.4	105.4	−202.0

[a] Energies in kcal/mole. Components as defined in the text. E_{AT} is the sum of all the attractive components. d(Å) is the equilibrium CN or CO distance. ΔE is computed with respect to the energy of the CH_3^+ ion in its *initial* form. $\Delta E'$ and the components are computed with respect to the energy of the final deformed ion.

appears between the two positions, namely that the repulsive term is larger for the approach towards the nitrogen than for the approach towards the oxygen, the difference counterbalancing the difference in attraction. A similar feature has been found recently [15] in the case of the approach of Na^+ towards O_2 and N_3 of cytosine and also towards O_6 and N_7 of guanine, where it was rationalized on the ground that the valence-shell orbitals of the oxygen atom extend less in space than

those of the nitrogen; hence at a given distance the orbital overlap with the electron cloud of an approaching ligand is smaller for oxygen than for nitrogen, yielding a smaller repulsion.

3.3. Ethylation

The results of ethylation confirm and emphasize the trend observed on passing from protonation to methylation. In the first stage of our computation the nearly planar

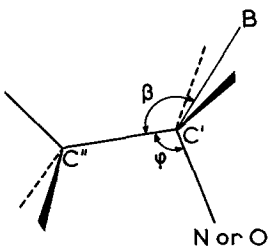

Fig. 3. Notations for the approach of the ethyl cation. *B* is the bissectrix of $\widehat{HC''H}$. Note that $\theta_1 = \widehat{C_2NC'}$; $\theta_2 = \widehat{C_2OC'}$

optimal arrangement of Fig. 1c [11] was assumed for the CH_2 group of $C_2H_5^+$. When this ion is allowed to approach cytosine in this state and with the C′C″ bond (see Fig. 3 for definition of the notations) in the plane of the ring, the approach towards N_3 is strongly hindered as indicated by the fact that a much more favorable energy of interaction is obtained when C′C″ is rotated 90° out of the plane of cytosine even for a distance NC′ as large as 1.60 Å. No such effect appears for an approach towards the oxygen atom, for which the in-plane location of the C′C″ bond is favored, even at a distance OC′ = 1.55 Å for the quasi-planar ion. The decomposition of the interaction energy into its previously defined components indicates clearly the phenomenon involved (Table 3): for the in-plane orientation of C′C″, the attractive terms are more favorable at 1.60 Å of the nitrogen atom than when C′ approaches the oxygen atom, even at a closer distance (1.55 Å), due to the stronger electrostatic

Table 3. Binding energies of $C_2H_5^+$ to N_3 and O_2 of cytosine[a]

Atom	*d*		*ΔE*	E_C	E_{DEL}	E_{AT}	E_{EX}	*ΔE′*
N	1.60	(I)	−9.9	−154.8	−162.9	−317.7	307.8	
		(II)	−97.5	−116.7	−161.8	−278.5	181.0	
O	1.55	(I)	−107.9	−95.5	−155.5	−251.0	143.1	
N	1.50	(I)	−124.0					
		(II)	−134.8	−136.6	−193.2	−329.8	178.8	−151.0
O	1.48	(I)	−140.2	−104.7	−187.8	−292.5	136.0	−156.4
		(II)	−140.5					

[a] Components (kcal/mole) as defined in the text. E_{AT} as in Table 2. *d*(Å) is the distance C′N or C′O. *ΔE* is computed with respect to the energy of $C_2H_5^+$ in its *most stable initial* form. *ΔE′* and the components in the final geometry are computed with respect to an isolated $C_2H_5^+$ in its final form. I: C′C″ in the plane of cytosine; II: C′C″ rotated by 90° out of the plane.

component, but at the same time the exchange repulsion is much larger for N_3 than for O_2, so that the final balance is largely in favor of the oxygen. The out-of-plane rotation of C′C″ by 90° decreases appreciably the repulsive component, but not sufficiently to make the total energy more favorable than in O_2.

In the second step of our exploration we have permitted the deformation of the valence and dihedral angles at the methylene carbon and have optimized the distances of approach C′N and C′O respectively and the orientation angles $\theta_1 = \widehat{C_2NC'}$ and $\theta_2 = \widehat{C_2OC'}$. The other distances and the geometry of the end methyl group were kept constant.

The most favorable geometry for nitrogen binding was found for $d_{C'N} = 1.50$ Å, $\beta = 138°$, $\varphi = 111°$, $\theta_1 = 120°$, with C′C″ rotated 90° out of the ring plane, with a binding energy of -134.8 kcal/mole (against -124.0 kcal/mole when C′C″ lies in the ring plane).

For oxygen binding, the most stable arrangement corresponds to $d'_{ON} = 1.48$ Å, $\beta = 138°$, $\varphi = 111°$, $\theta_2 = 120°$, with $\Delta E = -140.2$ kcal/mole for C′C″ in the cytosine plane, and $\Delta E = -140.5$ kcal/mole for C′C″ rotated by 90°.

