Theoretical study of the protonation and deprotonation of cytosine. Implications for the interaction of cytosine with water

ARTICLE *in* JOURNAL OF MOLECULAR STRUCTURE · FEBRUARY 2000

Impact Factor: 1.6 · DOI: 10.1016/S0022-2860(99)00276-8

3 AUTHORS:

Asit K. Chandra
North Eastern Hill University
134 PUBLICATIONS **2,487** CITATIONS
SEE PROFILE

Minh Tho Nguyen
University of Leuven
748 PUBLICATIONS **10,835** CITATIONS
SEE PROFILE

Thérèse Zeegers-Huyskens
University of Leuven
208 PUBLICATIONS **2,974** CITATIONS
SEE PROFILE

ELSEVIER

Journal of Molecular Structure 519 (2000) 1–11

Journal of
MOLECULAR
STRUCTURE

www.elsevier.nl/locate/molstruc

Theoretical study of the protonation and deprotonation of cytosine. Implications for the interaction of cytosine with water

A.K. Chandra, M.T. Nguyen, Th. Zeegers-Huyskens*

Department of Chemistry, University of Leuven, 200F, Celestijnenlaan, B-3001 Heverlee, Belgium

Received 22 March 1999; accepted 26 April 1999

Abstract

The geometries, harmonic vibrational frequencies and energies of the two stable cyclic structures of the keto tautomer of cytosine complexed with water are computed using density functional theory (B3LYP) combined with the 6-31++G(d,p) basis set. The effect of complex formation with water on the pyramidalization of the amino group is discussed. The proton affinities of the oxygen and nitrogen atoms and the deprotonation enthalpies of the three NH bonds of cytosine are computed at the same level of theory. The deprotonation enthalpies of the two NH bonds of the amino group differ by 23 kJ mol^{-1} from each other and this reflects the asymmetric deformation of the amino group. The most stable hydrogen bond between cytosine and water is formed at the acceptor atom characterized by the lowest proton affinity and at the NH group having the highest acidity. The results are compared with data obtained at the same level of theory for the uracil– and thymine–water complexes. For the three nucleobases, the intermolecular distances and the energies of the hydrogen bonds formed at the different sites depend on the proton affinity and the deprotonation enthalpy of these sites. The dominance of the proton donor capacity in determining the hydrogen-bond energies and the cooperativity in the cyclic structures are discussed.

Keywords: Density functional theory; Cytosine–water complex; Hydrogen-bond interactions

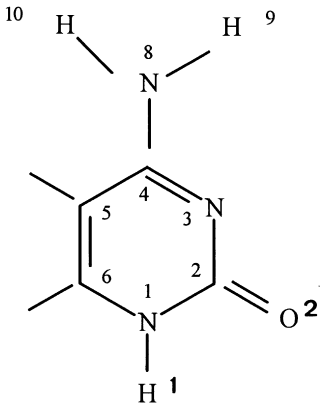
1. Introduction

The complex network of hydrogen-bond interactions that modulate recognition of DNA bases is based on the assumption of specific tautomeric and ionic states for the nucleic acid bases [1]. Minor tautomeric forms, i.e. the imino form of cytosine may be important in the formation of base mispairs. Alternatively, ionized and wobble base pairing have also been suggested to play a major role in mispair formation. The influence of acid–base equilibria on the stabilization of DNA structure has also received scant attention and recent data suggest that ionization may be relevant in determining mutagenic properties of analogues of nucleic acid bases [2]. The energetic provisions for Löwdin's DNA mutational mechanism [3] were investigated recently [4] for the cytosine–guanine Watson–Crick base pair. The basis of this mechanism that does not require the presence of rare tautomers to be formed, involves a concerted transfer of two protons in the interbase hydrogen bonds. Barrier height for the proton transfer reaction depends on the intrinsic acidity and basicity of the centres involved in the proton transfer reaction [5]. It is thus of primary importance to know the proton

* Corresponding author. Tel.: + 32-16-327477; fax: + 32-16-327991.

E-mail address: therese.zeegers@chem.kuleuven.ac.be (Th. Zeegers-Huyskens).

0022-2860/00/$ - see front matter
PII: S0022-2860(99)00276-8



Scheme 1. Atom labelling in the amino–oxo form of cytosine.

acceptor and proton donor abilities of these centres. Nucleobases are characterized by the presence of multiple proton donor and proton acceptor sites and experimental data on proton affinities (PA) [6,7] or deprotonation enthalpies [8] are related to the most stable protonated or deprotonated forms of the examinated species. For this reason, theoretical calculations can complement the experimental data.

The intrinsic acidity emerges as a common property in the protein-α helix and enzymes; this property may be in part responsible for the natural selection of these biological molecules [9,10] and this is likely to be the case for nucleo-bases. In our recent works, the proton affinities of the oxygen atoms and the deprotonation enthalpies of the NH bonds of uracil [11] and thymine [12] have been calculated by the B3LYP density functional theory using the 6-31++G(d,p) basis set. In the present work, we expand the scope of our analysis in considering the same properties for cytosine (Scheme 1) and their implications for the structure and energetics of the two stable cytosine–water complexes.

Several papers have discussed the tautomerism of the cytosine molecule [13–20] and the review of these studies goes behind the scope of this work. At 15 K in argon and nitrogen matrices, both amino–oxo and amino–hydroxy tautomers coexist [21] implying an almost zero free energy difference. In condensed phase, experimental evidence is in favor of a predominance of the amino–oxo form [22] which is of the same form as in the base-pairing scheme in DNA [1].

