ELSEVIER

Biophysical Chemistry 55 (1995) 231–237

Biophysical Chemistry

A theoretical and experimental investigation of the electronic spectra and tautomerization of nucleobases

Abdulrahman O. Alyoubi *, Rifaat H. Hilal

Chemistry Department, Faculty of Science, King Abdulaziz University, P.O. Box 9028, Jeddah 21413, Saudi Arabia

Received 6 September 1994; revised 14 December 1994; accepted 21 December 1994

Abstract

The electronic structures of all possible tautomers of uracil, thymine, cytosine, adenine and guanine have been carefully examined within the MNDO-MO frame-work. Equilibrium geometries are determined and the relative stabilities are discussed. Allowance for solvent effect on the stabilities is made by assuming a tetrahedral solvent cage with the DNA base occupying its centre. The electronic absorption spectra of the studied DNA bases, in solvents of different polarities are recorded and discussed. Assignments of the observed bands are facilitated using MNDO-CI computations. It is suggested that in solution the DNA bases are in some statistical mixtures of the most stable tautomers, and the Watson–Crick (WC) structure cannot account for the observed spectra alone.

Keywords: DNA bases; Tautomeric equilibria; MNDO; Electronic spectra

1. Introduction

Mutagenesis has remained an intriguing aspect to genetics since the beginning of this century, and its analysis has proceeded with the elucidation of gene replication and expression [1]. Although the bi-helical complementary base paired structure of DNA implies mispairing events as the principle means for generating spontaneous mutations, a general mechanism for mispairing has not emerged [2].

Watson and Crick [3] postulated that a base of DNA may undergo a tautomeric shift that alters its pairing properties. It has been elaborated [4] on this speculation by proposing that keto–enol or amino–imino and anti-syn-tautomerization (rotation of the base from its normal anti position through 180° to the uncommon syn position) enable a double helix to accommodate eight complementary mispairs without intolerable distortion.

A correlation between the geometrical isomerism of tautomeric forms of cytosine and its internal structural parameters was predicted by Ha and Gunthard [5].

By using matrix-isolation experiments Nowak et al. [6] have studied the tautomerism in the DNA bases. Base pair induced NMR shifts studies [7] in DNA indicated that rare tautomeric forms may be involved in mutagenesis.

A considerable number of theoretical studies have been published on the subject of base pair associations [8–11]. In those studies, focus was concen-

* Corresponding author.

SSDI 0301-4622(95)00002-X

trated on the relative stability of different associates in water and in nonpolar solvents. Little or no attention has been paid, however, to the effect of medium on the individual nucleobases [12].

Furthermore, no firm data exist concerning the frequency of the enol forms of guanine and thymine (or uracil), the imino forms of cytosine and adenine, either in neutral aqueous solution or in conditions that might be obtained in the immediate vicinity of the growing DNA chain.

The present work, presents theoretical as well as experimental investigation of the tautomeric equilibria for all DNA bases. All possible tautomeric forms (including rare tautomers) are theoretically considered. Electronic absorption spectra of DNA bases in solvents of different polarities are analyzed and quantitatively discussed.

2. Experimental

Merck analytical grade thymine, uracil, adenine, cytosine and guanine were used without further purification.

BDH spectral quality dioxane and absolute ethanol were used as solvents in all spectroscopic measurements which have been carried out on a 260-Shimadzu double beam spectrophotometer using 1 cm length quartz cells.

