THOERETICAL STUDY OF THE TAUTOMERISM OF NUCLEOTIDE BASES IN THE GAS PHASE AND IN AQUEOUS MEDIA

B. Ya. Simkin

UDC 547.854.4.81'857:541.623'67

The relative stabilities of the tautomers of nucleotide bases in the gas phase and in aqueous media were estimated by the self-consistent field (SCF) MO method within the π-electron approximation. It is shown that the solvation energy, which depends substantially on the order of the tautomers with respect to their energies, must be taken into account for the accurate evaluation of the most stable tautomeric structure. The probabilities of spontaneous mutations in the gas phase and in aqueous media were estimated. The ionization potentials and electron affinities of the normal and "rare" tautomeric structures were calculated.

The development of a certain amount of spontaneous and induced genetic mutations is associated with the conversion of the molecules of nucleotide bases to "rare" tautomeric structures [1]. Thus adenine in the energetically less favorable (as compared with the amino form) imino structure is paired with cytosine instead of thymine, and thymine in the "rare" enol form is paired with guanine instead of cytosine. A knowledge of the relative stabilities of the tautomeric forms makes it possible to estimate the probability of mutations for a spontaneous mechanism [2].

Quantum-chemical methods with different accuracies, beginning with the simple Hückel MO method [3] and ending with the ab initio method [4], have been used for the study of tautomeric bases. All of these calculations were made within the isolated molecule approximation and disregarded the effect of the solvent on the relative stabilities of the tautomers. In addition, the results of experimental studies provide evidence for a substantial change in the percentages of the "rare" tautomeric structures in media with different polarities [5, 6].

Our recently proposed simple method for taking into account the effect of the dielectric properties of the medium within the π-electron approximation [7], which is a development of the Klopman method [8], makes it possible to estimate with good accuracy the position of the tautomeric equilibrium in media with different polarities. Despite the fact that specific solvation is not taken into account in this method, our calculations have shown [7] that this approach is satisfactory for the estimation of the relative stabilities of tautomers in the azomethine series. In the present paper the proposed method is used to estimate the relative stabilities of all of the possible tautomeric forms of nucleotide bases I-IV (a-f) in the gas phase and in the aqueous media.

The results of calculations of the heats of atomization and solvation energies are presented in Table 1. In contrast to the three other bases, guanine (III) has 12 possible tautomeric structures. The results of the calculation of only six tautomers, which, according to the calculations, are the most stable forms, are presented in Table 1.

Gas Phase

For isolated molecules the calculations predict that the completely aromatic tautomers (Ia, IIa, IIIa, and IVa) are the most stable forms. A similar result was obtained for cytosine by Goddard and co-workers [4] by the ab initio method and by Breen and Flurry [9] by the CNDO/2 (complete neglect of differential overlap) method. We note that the order of the cytosine tautomers with respect to their energies and the differences in the energies obtained by the ab initio and Pariser–Parr–Pople (PPP) methods with the Dewar σ, π parametrization are in agreement with good accuracy: (see following page).

We note that the results of our calculation are not in agreement with Dewar's calculations [10], according to which the Ib, IIb, IIIc, and IVb tautomers are the most stable forms. In contrast to the method in [10], we did

Rostov State University. Scientific-Research Institute of Physical and Organic Chemistry, Rostov-on-Don 344006. Translated from Khimiya Geterotsiklicheskikh Soedinenii, No. 1, pp. 94-99, January, 1978. Original article submitted December 13, 1976.

0009-3122/78/1401-0077$07.50

Tautomer	a	b	c	d	e	f
ab initio	0,00	17,7	23,7	28,8	41,4	29,3
PPP	0,00	17,4	18,0	25,0	29,1	31,8

not take into account the effect of σ polarization on the atomic parameters in connection with the fact that the procedure for the determination of the σ charges on the atoms by means of the difference in the electronegativities is unsatisfactory and also in view of the fact that, as later noted by Dewar [11], allowance for σ polarization does not affect the magnitudes of the heats of formation.