The optimized geometries and energies of the different cytosine–water complexes have been previously reported from rather low level calculations (SCF calculations combined with the STO-3G basis set) [23]. More recently, the relative energies of the cytosine–water conformers (water interacting at the (N1,C2=O) side of cytosine) have have been computed at the MP2/6-31G(d) level and the results have shown that the cytosine(keto)– water complex is by 4.5 kJ mol^{-1} more stable than the cytosine(enol)–water complex [17]. It has also been recently shown that the cytosine–water complex is stable in the amino–oxo form against the proton transfer reaction that leads to the imino–oxo tautomeric form on the ground as well on the excited potential energy surface [24].

The vibrational spectra of cytosine in gas phase [25], in rare gas matrices [21,26–28], in aqueous solution [20,29] and in solid state at room temperature [30,31] have been extensively studied. Vibrational data have also been recently obtained from ineslastic neutron scattering [20]. The normal modes of free cytosine have been calculated using the density functional theory [13,15] or MP2 calculations using polarization functions [20]. To the best of our knowledge, no theoretical studies have addressed the assignment of the vibrational spectrum of cytosine complexed with *one* water molecule and no experimental data on this complex have been reported. This may be due to the existence, in low temperature materials, of two cytosine tautomers which can both interact with water and this makes the assignment of the vibrational modes rather speculative [32]. Owing to strong solvatation on all active sites of cytosine, the results obtained in aqueous solutions [20,29] do not reflect the *selectivity* of the hydrogen-bond interaction.

The main scope of this work is to discuss the optimized geometries, the characteristic vibrational modes and the energies of the two stable closed cytosine–water complexes as a function of the intrinsic basicities and acidities of the sites involved in the hydrogen-bond interaction and to generalize the correlations found previously for the uracil and thymine complexes. All the results are obtained using B3LYP functional combined with the 6-31++G(d,p) basis set. It must be stressed here that the scope of this work is not to compare the vibrational parameters of isolated cytosine obtained at different theoretical levels nor to discuss the optimal scaling factors. High level

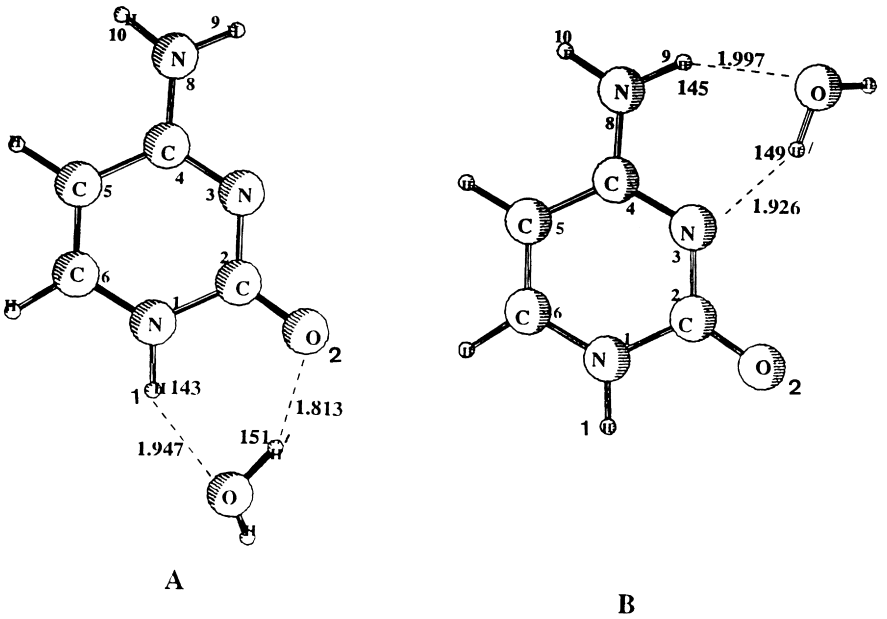


Fig. 1. Optimized structures for the complexes A and B between cytosine and water obtained from B3LYP/6-31++G(d,p) calculations (distances in Å, angles in degree).

density functional theory methods have significantly narrowed the gap between the computed and the experimental frequencies and this has been extensively discussed in earlier works [10,33–37].

2. Computational methods

The geometries of the isolated cytosine molecule and the cytosine–water complex were optimized by the density functional theory using B3LYP [38,39] exchange correlation functionals and 6-31++G(d,p) basis functions. The proton affinities and deprotonation enthalpies were computed by the same procedure, using the same 6-31++G(d,p) basis function. The GAUSSIAN 94 package [40] was used for all the calculations. Harmonic vibrational frequencies were calculated at the same level of theory to characterize the stationary points and to evaluate the frequency shifts due to complex formation with water. Anharmonic contributions whose evaluation requires the calculation of successive energy derivatives could not be evaluated for these relatively big systems.

3. Results and discussion

3.1. Geometry of the cytosine–water complexes

The B3LYP/6-31++G(d,p) optimized geometry of the two cyclic cytosine–water complexes A and B are shown in Fig. 1. Table 1 indicates characteristic geometrical parameters computed for free cytosine and cytosine–water complexes. The C5H and C6H distances and the angles of the ring are very insensitive to complex formation and are not listed in Table 1.[1] As discussed in recent works [13,41–43], amino group nonplanarity and deformation are essential for the explanation of many base–base interaction in DNA. The potential barrier separating the planar from the lowest pyramidal structure is however low and ranges between 0.3 and 1.3 kJ mol^{-1} [42]. We found for the planar structure of the amino form of free cytosine one imaginary vibration corresponding thus to a transition structure for inversion. The values of the sum of the angles around the N atom of the amino group ($\sum$AH) and the values of the dihedral

[1] All the optimized geometries and vibrational frequencies in the free molecule and in the complexes are available from the authors.