Stock solution of 10^{-3} M concentrations of the

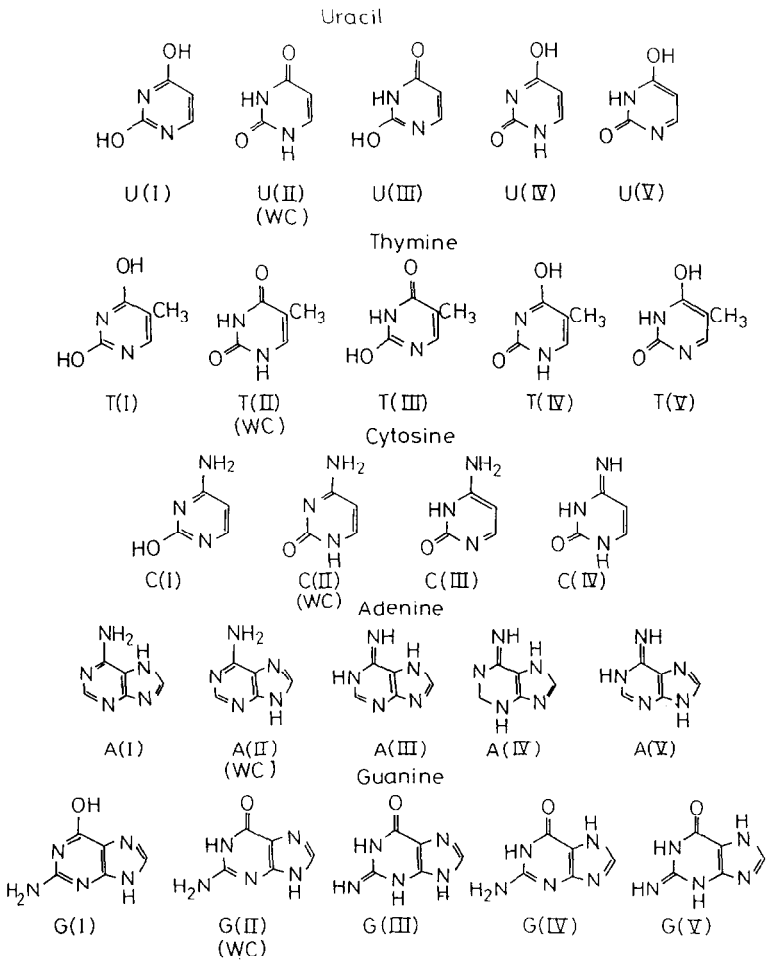

Fig. 1. All possible tautomeric structures of the nucleobases studied.

DNA bases studied were prepared by quantitatively dissolving the base in the given solvent. UV spectra were recorded for solutions in the $10^{-4} - 10^{-5}$ M concentration range prepared from stock solutions by direct dilution. No attempt has been made to adjust the pH of the aqueous solutions studied. The recorded spectra (absorbance vs. wavelength) are then recomputed and reported as molar absorptivity (ε) vs. wavelength. All spectra reported in this work are computer deconvoluted and Gaussian fitting of the broad observed bands were carried out. Deconvolution of the spectra is based on both the least squares fitting procedure and differentiation. Accurate peak parameters ($\lambda_{max}, \varepsilon_{max}$) are then extracted and reported [13,14].

2.1. Method of calculations

All calculations were performed within the MNDO frame-work [15] using the MOPAC program version 5.0. This version of MOPAC is supported by a MNDO, PM3 and AM1 Hamiltonians and calculation can be carried out using any of the three Hamiltonians quite simply. However, our previous experience [16] and various reports on MNDO computations have demonstrated success in the ability of this method to calculate isomerization and tautomerization energies, in various systems [17,18]. On the other hand, the PM3 approach is designed for the special case of hypervalent systems [19] and the AM1 appears in the literature with different parameters. Consequently results are heavily dependent on the parameters and/or program version used. Equilibrium geometries were determined by optimizing all ($3N$-6) variable parameters for each tautomeric form studied. Effect of electron correlation was taken into account using 100×100 CI computation that includes all singlet, doublet and highest multiplets arising from the higher three occupied MO's and the lowest three unoccupied MO's.

3. Results and discussion

3.1. Theoretical computations

Fig. (1) presents all tautomeric structures considered for the five nucleobases; equilibrium geometries are summarized in Table 1. It is evident that thymine and uracil possess the same main geometric features, indicating that methyl substitution has no major effect on the geometry in the tautomeric part. The equilibrium geometries obtained in this work using the MNDO-MO method are in excellent agreement with those obtained using ab initio MO methods [9].

