SEMIEMPIRICAL CALCULATIONS OF ALKYLATION AND PROTONATION ENERGIES OF BASES OF NUCLEIC ACIDS

Stanislav MIERTUŠ[a] and Magdaléna TREBATICKÁ[b]

[a] *Department of Analytical Chemistry, Slovak Institute of Technology, 812 37 Bratislava and*
[b] *Institute of Physics and Biophysics, Comenius University 842 15 Bratislava*

Received March 30th, 1983

The semiempirical quantum-chemical MINDO/3 method has been used to study protonation and alkylation of bases of nucleic acids: cytosine, thymine, uracil, adenine, and guanine. The optimum sites of attack of the reagents $H^{(+)}$, $CH_3^{(+)}$, and $C_2H_5^{(+)}$ have been found among selected nucleophilic positions of the mentioned bases. At the same time interaction energies of protonation and alkylation have been expressed, and relative affinity of the considered positions to the protonation and alkylation has been evaluated.

Investigation of relations between mutagenity and carcinogenity of some compounds showed that a change of a cell into a cancerous one is caused by changes in its genetic code. Alkylating agents represent a numerous group of such compounds and have attracted attention recently both in experimental and theoretical research. In the experimental field the most important papers are those by Singer[1-5] dealing with methylation and ethylation of DNA, RNA and their subunits *in vitro*, the papers[6,7] dealing with experimental *in vitro* alkylation and enzymatic release of the alkylated bases, and papers[8-10] by Rajewsky & Goth describing experiments carried out at the *in vivo* level.

The theoretical studies are focused on search for the sites in nucleic acids which are most suitable for the protonation or alkylation. These studies can use either the supermolecular approach, considering explicitely both the alkylated base and the alkylating agent, or a simplified approximative approach which only evaluates affinity to protonation from calculations of bases (*e.g.* from electrostatic potentials near the protonation sites). The supermolecular *ab initio* calculations of the protonation of the bases are dealt with in papers[11,12] by Mezey and Ladik and those[13-17] by Pullman and coworkers. The latter studies at the *ab initio* level involve calculations of topology of electrostatic potential near the molecule[13-16] as well as complete calculations of the *ab initio* energies of protonation, methylation and ethylation of cytosine[17]. So far, however, no complete study is available on the protonation and alkylation of all the bases which would enable an evaluation of influence of size

of alkylating agent. Therefore, the present communication gives results of calculations of the protonation and alkylation energies of the bases of nucleic acids and the optimum sites of attack by the reagents $H^{(+)}$, $CH_3^{(+)}$, and $C_2H_5^{(+)}$ in the protonation and alkylation.

METHODS AND CALCULATION

The semiempirical quantum-chemical MINDO/3 method[18] was used for the calculation of the protonation and alkylation energies of the bases of nucleic acids — cytosine, thymine, uracil, adenine, and guanine — with application of a version of the standard program GEOMO[19]. The individual interaction energies were computed from the relation:

$$E = E_{A-B} - (E_A + E_B), \qquad (1)$$

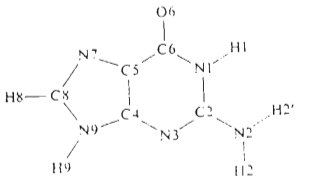
where E_A and E_B mean energies of the alkylating cation (proton) and the base, respectively, and E_{A-B} is the energy of the alkylated (protonated) base calculated in the optimum position of the reagent with respect to the base. The optimization of mutual positions of the reagent and base was carried out by the Rinaldi's procedure[20]. In all the cases optimized were the distance (d) of the attacking reagent and the attacked atom of the base and the bond angle φ always defined with respect to the atom with lower serial number (for numbering of atoms in the individual bases see Fig. 1). In well-grounded cases the dihedral angles ω and/or ω' were also optimized whose meaning is schematically given in Fig. 2. Also these angles are defined always with respect to the atoms with lower ordinal numbers.

Internal geometry of the bases was not optimized in the calculations, but the experimental geometry[21–23] was used. Such approach was chosen, because optimization of all geometry parameters of such large molecules has extremly high computer needs. In accordance with ref.[17], different geometries were considered for the alkyl groups before and after their binding to the bases. The choice of the protonation and alkylation sites followed the charge distribution, the positions with high electron densities being chosen.