Table 1
Results of B3LYP/6-31++G(d,p) geometry optimizations of free cytosine and the cytosine–water complexes A and B (bond lengths in Å, bond angles in degree)

	Free cytosine	Complex A	Complex B
N1–C2	1.427	1.415	1.422
C2–N3	1.370	1.363	1.370
N3–C4	1.321	1.325	1.331
C4–C5	1.442	1.438	1.442
C5–C6	1.361	1.363	1.356
C2=O	1.224	1.238	1.225
C4–N8	1.359	1.359	1.351
N1–H1	1.011	1.023	1.011
N8–H9	1.009	1.009	1.020
N8–H10	1.006	1.006	1.006
∠C4H8N9	116.5	118.2	118.8
∠C4N8H10	120.2	122.1	120.9
∠H9N8H10	119	119.6	120.0
∠H9N8C4N3	10.4	3.3	3.2
∠H10N8C4C5	−15.8	−5.0	−4.0
∑AH	355.7	359.9	359.7
δ	4.3	0.1	0.3
Δ	−5.4	−1.7	−0.8
	Free water	Complex A	Complex B
OH	0.965	0.964	0.964
OH′	0.965	0.987	0.985
∠HOH′	105.7	107.3	107.0
Intermolecular parameters			
Complex A		Complex B	
H1···Ow	1.947	H9···Ow	1.997
O2···H′	1.813	N3···H′	1.926
∠OwH′O2	151.0	∠N8H9Ow	145.0
∠N1H1Ow	143.0	∠N3H′Ow	149.0

angles H9N8C4N3 and H10N8C4C5 obtained in this work are 355.7, 10.4 and −15.8° respectively. These values are intermediate between the ones obtained from MP2/6-311G(2df,p) calculations (351.9, −21°, 13°) [41] and HF/6-31G(d,p) calculations (359.5, −5, 3°) which underestimate the pyramidalization of the amino group [43]. DFT computations without diffuse functions (6-31G(d,p) basis set) gave smaller dihedral angles of −10 and 5.5° respectively [13].

The two stable cytosine–water complexes are the ones in which water accepts one of the acidic NH protons while donating a proton to the O2 or N3 atoms. As shown by recent calculations carried out at the MP2/6-31G(d) level [17], the structure where water acts as a bidonor towards the O2 and N3 atoms is by 22.4 kJ mol^{-1} less stable than the complex A and the intermolecular distances markedly longer (H···N3 = 2.149 Å, H···O2 = 2.349 Å) than those computed for both complexes A and B. The calculations of the present work do not predict this structure to be a local minimum. This anticooperative structure [44,45] will no longer be discussed hereafter. The MP2/6-31G(d) computed H1···O distance of 1.774 Å in complex A is also markedly lower than the one calculated in this work 1.947 Å.

Not surprisingly, the bonds which are the most sensitive to interaction with water are those involved in the formation of the pseudo-ring structure, i.e. the N1H, N1C2 and C2=O2 bonds for structure A and the N3C4, C4N8 and N8H9 bonds for structure B. Complex formation results in moderate changes of the nonplanar geometry of the amino group. As

Table 2
Unscaled characteristic B3LYP/6-31++G(d,p) vibrational frequencies (cm^{-1}) in free cytosine and water and the cytosine–water complexes A and B

Assignment[a]	Free cytosine[b]	Complex A	$\Delta\nu$	Complex B	$\Delta\nu$
ν^{as}NH2	3744(49)	3756(55)	+12	3719(90)	−25
νN1H	3629(75)	3405(68)	−224	3632(84)	+3
ν^{s}NH2	3601(89)	3614(84)	+13	3447(2)	−154
νC2=O+δN1H1	1774(781)	1746(698)	−28	1774(750)	0
νC5=C6+δC6H+δNH2	1692(478)	1692(465)	0	1698(590)	+6
δNH2	1634(145)	1641(64)	+7	1663(18)	+29
νR+δN1H	1572(167)	1579(110)	+7	1565(178)	−7
δN1H+νN1C2+νN3C4	1444(82)	1477(76)	+33	1450(91)	+6
νC4N8+δC5H+δC6H	1358(52)	1370(74)	+12	1374(32)	+16
νC2N3+δC6H	1263(25)	1288(20)	+25	1276(39)	+13
νR+δC5H+δN1H	1219(51)	1254(45)	+35	1225(53)	+6
νR+δC5H+δN1H	1129(3)	1140(2)	+11	1131(3)	+2
rNH2	1096(47)	1099(25)	+3	1120(43)	+24
νN1C2+νC4C5+rNH2	927(3)	938(3)	+11	950(6)	+23
γC2=O+γR	767(24)	777(61)	+11	773(35)	+6
γC5H+γR	724(28)	715(5)	−9	758(4)	+34
γN1H	624(69)	834(158)	+210	611(32)	−13
δC2=O+δR	577(2)	588(2)	+11	584(2)	+11
δR+tNH2	546(2)	557(10)	+11	552(4)	+6
tNH2+δC2 = O	523(17)	537(14)	+14	676(105)	+153
γR	397(17)	416(10)	+19	406(55)	+9
γR+ωNH2	359(3)	373(26)	+14	387(45)	+28
ωNH2	266(246)	210(7)	–[c]	309(129)	+43
γR+ωNH2	250(15)	147(5)	–[c]	193(11)	–[c]
γR+ωNH2	133(3)	69(210)	–[c]	142(2)	–[c]
Water modes					
	Free water	Complex A	Complex B		
ν3	3927(54)	3890(79)	3899(79)		
ν1	3804(5)	3468(1105)	3498(999)		
ν2	1601(87)	1625(462)	1627(62)		
ωHOH′		747(235)	717(119)		