Table 1
Equilibrium geometries of the WC structure and the most stable conformer (MSC) of the studied DNA bases

DNA base	Bond	R (Å)		Angle	θ (°)	
		WC	MSC		WC	MSC
Uracil	C_1-N_2	1.403	1.366	CNC	124.88	114.90
	C_1-N_7	1.397	1.350	NCO	117.55	112.80
	N_2-N_3	1.429	1.365			
	C_3-O_4	1.22	1.346			
Thymine	C_1-N_2	1.405	1.361	CNC	124.65	115.70
	C_1-N_7	1.403	1.345	NCO	116.53	110.24
	N_2-C_3	1.393	1.366			
	C_3-O_4	1.225	1.346			
Cytosine	C_1-N_2	1.398	1.362	CNC	119.65	115.37
	N_2-C_3	1.336	1.368			
	C_3-N_4	1.373	1.372	NCO	124.45	114.74
	C_1-N_7	1.382	1.354			
	C_1-O_8	1.224	1.342			
Adenine	C_5-N_7	1.388	1.397	CNC	106.70	105.45
	N_7-C_8	1.412	1.341	NCN	112.23	112.37
	C_8-N_9	1.334	1.398			
Guanine	C_1-N_2	1.398	1.383	CNC	124.38	118.08
	N_2-C_3	1.454	1.354	OCN	116.86	113.75
	C_1-N_6	1.378	1.354			
	C_3-O_{10}	1.221	1.341			
	C_1-N_{11}	1.376	1.369			

Energetics for the tautomerization processes are summarized in Table 2 for the five DNA bases studied in this work. For adenine the Watson–Crick (WC) tautomer seems to be the most stable, however, for the four other DNA bases, structures other than the WC structure predominate. In the case of uracil and thymine the all-enol structure is the most stable, the relative stability over the WC (all keto) form is 0.18 and 0.16 eV for uracil and thymine, respectively. Careful examination of the energetics of the other tautomeric forms of uracil (or thymine) indicates that enolization of the oxygen atom attached to carbon number 1 is energetically favoured over the enolization of that attached to carbon 3

(numbering system is shown in Fig. 2). This is most probably due to the fact that the carbon atom 1 lies between the two more electronegative nitrogen atoms. This could facilitate the migration of charge away from C_1 and hence allows the keto–enol tautomerism to shift to the keto-side.

Cytosine, on the other hand, shows a relative high stability for its enol form, C(I). This seems to be associated with the relative stability of the aromatic amine structure over the nonaromatic one. Guanine shows this same trend for its most stable structure, G(I).

It is interesting to notice that the energy differences quoted in Table 2 between the different tautomeric forms of the five DNA bases are all less than 0.5 eV. Our previous dynamic NMR investigations on rotational energy barriers in amides [20] and various other theoretical and experimental [21] studies have come to the conclusion that energy barriers less than 1 eV can be easily overcome at room temperature [21]. Hence, one would expect that tautomerization of the DNA bases studied might very well take place between these forms, in the gas phase. Furthermore, the WC conformers although not predicted to be the most stable the present calculation yet shows that WC forms have dipole moments larger than that predicted for the most stable conformers for the DNA bases studied. This would indicate that in polar media, WC forms would be stabilized to an extent greater than that for the gas-phase most stable conformer. This additional solvent dependent stability might very well stabilize WC forms over all other possible tautomeric forms studied. It is well established that medium effects play an important role in DNA base associations

Table 2
Heat of formation (ΔH_f), relative stabilities (ΔE), ionization potential (I.P.), dipole moments (μ) and solvent stabilization energies (ΔE_{solv}) of all tautomeric structures of the DNA bases