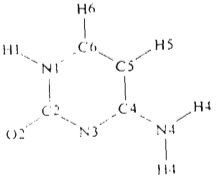
RESULTS

Two positions were considered for the protonation and alkylation of cytosine: N(3) and O(2). The optimization of the protonation and methylation involved the bord length d and bond angle φ, that of the ethylation also involved the dihedral angle ω' with respect to the second carbon atom of the ethyl cation. The individual quantities are explained in Fig. 2 for the case of ethylation of cytosine at N(3) position. The same approach was also chosen for the other bases. Table I gives the values of the protonation, methylation, and ethylation energies. The given energies correspond to the optimum values of geometry parameters summarized in Table II. Table II also gives values of the dihedral angles involving the first atoms of the attacking reagents, although these angles were not optimized in the calculations. In thymine we considered the interaction with the protonation and alkylation reagents at the O(2) and O(4) atoms. The calculated energy values are given in Table I, the corresponding values of the optimum geometry parameters are given in Table II.

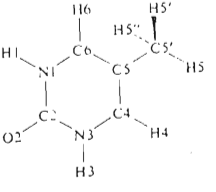
The uracil molecule was considered to be protonated (alkylated) at the O(2) and O(4) atoms. The calculated values of energies and geometry parameters are given in Tables I and II, respectively. The protonation and alkylation of adenine was considered to take place at the N(1), N(3), and N(7) atoms. Tables I and II give the respective



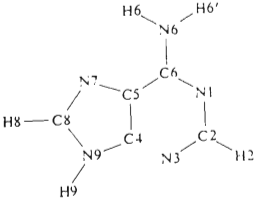
Guanine



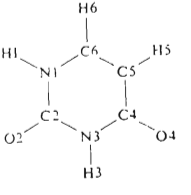
Cytosine



Thymine



Adenine



Uracil

FIG. 1
Numbering of the atoms in bases

calculated values of interaction energies and the optimum geometry parameters. The selected sites of the protonation and alkylation of guanine molecule were the atoms N(3), N(7), and O(6). Values of the calculated interaction energies and the optimum geometry parameters are given in Tables I and II, respectively. In the cases of the two nitrogen sites of guanine and all the sites of adenine, the values of dihedral angle ω' were not optimized, the value $\omega' = 90°$ is taken from ref.[17].

DISCUSSION

General problem to be solved is choice of suitable quantum-chemical method which would describe the studied phenomena with sufficient correctness. In the papers[13–17,24] by Pullman et al. the affinity to alkylation was determined by means of electrostatic potential and so called steric accessibility. This procedure has the advantage in that it works with the wave function obtained by *ab initio* method. On the contrary, this approach has obvious drawbacks, especially in that it expresses explicitely only the coulombic part of the interaction energy, whereas the other components (particularly the energy due to charge transfer, exchange repulsion, and polarisation) are not considered. Thereby applicability of such approach is limited to the protonation only, which was indicatively confirmed by the calculations[17] for cytosine.

It appears that supermolecular approach must be used, if differences in preference in methylation and ethylation are to be evaluated. Thereby, other components of the interaction energy (*i.e.* $E_{CT} + E_{ER} + E_{Pol}$) are explicitely involved, which makes it possible to evaluate quantitatively the preference of the bases when the alkylating reagent is changed. Moreover the method used by Pullman and coworkers[24] cannot evaluate simultaneously the preference caused by the values of electrostatic potentials and steric accessibility. Thus superiority of the supermolecular approach is even more distinct.

On the other hand, the enormous computer needs prevented application of the non-empirical *ab initio* method, so the greatest drawback of the approach chosen by us is the application of semiempirical methods. Therefore, our study is focused

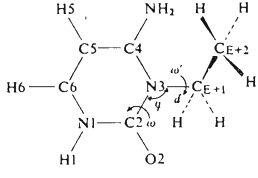


FIG. 2
Definition of the torsion angles

TABLE I

Energies of protonation, methylation, and ethylation of bases of nucleic acids (kJ mol^{-1})

Base	Atom	E_p[a]	E_m[a]	E_e[a]
Cytosine	N(3)	−845·8	−412·9	−176·7
	O(2)	−861·8	−497·1	−349·1
Thymine	O(2)	−772·1	−417·9	−271·0
	O(4)	−789·7	−438·2	−285·7
Uracil	O(2)	−755·0	−403·7	−252·8
	O(4)	−798·3	−440·1	−285·5
Adenine	N(1)	−941·8	−511·7	−275·6
	N(3)	−927·2	−509·5	−276·7
	N(7)	−90·53	−507·4	−269·8
Guanine	N(3)	−858·9	−424·8	−181·4
	N(7)	−910·1	−514·5	−289·0
	O(6)	−860·8	−494·7	−337·5

[a] E_p energy of the protonation, E_m energy of methylation, E_e energy of ethylation.