Aqueous Solutions

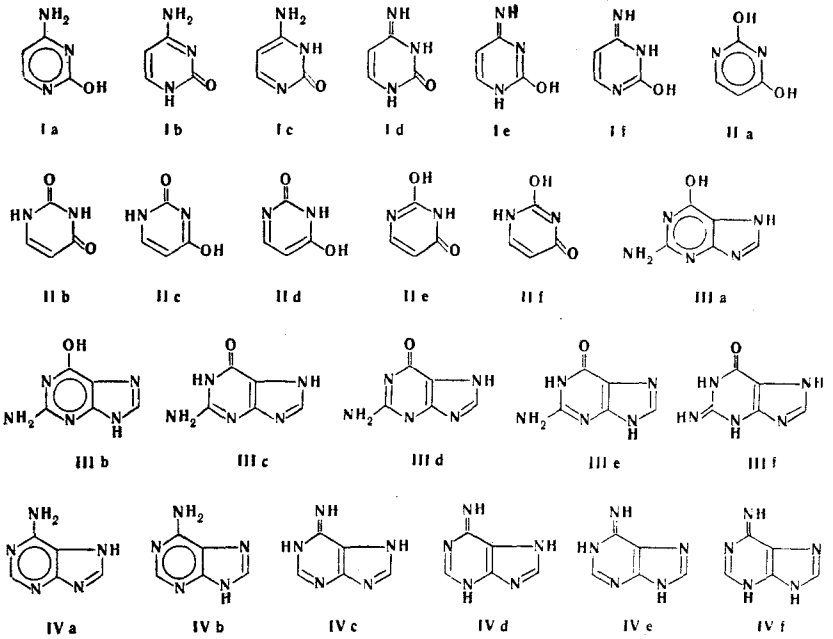
The available experimental data on estimation of the relative stabilities of nucleotide bases pertain primarily to aqueous solutions [12-18]. In this connection, we calculated the energies of atomozation of the tautomers and the solvation energies for $\varepsilon = 80$.

The most stable form of cytosine in water [16] is the "keto" tautomer Ib, which is confirmed by the results of our calculations (see Table 1).

In the case of uracil (thymine) in solution tautomer IIb becomes preferred over enol tautomer IIa due to the high solvation energy; this is in agreement with the data from the UV spectra [17].

Guanine also has a different relative order of tautomers (with respect to energy) in aqueous solutions than that observed for the gas phase. Structure IIIc becomes the most stable form; this is confirmed by the available experimental observations [16, 19].

In the case of adenine there is a discrepancy between the results of the calculations and the experimental data on the structure of the most stable tautomer. According to the data in Table 1, tautomer IVa should remain the most stable form also in aqueous solution; this contradicts the experimental NMR spectral data [15], according to which the $N_{(9)}H$ tautomer should have the lowest energy. The disagreement between the conclusions drawn on the basis of the calculations and the experiments is associated, first, with the failure to take into account in the calculations the energy of the HNH...N hydrogen bond, which develops in the $N_{(9)}H$ structure, and the rather strong H–H repulsion in the $N_{(7)}H$ tautomer, which also was disregarded in the π-electron calculation.



Of all of the nucleotide bases, only adenine in aqueous solutions has predominantly an aromatic structure; this is associated with the relatively low solvation energy in this case.

The data in Table 1 provide evidence that the simple method for taking into account the effect of the solvent on the relative stabilities of the tautomers leads to conclusions that are in agreement with the results of calculations with respect to the scheme in [7], which requires an increase in the counting time by a factor of two to three.