DNA base	Structure	$-\Delta H_f$ (kJ mol^{-1})	ΔE (eV)	I.P. (eV)	μ (D)	ΔE_{solv} (eV)
Uracil	U(I)	289.36	−0.18	9.86	3.61	0.11
	U(II) [a]	271.71	0.0	9.93	4.12	0.25
	U(III)	277.15	−0.05	9.47	2.02	
	U(IV)	245.93	0.26	9.73	6.93	
	U(V)	250.62	0.21	9.46	5.99	
Thymine	T(I)	317.30	−0.16	9.70	3.49	0.13
	T(II) [a]	302.04	0.0	9.77	4.03	0.21
	T(III)	308.95	−0.07	9.32	2.63	
	T(IV)	271.21	0.32	9.62	6.85	
	T(V)	280.70	0.22	9.38	5.85	
Cytosine	C(I)	89.12	−0.42	9.26	2.88	0.09
	C(II) [a]	48.03	0.0	9.41	6.02	0.29
	C(III)	31.63	0.17	8.98	6.91	
	C(IV)	32.84	0.16	9.50	4.47	
Adenine	A(I)	−265.47	0.35	9.07	0.10	
	A(II) [a]	−231.33	0.0	8.85	2.01	0.31
	A(III)	−298.57	0.69	8.84	2.92	
	A(IV)	−294.76	1.00	8.61	8.38	
	A(V)	−275.39	0.45	8.65	3.39	
Guanine	G(I)	−14.39	−0.28	8.74	2.92	0.11
	G(II) [a]	−42.38	0.0	8.56	5.69	0.28
	G(III)	−94.68	0.54	9.10	7.79	
	G(IV)	−56.48	0.15	8.68	1.99	
	G(V)	−74.35	0.33	9.12	3.89	

[a] The WC conformation (cf. Fig. 1).

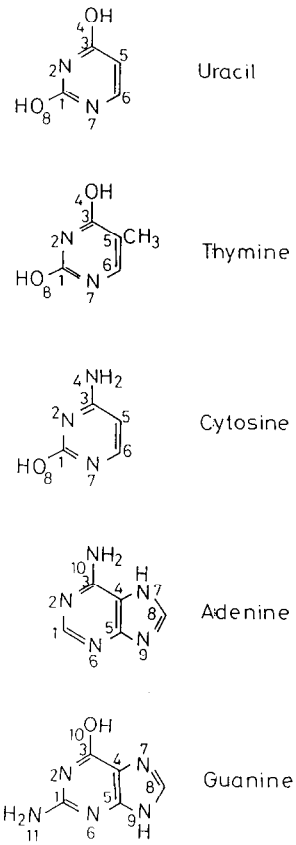

Fig. 2. Numbering system for the studied bases.

[22]. Various studies suggest that in the gas-phase and in nonpolar solvents e.g. CCl_4, $CHCl_3$, base associate mainly by H-bonding, whereas in water solution stacking configurations are preferred [23,24]. Efforts were undertaken to reflect the experimentally observed differences between the most stable conformations of solvated molecules and those in the gas-phase. The complexity of the theories applied ranges from simple correlations of the approximated molecular volumes with the free energy of solution [25], to quantum chemical calculations on stimulated associations of a single solute with some solvent molecules: the super molecule frame work [26]. This super molecule approach has been adopted in the present work to evaluate the solvent effect on the tautomeric equilibria studied.

In the present work, a cost-effective, yet symmetric supermolecule is proposed, in which a DNA base occupies the center of a tetrahedral solvent cage with four water molecules occupying the apices. The volume of the solvent cavity, that has to be created to insert a single solute into the solvent cage, has been calculated using the method proposed by Birnstock et al. [27]. Computation for this solvated DNA base has been carried out for tautomeric forms studied for the five DNA bases. Full geometry optimization (see Table 2) have been performed for the tetrahedral cage assuming a frozen DNA base geometry. It is evident that in each and every case, the WC conformer has been stabilized to an extent much greater than any other conformer. This solvent stabilization exceeds the gas-phase energy differences between the WC and the most stable structures.