It thus appears that the fixation on O_2 is preferred over fixation on N_3 by a few kcal/mole and, further, that the binding energy of the ethyl cation to the oxygen atom is practically indifferent to the orientation of the C′C″ bond, a situation which does not occur at the nitrogen position where the in-plane orientation remains strongly disfavored. The energy decomposition at equilibrium confirms the trends observed initially, particularly the rôle played by the exchange repulsion component of the binding energy in the preference observed (see Table 3).

4. Conclusions

In conclusion, within the limits of accuracy of this study, it appears that the intrinsic preferences of the H^+, CH_3^+ and $C_2H_5^+$ cations towards the two possible essential binding sites on cytosine, N_3 and O_2, undergo a continuous modification with the increase of size and complexity of the cation. N_3 is the preferred site of fixation of H^+, O_2 appears as the preferred site of fixation of $C_2H_5^+$, while the CH_3^+ ion does not seem to have a marked preference for any of these sites. This conclusion throws some light on the mechanism of attack by the various alkylating agents: if the reaction were of the S_N1 type (direct attack by the cation) ethylation should occur preferentially on O_2 but methylation should be observed in similar amounts on either O_2 or N_3. This proposition may be used to interpret the experimental data of Ref. [4] giving the percentage of products of cytidine alkylation by various agents. Among the ethylating agents, ethylnitrosourea shows a clear preference for O_2-ethylation over N_3-ethylation, whereas ethylmethanesulfonate and diethylsulfate show a definite preference for N_3, with some proportion of O_2-ethylation. This, together with our conclusion above, suggests an S_N1 mechanism in the case of the nitroso compounds, but rather an S_N2, or better, a mixture S_N1, S_N2 for the two other agents.

On the other hand, the fact that dimethylsulfate yields 99% of N_3-methylcytidine appears to indicate that there is no participation of an S_N1 mechanism in this case.

Finally, methylnitrosourea yields about three times as much N_3-methylation as O_2-methylation of cytidine [4]: this seemingly indicates that the action of the methylnitroso-agent does not obey a pure S_N1 mechanism.

Our conclusions concerning the sulfates and sulfonates appear in agreement with the general contention on their mechanism of action [16–18]. As to the nitrosoureas, their classification [17] must probably be more delicately shaded.

On the whole the results of the present work are consistent with the effects observed upon cytidine alkylation. They are also consistent with the observation that ethylating agents react to a higher extent with the ring oxygen of guanine than do methylating agents both in TMV-RNA [3] and in DNA *in vitro* [19] and *in vivo* [20] and with the fact that ethylnitrosourea shows more affinity for O_6 than for N_7 of guanine, in contrast to methylnitrosourea [3, 20].

References

1. Lawley, P. D.: Progr. Nucl. Acid. Res. Mol. Biol. **5**, 89 (1966)
2. Singer, B.: Progr. Nucl. Acid. Res. Mol. Biol. **15**, 219 (1975)
3. Singer, B., Fraenkel-Conrat, H.: Biochem. **14**, 772 (1975)
4. Singer, B.: FEBS Letters **63**, 85 (1976)
5. Bonaccorsi, R., Scrocco, E., Tomasi, J.: J. Chem. Phys. **52**, 5270 (1970)
6. Pullman, A.: Chem. Phys. Letters **20**, 29 (1973)
7. Dreyfus, M., Pullman, A.: Theoret. Chim. Acta (Berl.) **19**, 20 (1970)
8. Bonaccorsi, R., Pullman, A., Scrocco, E., Tomasi, J.: Theoret. Chim. Acta (Berl.) **24**, 51 (1972)
9. Bonaccorsi, R., Scrocco, E., Tomasi, J., Pullman, A.: Theoret. Chim. Acta (Berl.) **36**, 339 (1975)
10. Roothaan, C. C. J.: Rev. Mod. Phys. **23**, 69 (1951)
11. Hehre, W. J., Stewart, R. F., Pople, J. A.: J. Chem. Phys. **51**, 2657 (1969)
12. Barker, D. L., Marsh, R. E.: Acta Cryst. **17**, 1581 (1964)
13. Lathan, W. A., Hehre, W. J., Pople, J. A.: J. Am. Chem. Soc. **93**, 810 (1971)
14. Kindly communicated to us by Prof. J. Tomasi
15. Perahia, D., Pullman, A., Pullman, B.: Theoret. Chim. Acta (Berl.) **43**, 207 (1977)
16. Ross, W. C. J.: Biological alkylating agents. London: Butterworths 1962
17. Lawley, P. D.: Molecular mechanisms for repair of DNA, Part A, p. 25, P. C. Hanwalt and R. B. Setlow, Eds. New York: Plenum Press 1974
18. Organic chemistry of nucleic acids, N. K. Kochetkov and E. I. Bodovskii, Eds., English Translation. London–New York: Plenum Press 1972
19. Lawley, P. D., Shah, S. A.: Chem. Biol. Interactions **7**, 115 (1973)
20. Sun, L., Singer, B.: Biochem. **14**, 1802 (1975)

Received March 28, 1977