TABLE II

Values of the optimum geometry parameters for the protonation, methylation, and ethylation

Base	Atom	Protonation[a]			Methylation[a]			Ethylation[a]			
		d	φ	m	d	φ	ω	d	φ	ω	ω'
Cytosine	N(3)	1·028	117·3	180	1·490	116·9	180	1·584	114·8	180	270·2
	O(2)	0·953	121·6	180	1·364	144·8	180	1·414	139·2	180	0·0
Thymine	O(2)	0·951	126·0	180	1·362	150·0	180	1·418	144·1	180	179·3
	O(4)	0·951	125·0	180	1·366	149·9	0	1·422	144·0	0	177·3
Uracil	O(2)	0·951	126·8	180	1·364	152·3	180	1·433	141·8	180	151·9
	O(4)	0·953	121·8	180	1·365	149·7	0	1·419	146·7	0	140·7
Adenine	N(1)	1·029	116·9	180	1·480	113·9	180	1·560	112·8	180	90
	N(3)	1·026	124·5	180	1·473	120·8	180	1·548	121·5	180	90
	N(7)	1·012	133·1	180	1·446	135·2	180	1·518	136·8	180	90
Guanine	N(3)	1·023	122·9	180	1·477	124·2	180	1·564	126·0	180	90
	N(7)	1·016	128·1	180	1·452	131·4	180	1·523	130·7	180	90
	O(6)	0·952	125·2	0	1·263	151·4	0	1·419	142·6	0	179·3

[a] The bond length values in 10^{-10} m, the angle values in degrees.

on evaluation of relative changes in the alkylations of the individual bases with the individual attacking reagents.

The results given in this paper were obtained by the MINDO/3 method. The initial calculations were carried out by two methods: CNDO/2 and MINDO/3; the former, however, appeared not to be able to describe correctly the trend of decreasing interaction energy going on from the protonation to alkylations, so the MINDO/3 results are only given. The absolute values of the protonation energies obtained by the MINDO/3 method are higher than those deduced from experimental data. However, with respect to the fact that the MINDO/3 method was parametrized for calculation of heats of formation, it can be expected that relative changes of the protonation and alkylation energies are expressed by this method with qualitative correctness. Moreover, the MINDO/3 method gives lower values as compared with the *ab initio* calculations of the protonation energies[11,12].

Cytosine

Two interaction sites in cytosine were considered: N(3) and O(2). From Table III it is seen that oxygen has greater affinity to the proton. The electrostatic potentials indicate a deeper minimum at N(3) of cytosin. The *ab initio* calculations[12,17] gave

TABLE III

Relative values of energies of the protonation (E_p), methylation (E_m), and ethylation (E_e) in kJ . mol^{-1}

Base	Atom	E_p	E_m	E_e
Cytosine	N(3)	16·0	87·4	172·4
	O(2)	0	0	0
Thymine	O(2)	17·6	20·3	14·7
	O(4)	0	0	0
Uracil	O(2)	43·3	36·4	32·7
	O(4)	0	0	0
	N(1)	0	0	1·1
Adenine	N(3)	14·6	2·2	0
	N(7)	36·5	4·3	6·9
Guanine	N(3)	51·2	89·7	156·1
	N(7)	0	0	48·5
	O(6)	49·3	19·8	0

different results in spite of using the same STO 3G basis set. In the paper[17] the protonation of N(3) position appears more stable by 7·5 kJ mol^{-1}, whereas the other results[12] prefer the O(2) position by 42·1 kJ mol^{-1}. These differences can be due to differences in the geometries used. The *ab initio* calculations also showed that in the protonation the most significantly stabilizing component is the charge-transfer energy[17]. A similar trend was obtained in our calculations, too. Methylation and ethylation were considered at the same position of cytosine. The affinity order was maintained, the preference of the affinity of oxygen being gradually increased. Repulsive component of the interaction energy becomes more important here and makes the alkylation at O(2) more favourable. The same trend was also obtained from *ab initio* calculations[17] which show increasing affinity of O(2) with increasing size of the alkylating agent. However, the overall affinity increase is much milder than in the MINDO/3 method. Our calculation also gave the optimum geometry of the adducts, *i.e.* the protonated and alkylated bases (Table II). In the systems studied by us the calculated optimum geometries can be considered qualitatively correct. The obtained optimum N—H$^+$, and O—H$^+$ bond lengths agree well with the *ab initio* calculations[12]. Both in the methylation and ethylation the N—C bond appears longer than the O—C bond, but the N—C bond length obtained in the ethylation seems much too long as compared with the real values. The optimum bond angles φ calculated by the MINDO/3 method for the protonation at N(3) are very similar to those of the alkylation. When the reagent is bound to the O(2) position, the obtained values of the bond angles φ are greater in the methylation and ethylation than those expected *e.g.* on the basis of the *ab initio* calculations[17] for the protonation and alkylation of cytosine.