3.2. Electronic absorption spectra

The electronic spectra of the different DNA bases reveal a problem not encountered in the study of hydrocarbons even those having many more atoms. First, all have more π electrons than centers making them isoelectronic with unknown hydrocarbon di- and tri-anions and therefore negating the possibility of simple correlations with better understood systems. Second, the presence of numerous non bonded pairs and mobile protons allows for tautomeric equilibria, both real and imagined for these DNA bases. The spectrum of uracil in dioxane, as a non polar solvent, shows one main envelope (Fig. 3) covering

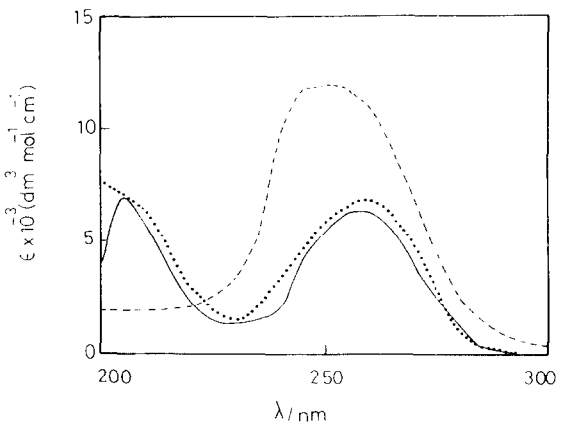

Fig. 3. Electronic absorption spectra of uracil in: (—) ethanol, (- - -) dioxane and (· · · · · ·) water.

the 280–235 nm region with a maximum at 250 nm. Upon increasing solvent polarity, using ethanol as a solvent, a red shift, a decrease in the intensity of the 250 nm envelope and a new band centered at 215 nm are observed. As one increases the polarity of the solvent the main 250 nm envelope seems to show a more defined shoulder at 240 nm.

Table 3
Experimentally observed and theoretically computed transition energies (ΔE) intensities (f) of the WC and MSC structures of the studied DNA bases

DNA base	Observed		Calculated WC		Calculated MSC	
	ΔE(eV)	f	ΔE(eV)	f	ΔE(eV)	f
Uracil	4.73	0.12	4.75	0.15	–	–
	5.12	0.07	–	–	5.01	0.11
	5.72	0.10	5.63	0.10	6.18	0.15
Thymine	4.56	0.17	4.55	0.21	–	–
	5.02	0.13	–	–	4.92	0.25
	5.99	0.15	5.86	0.18	5.72	0.19
Cytosine	4.47	0.13	4.40	0.17	–	–
	5.23	0.11	–	–	5.02	0.18
	5.99	0.15	5.75	0.25	5.86	0.30
Adenine	4.74	0.30	4.64	0.35	–	–
	5.10	0.11	–	–	4.93	0.21
	5.72	0.28	5.59	0.30	5.85	0.40
Guanine	4.39	0.16	4.31	0.22	–	–
	4.82	0.20	4.82	0.29	4.55	0.31
	5.23	0.05	–	–	5.19	0.12
	5.99	0.44	6.15	0.15	5.86	0.38

Table 3 presents the energies of the allowed $\pi-\pi^*$ transitions observed experimentally and computed theoretically for both structures U(I) and U (II) of uracil. It is quite obvious that the computed spectrum based on structure U(II), WC structure, fits to a remarkable extent the experimentally observed pattern, with the exception of the 240 nm band. Circular dichroism (CD) and magnetic circular dichroism (MCD) spectra indicate [28] the existence of the 240 nm absorption, in the spectra of uracil and thymine. However, all calculations together do not predict any transition in this region. Hug and Tinoco [29] suggested that an $n \rightarrow \pi^*$ transition might very well be the source of the 240 nm band. In the present work, the spectrum of uracil (and of thymine) has been recorded for a 0.1 M HCl solution of the base, and no detectable difference from the spectrum in pure water has been observed and the 240 nm band still persisted. This rules out the assumption of Hug and Tinoco that the 240 nm band is an out-of-plane $n-\pi^*$ transition, Theoretical computation of the spectrum of uracil, based on structure U(I), predicts a singlet $\pi-\pi^*$ transition at 242 nm. The aforementioned observations seems to suggest a statistical mixture of both structures U(I) and U(II) of uracil and a detectable shift of the tautomeric equilibrium may be achieved by varying solvent polarity. This same trend has been observed in the spectra of thymine (Fig. 4) and cytosine (Fig. 5). Comparison between the theoretically computed and the experimentally observed spectra is reported in Table 3.