[a] ν = stretching, δ = in-plane deformation, γ = out-of-plane deformation, r = rocking, t = torsion, ω = wagging, R = ring vibration.
[b] The intensities in km mol^{-1} are indicated in parentheses.
[c] See discussion in the text.

discussed recently [46], there are two sources on nonplanarity for the DNA bases: (1) a partial sp^3 hybridization of the amino group, and (2) the interaction of the amino group hydrogen atoms with the closest atom(s) belonging to the base (in the case of cytosine, the H5 atom). The simplest way to analyse the planarity of the first type is to consider the deviation from 360° of the sum of the angles around the nitrogen atom of the amino group (δ in Table 1). The second type of planarity can be estimated as the sum of the dihedral angles H10N8C4C5 and H9N8C4N3, provided that the only the amino group hydrogen atoms are nonplanar (Δ in Table 1). An unconstrained full geometry optimization yields a Δ value about 2° higher [41]. This small difference will probably not influence the relative changes of the angles resulting from the interaction with water. The results of Table 1 indicate that both types of nonplanarity are equally operating in the complexes. The differences between the δ and Δ values in free cytosine and in the two complexes are about 4°. In the two complexes, the N atom of the amino group is nearly sp^2 hybridized and this parallels a shorter C4N8 distance in complex A. It is also worth noticing that amino groups which adopt a pyramid like structure in isolated guanine and cytosine become nearly planar in the base pair [47].

The elongation of the OH′ bond of water involved

in hydrogen-bond formation is somewhat larger for complex A (0.021 Å) than for complex B (0.019 Å). For both structures, a small decrease of the distance of the free OH bond is predicted. This decrease parallels a small increase of the force constant of the free OH bond [48].

3.2. Vibrational spectrum of the cytosine–water complexes

Table 2 contains the vibrational frequencies in free cytosine and the two cyclic structures of the cytosine–water complexes. Vibrations weakly sensitive to complex formation (νC5H, νC6H, the in-plane ring deformation at 989 cm^{-1}, the out-of-plane C6H deformation at 961 cm^{-1} and the coupled out-of-plane ring and C5H deformation at 763 cm^{-1}) are not mentioned in this table.[1] The experimental and calculated vibrations of free cytosine have been compared in recent works [13,15,20]. For calculations performed at the HF/6-311G(d,p) level, the scale factor is about 0.90 [29]. When using the B3LYP density functional and the 6-31G(d,p) basis set 15, the scale factor is about 0.97. As shown by the present calculations, addition of two diffuse functions results in a slight increase in the frequency of the νNH2 vibrations. The agreement between the experimental frequency of the νC=O vibration (1730 cm^{-1}) is better when the calculations are carried out at the 6-31++G(d,p) level (1774 cm^{-1}) than at the 6-31G(d,p) one (1818 cm^{-1}). In complex A where both the N1H and C=O bonds of cytosine are involved in complex formation, a strong decrease of the νN1H stretching frequency ($\Delta\nu = -224\ cm^{-1}$), a strong frequency increase of the N1H out-of-plane deformation ($\Delta\nu = +210\ cm^{-1}$) and a moderate decrease of the νC=O stretching frequency are predicted. It must be mentioned here that the νN1H mode is coupled with the νH2O vibrations and the νC=O and δNH2 modes are coupled with the δH2O vibration. The main component of the δNH1 vibration computed at 1444 cm^{-1} in the free cytosine molecule is shifted upward by 33 cm^{-1}. The other modes involving a δN1H vibration at 1572, 1219 and 1129 cm^{-1} are also red shifted by 7, 35 and 11 cm^{-1} respectively. The out-of-plane ring vibrations computed at 198 and 133 cm^{-1} and the wagging vibration of the NH2 group are strongly coupled with the low frequency water vibrations and will no longer be discussed in this work. It is interesting to mention that the stretching vibrations of the NH2 group which is not involved in complex formation are shifted to higher frequencies, the ν^{as}NH2 being shifted by 12 cm^{-1} and the ν^{s}NH2 vibration by 13 cm^{-1}. This effect results from the decrease of the pyramidal character of the N8 atom. A similar effect has been observed in other complexes such as those between 3-aminoquinoline and phenols [49]. Complex formation takes place on the nitrogen atom of the heteroaromatic ring and results in an increase of the νNH2 frequencies and the NH2 angle. In complex B where the N8H9 group of cytosine is hydrogen bonded to the oxygen atom of water, the νNH2 vibrations which are almost 100% pure, are shifted downwards by 25 and 154 cm^{-1}. The perturbation of the ν^{as}NH2 vibration is smaller than that of the ν^{s}NH2 vibration because, as shown by the calculations of the amplitudes, the ν^{as}NH2 vibration in the complexes can be approximately considered as the free one and the ν^{s}NH2 vibration as the bonded one. This has been discussed thoroughly for aniline derivatives complexed with proton acceptors [50,51].