TABLE 1. Heats of Atomization and Solvation Energies of the Tautomers of Nucleotide Bases

Nucleotide base	Tautomer	E_{at}, kcal/mole	E_{solv}, kcal/mole	$E_{at}+E_{solv}$	$E_{at}+E_{solv}$*
Cytosine	a	1350,1	24,8	1374,9	1395,1
	b	1332,7	43,5	1376,2	1396,1
	c	1332,1	42,5	1374,6	1394,3
	d	1325,1	36,5	1361,6	1379,6
	e	1321,0	22,8	1343,8	—
	f	1318,3	22,3	1340,6	—
Uracil	a	1273,5	24,5	1298,0	1317,6
	b	1258,1	53,1	1311,2	1325,6
	c	1256,0	43,0	1299,0	1318,8
	d	1255,4	42,2	1297,6	1318,6
	e	1255,0	40,9	1295,9	1318,0
	f	1251,9	40,4	1292,3	1310,3
Guanine	a	1717,3	37,3	1754,6	1784,7
	b	1714,8	33,0	1747,8	1775,3
	c	1702,3	54,8	1757,1	1788,2
	d	1700,1	54,1	1754,2	1783,9
	e	1698,0	47,9	1745,9	1774,0
	f	1696,0	50,7	1747,3	1775,6
Adenine	a	1620,8	35,9	1656,7	1686,9
	b	1618,4	31,8	1650,2	1677,6
	c	1594,6	33,8	1628,4	1656,1
	d	1592,8	33,3	1626,1	—
	e	1591,3	28,6	1619,9	1651,4
	f	1589,3	29,8	1619,1	

* The calculation was carried out with minimization of the total energy of the molecule in a medium with a dielectric permeability (ε) of 80.

TABLE 2. Ionization Potentials and Electron Affinities of the Bases in the Gas Phase and in Aqueous Solutions

Nucleotide base	Tautomer	Gas		$\varepsilon=80$		Exptl. [23]
		I, eV	A, eV	I, eV	A, eV	I, eV
Cytosine	b	11,45	2,29	11,79	0,91	8,94
	d	11,73	2,02	11,81	1,14	
Uracil	b	12,30	2,15	11,90	1,08	9,50
	e	11,01	1,96	11,48	0,52	
Guanine	c	10,44	1,75	11,05	0,36	8,24
	b	10,16	1,04	10,56	0,22	
Adenine	b	10,64	1,55	11,15	0,51	8,44
	e	10,06	1,67	10,28	0,41	

The results of the calculations make it possible to observe the probability of the mutations that arise when one of the tautomers in a pair of bases exists in an unfavorable tautomeric structure. Taking into account the fact that one of the hydrogen atoms in DNA is replaced by a carbon residue, it is sufficient to observe, respectively, the b→d tautomerism for cytosine, the b→c tautomerism for uracil, the c→b tautomerism for guanine, and the a→c tautomerism for adenine. Using the data in Table 1, one may conclude that the probability of the development of the structures necessary for the mutations decreases in the order uracil > guanine > cytosine > adenine or in the order of differences in energies of 6.8, 12.9, 16.8, and 20.8 kcal/mole. Thus, uracil should be considered to be the base most liable to undergo tautomeric transformations in aqueous media, under the assumption that the probabilities of barrier-surmounting and tunneling mechanisms for proton transfer are identical for all of the nucleotide bases.

The corresponding ratios of the percentage of the stable tautomeric structure to the percentage of the "rare" form are as follows:

$$k_{urac}=\frac{[b]}{[c]}=10^{4.8};\quad k_{guan}=\frac{[c]}{[b]}=10^{9.2};\quad k_{cyt}=\frac{[b]}{[d]}=10^{12.0};\quad k_{aden}=\frac{[a]}{[c]}=10^{14.8}.$$

The relative orders of the values for uracil and cytosine are in agreement with the experimental conclusions [20, 21], according to which tautomerism should develop more easily for uracil than for cytosine, and refute the data of Lee and Chan [12, 13], which were criticized by Wong [14]. The calculated values are substantially higher than the experimental values.