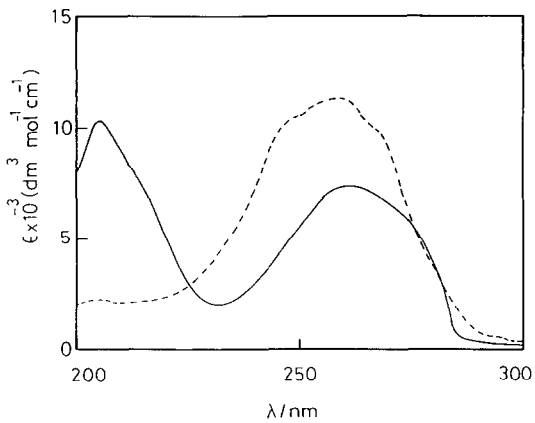

Fig. 4. Electronic absorption spectra of thymine in: (—) ethanol and (---) dioxane.

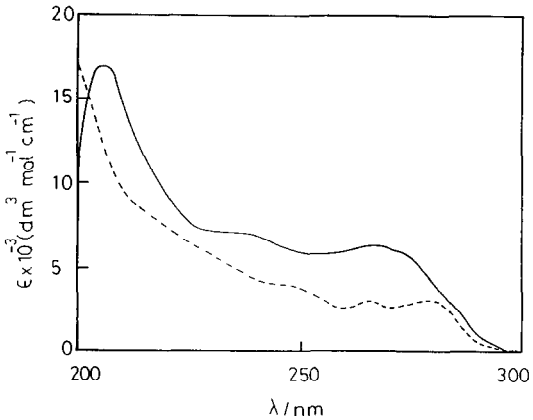

Fig. 5. Electronic absorption spectra of cytosine in: (—) ethanol and (---) dioxane.

The spectra of adenine and guanine show a very similar trend. A shoulder at 240 nm, which at first glance, was assigned as an out-of-plane polarization and thus an $n \rightarrow \pi^*$ classification [30]. Results of the present work indicates clearly that such an $n \rightarrow \pi^*$ assignment is quite misleading. Changing the solvent gradually from nonpolar to a polar hydrogen bonding, the 240 nm band appears to red shift, in contrast to the well known behaviour of $n \rightarrow \pi^*$ bands. The results of the present work elaborates on the polarized phosphorescence experiment of Cohen and Goodman [31] on purine which show no hint of out-of-plane polarization under 240 nm excitation. MNDO-MO computation on the WC conformation of adenine and guanine predicted the experimentally observed spectra successfully except for the 240 nm band which has not been theoretically predicted for both compounds. This same behaviour has been reported for other theoretical models (e.g. CNDO) and has been attributed to an inherent draw back of the semi empirical models. However, this seems quite misleading since these models were able to satisfactory predict the spectra of species very similar to the studied DNA bases. Theoretical computation of the spectra of adenine and guanine based on structures A(I) and G(I) predict moderate intensity band at the 240–250 nm region. This would lead to the conclusion that in solution, statistical mixtures, of WC and the most stable conformers of various DNA bases, exist and solvent polarity plays a pronounced role in shifting these tautomeric equilibria.