The ν^{s}NH2 vibration predicted at 3447 cm^{-1} is strongly mixed with the νH2O vibrations. This is also the case of the δNH2 mode predicted at 1663 cm^{-1} which is shifted upward by 29 cm^{-1} from free cytosine and which is strongly mixed with the δH_2O vibration. Other modes involving a vibration of the NH2 group are also shifted to higher frequencies: the rocking vibration, the torsion vibration and the wagging vibration. The mixed γC5H + γR mode computed at 724 cm^{-1} in free cytosine is mixed with the tNH2 mode in the B complex. In free cytosine, the inversion mode at the NH2 group is predicted at 266 cm^{-1} and is also mixed with the γR modes at 250 and 133 cm^{-1}. In complex B where one of the hydrogen of the NH2 group is involved in complex formation, the inversion mode is blue-shifted by 43 cm^{-1}. In complex A, no clear inversion mode could emerge from our calculations, this mode being coupled with the water vibrations and with the intermolecular modes. The mode predicted at 69 cm^{-1} and characterized by an high infrared intensity involves predominantly this inversion mode. It is interesting to note here that in free cytosine, HF/6-31G(d) calculations predict the predominant contribution of the NH2 inversion mode at

Table 3
Total energies (E_T) (in Hartree) and zero-point vibrational (ZPE) (in kJ mol^{-1}) for cytosine and water monomers and the cytosine–water complexes, BSSE values and relative energies (in kJ mol^{-1}) of complexes A and B between cytosine and water obtained from B3LYP/6-31++G(d,p) calculations

	E_T	ZPE	BSSE	Relative energies
Cytosine monomer	− 394.963309	257.9	–	
Water monomer	− 76.434124	55.8	–	
Complex A	− 471.416968	324.1	3.6	− 37.2
Complex B	− 471.415780	323.5	3.5	− 34.8

218 cm^{-1}; the mode at 143 cm^{-1} contains also some contribution to this mode and is coupled with ring torsional vibrations. In cytosine embedded in a cavity of polarizable continuum of dielectric constant of 40, the first mode is red-shifted by 43 cm^{-1}, the second mode remaining almost unaltered [47].

The computed infrared intensities represent only a quantitative trend. The experimental intensity of a NH group involved in hydrogen-bond formation with a proton acceptor generally increases. The decrease of the intensity of the νN1H vibration in complex A and of the ν^{s}NH2 vibration in complex B can be qualitatively accounted for by a strong coupling of these modes with the ν1 mode of water. This clearly appears when comparing the intensity of the ν1 vibration in free water (5 km mol^{-1}) and in the A (1105 km mol^{-1}) and B (999 km mol^{-1}) complexes. The same remark also holds for the νC=O vibration in complex A. Hydrogen-bond formation between a C=O group and a proton donor generally results in a small increase of the intensity of the νC=O vibration. The intensity decrease from 780 km mol^{-1} in free cytosine to 698 km mol^{-1} in complex A can be explained by a strong mixing between the νC=O vibration and the δH2O mode in the complex. This coupling also results in a very strong intensity enhancement (from 5 to 1105 km mol^{-1}). The same effect can be observed in complex B where the δNH2 and δH2O vibrations are strongly mixed.

The infrared spectrum of polycrystalline cytosine monohydrate [16] recently investigated between 1800 and 400 cm^{-1} cannot be compared with the present data. Indeed, in this hydrate, two cytosine molecules form nonsymmetric closed dimers. The lengths of N1H···N3 and N8H9···O hydrogen bonds amount to 2.95 and 2.99 Å and the N8H···water hydrogen bond 2.97 Å [52]. The νC=O vibration of the bonded C=O group in the dimer is observed at 1659 cm^{-1} and a broad band assigned as mixed wagging and torsion of the amino group is observed at 664 cm^{-1}. In the experimental FT-Raman spectra of cytosine dissolved in D2O [20], the νC=O vibration observed at 1645 cm^{-1}. This low frequency results from the strong solvatation of cytosine and the high dielectric constant of the medium. A blue shift of the νC=O vibration of 55 cm^{-1} is predicted when cytosine is embedded in a cavity formed in polar continuum of dielectric constant of 40 [47].

3.3. Relative energy of the cytosine–water complexes and proton affinity and deprotonation enthalpy of the different sites of cytosine

Table 3 lists the total energies (E_T), zero-point vibrational energies (ZPE), the basis set superposition errors (BSSE) values calculated by the counterpoise method [53] and relative energies of both complexes A and B. Complex A is calculated to be about 2.5 kJ mol^{-1} more stable than complex B. The intermolecular distances and relative energies will now be discussed as a function of the proton donor and proton acceptor abilities of the different sites of cytosine. It is useful to remember here that the proton affinity, PA(B), is defined as the negative enthalpy change associated with the gas phase protonation reaction $B + H^+ \leftrightarrow BH^+$ and the deprotonation enthalpy, PA(A^-), as the enthalpy change associated with the gas phase deprotonation reaction $AH \leftrightarrow A^- + H^+$. The PA(B) values of the O and N3 atoms and the PA(A^-) values of the N1H, NN8H9 and N8H10 bonds of cytosine are listed in Table 4. These data indicate that the PA(B) values of the O (N3 side) and N3 atoms are very similar and in relatively good agreement with the experimental value [7] of

Table 4
B3LYP/6-31++G(d,p) proton affinities (PA(B)) and deprotonation enthalpies (PA(A^-)) of cytosine(kJ mol^{-1}) ($T = 298$ K)

PA(B)		PA(A^-)	
O (N1 side)	922.1	N1H	1444.2
O (N3 side)	957.2	N8H9	1481.4
N3	955.7	N8H10	1457.9

944.2 kJ mol^{-1} which probably corrresponds to the protonation of the most basic site. Previous calculations carried out at the HF/4-31G level with STO-3G optimized geometries overestimate the PA(B) values, especially for the N3 atom (1042 kJ mol^{-1}) [54,55].