According to the mutation mechanism proposed by Lowdin [22], proton exchange does not take place within the base as proposed by Watson and Crick but rather between pairs of bases, and this leads to the simultaneous development of two "rare" tautomeric forms. Assuming that the energy of interaction in the pair does not change substantially on passing to the tautomeric forms, which requires verification by theoretical calculations, it may be concluded that the probabilities of the development of rare pairs of bases are practically identical for adenine–uracil and guanine–cytosine structures (ΔE 27.6 and 29.7 kcal/mole, respectively)

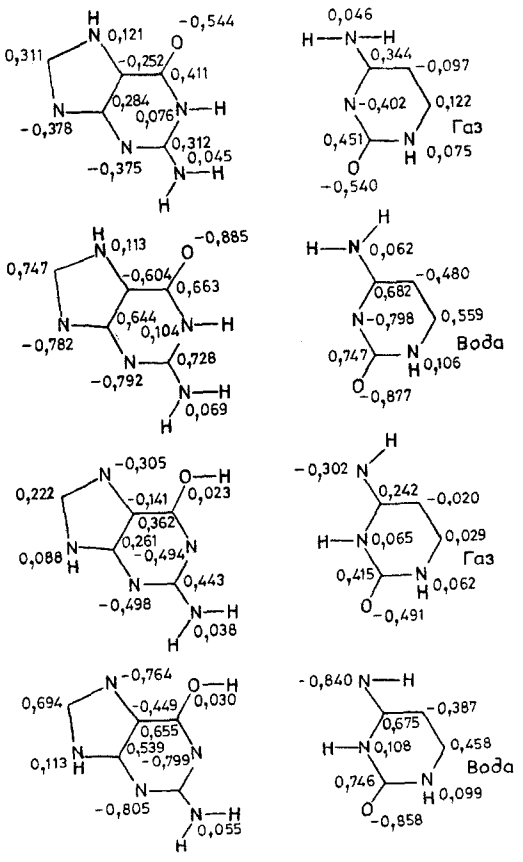
In the molecular diagrams presented below there are charges on the atoms in the guanine–cytosine pair for the gas phase and for the aqueous solution. Similar results for a "rare" tautomeric pair are also presented in the diagrams. It is easy to see that substantial rearrangement of the electronic structure occurs in a polar solvent; in particular, the charges on the heteratoms that introduce one π electron increase very markedly. The hydrogen bonding in aqueous media should become much stronger because of these effects.

Ionization Potentials and Electron Affinities

The ionization potentials and electron affinities play an important role in charge transfer processes between the nucleotide bases. The calculated ionization potentials and electron affinities of the nucleotide bases for the gas phase and aqueous solutions are presented in Table 2. The order of the bases with respect to increasing ionization potential does not change on passing from the gas phase to solution and is in agreement with the experimental order [23]: I(uracil) > I(cytosine) > I(adenine) > I(guanine). The ionization potentials were calculated by means of Koopman's theorem and are ~ 2.50 eV higher than the experimental values.

The order of the bases with respect to the ionization potentials changes substantially for the "rare" tautomeric forms: I(cytosine) > I(uracil) > I(guanine) > I(adenine).

Thus the calculations provide evidence for the substantial effect of the solvent on the tautomerism of the nucleotide bases and, despite the crudeness of the solvent model, make it possible to give the correct order of the tautomeric forms with respect to increasing energy.



Methods of Calculation

The heats of atomization and the solvation energies were calculated by the self-consistent field (SCF) MO method within the π-electron approximation with the Dewar σ, π parametrization [24]. The parameters and calculations were given by us earlier [7, 25]. The effect of the solvent was taken into account through the dielectric permeability (ε) of the medium; the matrix elements of the Fock operator have the following form [7]:

$$F_{\mu\mu}=W_\mu+\frac{1}{2}P_{\mu\mu}\gamma_{\mu\mu}+\sum_{\nu\neq\mu}(P_{\nu\nu}-Z_\nu)\gamma_{\mu\nu}-\left(1-\frac{1}{\varepsilon}\right)\left[\frac{1}{r_\mu}(P_{\mu\mu}-Z_\mu)+0{,}75\sum_{\nu\neq\mu}(P_{\nu\nu}-Z_\nu)\gamma_{\mu\nu}\right]. \quad (1)$$

$$F_{\mu\nu}=\beta_{\mu\nu}-\frac{1}{2}P_{\mu\nu}\gamma_{\mu\nu}.$$

The W_μ, $P_{\mu\nu}$ and Z_μ values have their usual significance (for example, see [24], and r_μ are the van der Waals atomic radii).