References

[1] P.O. Lowdin, Adv. Quantum Chem., (1965) 213.
[2] B. Pullman and A. Pullman, Quantum Biochemistry, Interscience Publishers, New York, 1963.
[3] J.D. Watson, and F.H. Crick, Nature (London), 171 (1953) 737.
[4] J. Donohue, J. Proc. Natl. Acad. Sci. USA, 42 (1956) 60; J. Donohue and K.N. Trueblood, J. Mol. Biol., 2 (1960) 363.
[5] T.K. Ha, and H.H. Gunthard, J. Mol. Struct. Theochem., 95 (1992) 209.
[6] M.J. Nowak, L. Lapinsk, J.S. Kwiatkowski, and J. Leszynski, Spectrochim. Acta, 47A (1991) 87.
[7] G.V. Fazakerly, Z. Gdaniec, and L.C. Sowers, J. Mol. Biol., 230 (1993) 6.
[8] I.V. Poltev and N.V. Shulyapina, J. Biomol. Struct. Dyn., 3 (1986) 739.
[9] P. Hobza and C. Sandorfy, J. Am. Chem. Soc., 109 (1987) 1302.
[10] A. Pohorille, R.L. Pratt, S.k. Burt and R.D. MacElroy, J. Biomol. Struct. Dyn., 1 (1984) 1257.
[11] A. Pohorille, S.K. Burt and R.D. MacElory, J. Am. Chem. Soc., 106 (1984) 402.
[12] P. Cieplak and P.A. Kollman, J. Am. Chem. Soc., 110 (1988) 3734.
[13] B.F. Barker, M.F. Fox, E. Hyon and E.W. Loss, Anal. Chem., 46 (1979) 1785.
[14] B.F. Barker and M.F. Fox, Chem. Soc. Rev., 9 (1980) 143.
[15] M.J.S. Dewar, and W.J. Thiel, J. Am. Chem. Soc., 99 (1977) 4899.
[16] H. Moustafa, S. El-Taher, M. Hamed and R.H. Hilal, Int. J. Quantum Chem., 49 (1994) 805.
[17] T. Choudhury and S. Scheiner, J. Mol. Struct. Theochem., 18 (1984) 373; R.O. Angus Jr. M.W. Schmidt and R.P. Johnson, J. Am. Chem. Soc., 107 (1985) 532.
[18] S. Sleltzer, J. Am. Chem. Soc., 109 (1987) 1627; G.J.M. Dormaus, G.C. Groenenboom, W.C.A. Van Dorst and H.M. Buck, J. Am. Chem. Soc., 110 (1988) 1406.
[19] M.J.S. Dewar and J.J.P. Stewart, Chem. Phys. Lett., 111 (1984) 416.
[20] R.H. Hilal, M. Hamed and H. Mustafa, Biophys. Chem., 24 (1986) 171
[21] H. Kesster, Angew. Chem. Int. Ed., 9 (1970) 219; and references therein.
[22] W. Saenger, Principles of Nucleic Acid Structures, Springer-Verlag, New York, 1984.
[23] I.K. Yanson, A.B. Teplitsky and L.F. Sukhodub, Biopolymers, 18 (1979) 1149.
[24] T. Imoto, Biochim. Biophys. Acta, 475 (1977) 409.
[25] R.B. Hermann, J. Phys. Chem., 76 (1972) 2754.
[26] A. Pullman and B. Pullman, Quant. Rev. Biophys., 7 (1975) 505.
[27] F. Birnstock, H. Hofmann and H. Kohler, Theoret. Chim. Acta (Berlin), 42 (1976) 311.
[28] W. Voelter, R. Records, E. Bunnenberg and C.J. Dierassi, J. Am. Chem. Soc., 90 (1968) 6163.
[29] W. Hug and J. Tinoco, J. Am. Chem. Soc., 96 (1974) 665.
[30] H.H. Chen and L.B. Clark, J. Chem. Phys., 51 (1978) 1862.
[31] B.J. Cohen and L. Goodman, J. Am. Chem. Soc., 87 (1965) 5487.