The present results are in good agreement with recent calculations showing that O2-protonated cytosine should slightly prevail in the gas phase, whereas in aqueous solutions, the N3-protonated form of cytosine is 12.5 kJ mol^{-1} more stable than the enol tautomer [16]. Also recent MP4/6-311++G(d,p) calculations carried out on the protonation of the amino–enol and amino–keto forms of cytosine have shown that three atoms in neutral cytosine (N1,N3 and O) are susceptible to be protonated all within a range of 4.2 kJ mol^{-1}, the PA(B) value being 948.9 kJ mol^{-1} [19]. No accurate determination of the deprotonation enthalpies are available in the literature; the experimental results of Fourier-transform ion cyclotron resonance mass spectroscopy give only the relative sequence of acidities of the nucleobases [8]. Our calculations of the deprotonation enthalpies can be verified by the fact that we obtained for 2-aminopyridine, a PA(A^-) value of 1517.7 kJ mol^{-1} for the NH bond vicinal to the CH bond, which differs by only 1.7 kJ mol^{-1} from the experimental value of 1516 kJ mol^{-1} [56]. The deprotonation enthalpy of the N1H bond of cytosine is very similar to that of the N3H bond of uracil recently calculated at the same level [11,12] and to the experimental value of diacetamide (1449.6 kJ mol^{-1}) [57–59]. The deprotonation enthalpies of the two N8H bonds differ by more than 20 kJ mol^{-1}, the N8H10 bond having a higher acidity than the N8H9 one. This effect results from the repulsion between the N8H10 bond and the adjacent hydrogen atom and illustrates very nicely the non-equivalence of the two NH bonds of the amino group.

Comparison of the results reported in Tables 3 and 4 reveals that *the most stable complex is formed at the site characterized by the lowest PA(B) value*, namely the lone pair of the O atom at the N1 side and at the NH site *having the highest acidity*, namely the N1H bond.

Correlations between complexation energies and proton affinities are well documented in the literature [60–64] and for proton-held dimer cations $(A\cdots H\cdots B)^+$, linear correlations between the hydrogen-bond energies and the differences in proton affinities (ΔPA(B)) of the monomers forming the bond have been observed [63], in agreement with theoretical predictions [65]. Except for uracil and thymine complexed with water [12], these correlations have not been investigated for closed structures where one proton donor site and a proton acceptor site of the "host" molecule are both involved in complex formation.

Table 5 allows one to compare the $r(H'\cdots O)$ distances between the oxygen carbonyl and the H' atom of water, the hydrogen-bond energies (E_{HB}) and the PA(B) and PA(A^-) values of the corresponding sites for the uracil [11], thymine [12] and cytosine complexes. These parameters refer to structures where water is hydrogen bonded to the carbonyl oxygen and the intracyclic NH group. It is important to note here that these structures are planar and that the (Ow)H···O and (N)H···Ow angles are all between 143 and 151°.

These data show that the shortest $r(H'\cdots O)$

Table 5
Intermolecular distances $r(H'\cdots O)$ (in Å) and relative energies (in kJ mol^{-1}) for uracil, thymine and cytosine complexed with water. PA(B) and PA(A^-) values of the corresponding sites

Complex	$r(H'\cdots O)$	E_{HB}	PA(B)	PA(A^-)
Uracil (A)[a]	1.941	−32.8	815.1	1391.0
Uracil (B)[b]	1.975	−24.5	820.1	1447.1
Uracil (C)[c]	1.921	−26.7	849	1447.1
Thymine (A)[a]	1.929	−32.1	830.1	1397.8
Thymine (B)[b]	1.947	−24.7	835.2	1449.2
Thymine (C)[c]	1.914	−26.7.0	854.8	1449.2
Cytosine (A)[d]	1.813	−37.2	922.1	1442.2

[a] Cyclic complex formed between the N1H bond and the O2 atom (Ref. [12]).
[b] Cyclic complex formed between the N3H bond and the O2 atom (Ref. [12]).
[c] Cyclic complex bond formed between the N3H bond and the O4 atom (Ref. [12]).
[d] This work.

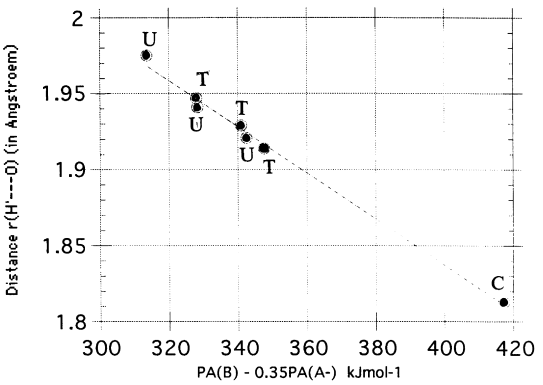


Fig. 2. $r(H'\cdots O)$ (in Å) as a function of $[PA(B) - 0.35PA(A^-)]$ (in kJ mol^{-1}) for closed complexes involving water and uracil (U), thymine (T) and cytosine (C).