Thymine

The protonation and alkylation were considered at two positions: O(2) and O(4). The preference of O(4) to O(2) (Table III) agrees with the order of electrostatic potential values[13] and also agrees qualitatively with the ab initio calculations of the protonation[12] where the preference difference is up to 46·6 kJ mol^{-1}. The obtained result is noteworthy with respect to the presence of methyl group adjacent to O(4). The decisive factor of the preference seems to be basicity of the individual atoms. When changing the reagent from $H^{(+)}$ to $CH_3^{(+)}$ and $C_2H_5^{(+)}$, the preference order is maintained, the difference being increased for methyl cation (Table III) but decreased in the ethylation. Similar trends to those of cytosine were obtained in the calculation of the optimum geometries. The MINDO/3 method gives greater bond angles in the case of binding methyl and ethyl cation. Noteworthy is the change in the optimum position of the alkyls at O(4) (Table II, the dihedral angle ω). Whereas the optimum position of the proton is closer to the methyl group than to H(3), in the case of methyl and ethyl cations the optimum position is on the opposite

half-plane, which is obviously connected with repulsion between the methyl group and alkylating agent.

Uracil

Structurally uracil resembles thymine, but it has no methyl group at C(5) position. In this case the protonation of O(4) is preferred again, the difference between the interaction energies of the protonation at O(2) and O(4) being still greater than that in the case of thymine, which seems to be the consequence of the absence of methyl group in the uracil molecule. The same affinity order is also obtained from values of the electrostatic potentials[14]. Increasing size of the reagent (from $H^{(+)}$ to $CH_3^{(+)}$ and $C_2H_5^{(+)}$) is accompanied by decrease in the difference of the interaction energies. The geometry optimization gives similar results for uracil and thymine. Only the dihedral angles ω are different, representing a certain deviation of the second carbon atom of ethyl group out of the plane of the uracil molecule.

Adenine

Three suitable sites in adenine were considered for the protonation and alkylation: N(1), N(3), and N(7). When modelling the protonation by our approach we obtained the order N(1) > N(3) > N(7). The same result was also obtained by the *ab initio* calculation[11] which, however, gives somewhat greater energy differences between the protonations of the individual position as compared with the values obtained by us. Besides other factors, this can also be due to the fact that in the cited *ab initio* calculations mutual positions of the proton and base were not optimized. The affinity obtained from the electrostatic potentials[15] shows another order: N(3) > N(1) > N(7), which supports the opinion proposed at the beginning of the discussion that, even in the case of protonation, neglection of other than coulombic contributions can lead to different results. The calculations taking into account size of the reagent showed that differences in affinities of the individual positions are considerably decreased. The affinity order is changed on going to the bulkiest reagent – ethyl cation: N(3) > N(1) > N(7) (Table III). This trend seems to agree well with the experimental work[8] by Goth & Rajewsky who showed that in ethylation of DNA the N(3) position of adenine is more favourable. At the same time, the calculations gave the optimum geometries of the adducts. After the protonation of the positions considered, no significant differences were found in bond lengths between the individual positions. The alkylation is accompanied by the expected increase of bond lengths. In the case of N—C bond, however, the MINDO/3 method seems to overestimate this increase, especially in the ethylation. It is noteworthy to compare the bond lengths between the individual positions after the protonation and the both alkylations: the shortest bond length is obtained after the attack at N(7) atom, this attack being, at the same time, most unfavourable energetically.