In addition, to ascertain the possibility of taking the medium into account for the relative stabilities of the tautomers without variation of the electronic and geometrical structures of the molecule we calculated the solvation energy from simple formula (2), which is a development of the Born equation for the solvation of ions:

$$E_{\text{solv}}=-\left(\sum_\mu q_\mu^2\gamma_{\mu\mu}+\sum\sum_{\mu<\nu}q_\mu q_\nu\gamma_{\mu\nu}\right)\left(1-\frac{1}{\varepsilon}\right); \quad (2)$$

where q_μ is the charge on atom μ, and $\gamma_{\mu\nu}$ are the Coulombic integrals.

LITERATURE CITED

1. G. S. Stent, Molecular Genetics: An Introductory Narrative, Freeman (1971).
2. R. Rein, Y. Coeckelenbergh, and J. T. Egan, Int. J. Quantum Chem., Symp., 145 (1975).
3. B. Pullman and A. Pullman, Quantum Biochemistry, Wiley-Interscience (1963).
4. J. D. Goddard, P. G. Mezey, and I. G. Gsizmadia, Theor. Chim. Acta, 39, 1 (1975).
5. D. M. Brown and M. J. E. Hewlins, J. Chem. Soc., C, 2050 (1968).
6. G. A. Jeffrey and Y. Kinoshita, Acta Cryst., 16, 20 (1963).
7. B. Ya. Simkin, V. I. Natanzon, and V. I. Minkin, Teor. Eskp. Khim., 13, (1977).
8. G. Klopman, Chem. Phys. Lett., 1, 200 (1967).
9. D. L. Breen and R. L. Flurry, Theor. Chim. Acta, 23, 138 (1971).
10. N. Bodor, M. J. S. Dewar, and A. J. Harget, J. Am. Chem. Soc., 92, 2929 (1970).
11. M. J. S. Dewar and A. J. Harget, Proc. Roy. Soc., Ser. A, 315, 443 (1970).
12. G. C. Y. Lee, J. H. Prestegard, and S. I. Chan, J. Am. Chem. Soc., 94, 951 (1972).
13. G. C. Lee and S. I. Chan, J. Am. Chem. Soc., 94, 3218 (1972).
14. Y. P. Wong, J. Am. Chem., 95, 3511 (1973).
15. M. Dreyfus, G. Dodin, O. Bensaude, and J. E. Dubois, J. Am. Chem. Soc., 97, 2369 (1975).
16. C. L. Angell, J. Chem. Soc., 504 (1961).
17. J. R. Marshall and J. Walker, J. Chem. Soc., 1004 (1951).
18. G. S. Parry, Acta Cryst., 7, 313 (1959).
19. M. Spenser, Acta Cryst., 12, 59 (1959).
20. A. R. Katritzky and A. J. Waring, J. Chem. Soc., 1540 (1962).
21. G. W. Kenner, C. B. Reese, and A. R. Todd, J. Chem. Soc., 855 (1955).
22. P. O. Lowdin, Adv. Quant. Chem., 2, 213 (1965).
23. N. S. Hush and A. S. Cheung, Chem. Phys. Letters, 34, 11 (1975).
24. M. J. S. Dewar, Molecular Orbital Theory of Organic Chemistry, McGraw-Hill (1969).
25. V. I. Minkin, V. A. Kosobutskii (Kosobutski), B. Ya. Simkin, and Yu. A. Zhdanov, J. Mol. Structure, 24, 237 (1975).