intermolecular distance is predicted for the cytosine–water complex A where the hydrogen bond is formed on the oxygen lone pair having the highest PA(B) value. Comparison of the uracil–, thymine– and cytosine–water complexes shows, however, that there is no correlation between the intermolecular distances and the PA(B) values of the corresponding oxygen atoms. This can be accounted for by the fact that the NH$\cdots$O bond formed in the closed complex will reinforce the OH$'\cdots$O bond by cooperative effect. The overall effect can be expressed by the following exponential expression illustrated in Fig. 2:

$$r(H\cdots O) = 2.542e^{-0.00812[PA(B)-0.35PA(A^-)]} \quad (1)$$

$$(r = 0.9962)$$

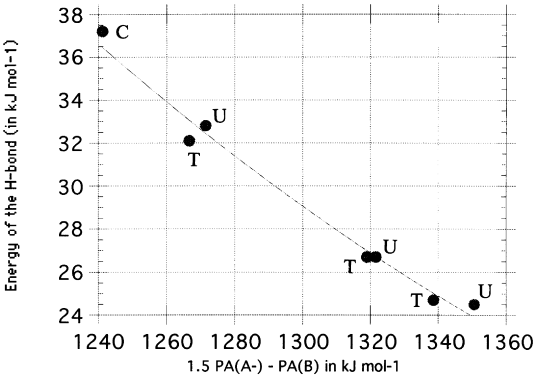


Fig. 3. E_{HB} as a function of $[1.5PA(A^-) - PA(B)]$ (both in kJ mol^{-1}) for closed complexes involving water and uracil (U), thymine (T) and cytosine (C).

The cooperativity which, in the present case, can be estimated from the ratio of the coefficients of $PA(A^-)$ and PA(B) in Eq. (1) is 0.35.

The data of Table 5 also indicate that the hydrogen-bond energies (E_{HB}) in the cyclic dimers are not correlated to the separate values of PA(B) or $PA(A^-)$. The energies are expressed by the following exponential expression:

$$E_{HB} = 4401e^{-0.00386[1.5PA(A^-)-PA(B)]} \quad (r = 0.9926) \quad (2)$$

where both the complexation energies and PAs are expressed in kJ mol^{-1}. The correlation is illustrated in Fig. 3. In a smaller E_{HB} − PA range, Eq. (2) can be considered as linear [12]. Somewhat similar results have been obtained for complexes involving diacetamide and guest molecules such as water, methanol or ammonia [66]. In this case also, the proton acceptor ability of the "guest" molecule plays a more important role than its proton acceptor ability. In this respect, it is interesting to note here that the energies of the complexes B involving water and cytosine (−34.8 kJ mol^{-1}) and 1-CH_3 cytosine (−34.5 kJ mol^{-1} from MP2/6-31G++G(d,p) calculations [67] or −34.4 kJ mol^{-1} from DFT/6-31+G(d,p) calculations) [68] are very similar. The PA of the N3 atom of 1-CH_3 cytosine is 973 kJ mol^{-1}, thus 16 kJ mol^{-1} higher than in cytosine but the deprotonation enthalpy of the N8H9 bond of 1-CH_3 cytosine is 1485.6 kJ mol^{-1} [69], about 4 kJ mol^{-1} higher than for the corresponding bond in cytosine. For structures B, the acidity of the NH2 group seems also to be a key factor.

Eq. (2) illustrates the dominance of the proton donor in determining the hydrogen-bond energies. This is in agreement with the theoretical considerations of Desmeules and Allen [65] who have shown that the charge rearrangement occurring upon hydrogen-bond formation reflects the dominance of the proton donor in determining the properties on the hydrogen bond. It is interesting to note here that in all the theoretical or experimental correlations between the hydrogen-bond energies and ΔPA, the coefficients of the PA of the two partners are taken as one. It can be anticipated that in a homologous series of hydrogen-bond complexes, *the energies at* $\Delta PA = 0$ should be the same. This obviously is not the

case. For symmetric proton-held dimer cations $H^+(B)_2$ as for example, the bond energy decreases as the PA of base B increases. This is nicely illustrated for $(O\cdots H\cdots O)^+$ bonds: when the proton is bonded to two water molecules, the binding energy is 146 kJ mol^{-1} and the energy drops to 124 kJ mol^{-1} when the two water molecules are replaced by two diethylether molecules. The electron density on the proton reflects the energetic trend [69]. The predominance of the proton donor in determining the hydrogen-bond parameters is not a specific feature of closed complexes discussed in this work but seems to be a more general characteristic of the hydrogen bonds [70].

3.4. Concluding remarks and perspectives

In this work, the deprotonation enthalpies of the NH bonds of cytosine are computed for the first time at a good level of theory. The structure, energy and vibrational properties of the two stable, closed cytosine–water complexes are also calculated. Comparison between the uracil and cytosine complexes show the greater importance of the proton donor in the determination of the hydrogen-bond energies. The hydrogen-bond interaction which determine the specificity of recognition between the nucleobases and the formation of rare base-pair tautomers resulting from the antiparallel simultaneous transfer of two protons are also expected to depend on the protonation and deprotonation enthalpies of the sites involved in the interaction. This will be discussed in a forthcoming publication dealing with the protonation and deprotonation enthalpies of guanine and adenine [68].

Acknowledgements

The authors thank the Fund for Scientific Research (FWO-Vlaanderen) for financial support and the Leuven Research Council for a postdoctoral fellowship to A.K.C.