Guanine

Three positions of considerable affinity were considered in guanine: N(3), N(7), and O(6). The protonation gave the order N(7) > O(6) > N(3) (Table III) which agrees with that of the minima of electrostatic potentials[16]. The *ab initio* calculations[11] using the supermolecular approach did not consider the protonation of the N(3) position. Therefore, it is impossible to carry out a complete comparison. The affinity of the N(7) position is, however, greater than that of O(6) according to these calculations, too. Effect of size of the attacking reagent was also studied. Replacement of $H^{(+)}$ by $CH_3^{(+)}$ does not affect the affinity order, but the affinity to O(6) is considerably increased (Table III). The same trend continues when replacing $CH_3^{(+)}$ by $C_2H_5^{(+)}$: the values of relative affinity are rather changed, which also results in a changed affinity order in the ethylation: O(6) > N(7) > N(3). It is interesting to compare these results with the experimental results[1,2]. It was found[1] that N(7) atom of guanine represents the preferred position for the protonation and methylation whereas for some ethylation reagents the O(6) position becomes the most reactive. The geometry optimization found the same trends as those of the above-studied molecules. According to expectation, the obtained optimum O—X bond lergths were shorter than N—X, the N(7)—X bond length being always shorter than N(3)—X. This trend is maintained for all the three attacking reagents (*i.e.* $H^{(+)}$, $CH_3^{(+)}$, $C_2H_5^{(+)}$).

CONCLUSION

The main aim of the present paper consists in evaluation of preference of individual positions of bases of nucleic acids to the protonation and alkylations. It is seen that application of the semiempirical MINDO/3 method gives qualitatively equivalent information to those given by more exact *ab initio* method in the extent in which the latter were available for comparison. The qualitative agreement with the indirect experimental data obtained from the protonation and alkylation of the bases or DNA is also noteworthy, in spite of the fact that our study used an approximative approach – *e.g.* internal geometry of the bases was not optimized, solvation effect and influence of the macromolecular DNA chain were not considered.

REFERENCES

1. Singer B., Fraenkel-Conrat H.: Biochem. J. *14*, 772 (1979).
2. Sun L., Singer B.: Biochem. J. *14*, 1795 (1979).
3. Singer B.: Nature (London) *264*, 333 (1976).
4. Singer B.: Progr. Nucl. Acid. Res. Mol. Biol. *15*, 219 (1975).
5. Singer B.: FEBS (Fed. Eur. Biochem. Soc.) Lett. *63*, 85 (1976).
6. Kirtikar D. M., Goldthwait D. A.: Prov. Nat. Acad. Sci. U.S.A. *71*, 2022 (1974).
7. Friedberg E. C., Goldthwait D. A.: Proc. Nat. Acad. Sci. U.S.A. *62*, 934 (1968).

8. Rajewsky M. F., Goth R.: *Screening Tests in Chemical Carcinogenesis* (R. Montesano, Ed.). Lyon, 1976.
9. Goth R., Rajewsky M. F.: Cancer Res. *32*, 1501 (1972).
10. Goth R., Rajewsky M. F.: Proc. Nat. Acad. Sci. U.S.A. *71*, 639 (1974).
11. Mezey P. G., Ladik J. J., Barry M.: Theor. Chim. Acta *43*, 251 (1980).
12. Hass E. C., Mezey P. G., Ladik J. J.: Theor. Chim. Acta *60*, 283)1981).
13. Perahia D., Pullman A: Theor. Chim. Acta *50*, 351 (1979).
14. Perahia D., Pullman A.: Theor. Chim. Acta *42*, 23 (1976).
15. Bonaccorsi R., Pullman A., Scrocco E., Tomasi J.: Theor. Chim. Acta *24*, 51 (1972).
16. Bonaccorsi R., Pullman A., Scrocco E., Tomasi J.: Theor. Chim. Acta *36*, 339 (1975).
17. Pullman A., Armbruster A. M.: Theor. Chim. Acta *45*, 249 (1977).
18. Bingham R. C., Dewar M. J. S., Low D. H.: J. Amer. Chem. Soc. *97*, 1285 (1975).
19. GEOMO Program, QCPE Version N. 295. Bloomington, Indiana.
20. Rinaldi D.: Computers Chemistry 1, 109 (1976).
21. Voet D., Rich A.: Progr. Nucleic Acid Res. Mol. Biol. *10*, 247 (1970).
22. Cochran J. W.: Acta Crystallogr. 4, *88* (1951).
23. Broomhead J. M.: Acta Crystallogr. *4*, 97 (1951).
24. Lavery R., Pullman A., Pullman B.: Int. J. Quantum Chem. *20*, 49 (1981).

Translated by J. Panchartek.