References

[1] G.A. Jeffrey, W. Saenger, Hydrogen Bonding in Biological Structures, Springer, Berlin, 1991 and references cited therein.
[2] L.C. Sowers, M.F. Goodman, R. Eritja, B.E. Kaplan, G.V. Fazakerley, J. Mol. Biol. 205 (1989) 437.
[3] P.O. Löwdin, Rev. Mod. Phys. 35 (1963) 724.
[4] J. Florian, J. Leszczynski, J. Am. Chem. Soc. 118 (1996) 3010.
[5] Th. Zeegers-Huyskens, P. Huyskens, in: H. Ratajczak, J.W. Orville-Thomas (Eds.), Molecular Interactions, 2, Wiley, New York, 1991, pp. 1–106.
[6] M. Meotner, J. Am. Chem. Soc. 101 (1979) 2399.
[7] F. Greco, A. Liguori, G. Sindona, N. Ucella, J. Am. Chem. Soc. 112 (1990) 9092.
[8] M.T. Rodgers, S. Campbell, E.M. Marzluff, J.L. Beauchamp, Int. J. Mass Spectrosc. Ion Proc. 137 (1994) 121.
[9] M. Meotner, J. Am. Chem. Soc. 110 (1988) 3071.
[10] M. Meotner, J. Am. Chem. Soc. 110 (1988) 3075.
[11] M.T. Nguyen, A.K. Chandra, Th. Zeegers-Huyskens, J. Chem. Soc. Faraday Trans. 94 (1998) 1277.
[12] A.K. Chandra, M.T. Nguyen, Th. Zeegers-Huyskens, J. Phys. Chem. 102 (1998) 6010.
[13] D.A. Estrin, L. Paglieri, G. Corongiu, J. Phys. Chem. 98 (1994) 5653.
[14] I.R. Gould, N.A. Burton, R.J. Hall, I.H. Hillier, J. Mol. Struct. 331 (1995) 147.
[15] J.S. Kwiatkowski, J. Leszczynski, J. Phys. Chem. 100 (1996) 941.
[16] J. Florian, V. Baumruk, J. Leszczynski, J. Phys. Chem. 100 (1996) 5578.
[17] A. Broo, A. Holmen, J. Phys. Chem. A 101 (1997) 3589.
[18] G. Fogarasi, J. Mol. Struct. 413 (1997) 271.
[19] C. Colominas, F.J. Luque, M. Orozco, J. Am. Chem. Soc. 118 (1996) 6811.
[20] A. Aamouche, M. Ghomi, L. Grajcar, M.H. Baron, F. Romain, V. Baumruk, J. Stepanek, C. Coulombeau, H. Jobic, G. Berthier, J. Phys. Chem. 101 (1997) 10 063.
[21] M. Szczesniak, K. Szczepaniak, J.S. Kwiatowski, K. KuBulat, W.B. Person, J. Am. Chem. Soc. 110 (1988) 8319.
[22] A.R. Katritzky, M. Karelson, J. Am. Chem. Soc. 113 (1991) 1561.
[23] J.E. Del, Bene, J. Comp. Chem. 4 (1983) 226.
[24] A.L. Sobolewski, L. Adamowicz, J. Chem. Phys. 102 (1995) 5708.
[25] D. Ferro, L. Bencivenni, R. Teghil, R. Mastromarino, Thermochim. Acta 42 (1980) 75.
[26] E.D. Radchenko, G.G. Sheina, N.A. Smorygo, Y.P. Blagoi, J. Mol. Struct. 116 (1984) 387.
[27] M.J. Nowak, L. Lapinski, J. Fulara, Spectrochim. Acta A 45 (1989) 229.
[28] I.R. Gould, M.A. Vincent, I.H. Hillier, L. Lapinski, M.J. Nowak, Spectrochim. Acta A 48 (1992) 811.
[29] R.C. Lord, G.J. Thomas Jr, Spectrochim. Acta A 23 (1967) 2551.
[30] H. Susi, J. Ard, J.M. Purcell, Spectrochim. Acta A 29 (1967) 2551.
[31] H. Susi, J.S. Ard, Spectrochim. Acta A 30 (1974) 1843.
[32] K. Szczepaniak, personal communication.
[33] Z. Latajka, Y. Bouteiller, J. Chem. Phys. 101 (1994) 9763.
[34] G. Rauhut, P. Pulay, J. Phys. Chem. 99 (1995) 3093.
[35] Z. Latajka, Y. Bouteiller, S. Scheiner, Chem. Phys. Lett. 234 (1995) 159.

[36] V. Barone, C. Adamo, J. Phys. Chem. 100 (1996) 941.
[37] K. Szczepaniak, W.B. Person, J. Leszczynski, J.S. Kwiatkowski, Pol. J. Chem. 72 (1998) 402.
[38] C. Lee, W. Yank, R.G. Parr, Phys. Rev. 37B (1988) 785.
[39] A.D. Becke, J. Chem. Phys. 98 (1993) 5648.
[40] GAUSSIAN 94, Revision C3, M.F. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Peterson, J.A. Montgomery, K. Raghavachary, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Ciolowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Repogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzales, J.A. Pople, Gaussian Inc., Pittsburgh, PA, 1995.
[41] J. Sponer, P. Hobza, J. Phys. Chem. 98 (1994) 3161.
[42] O. Bludsky, J. Sponer, J. Leszczynski, V. Spirko, P. Hobza, J. Chem. Phys. 105 (1996) 11 042.
[43] J. Sponer, J. Florian, P. Hobza, J. Leszczynski, J. Biomol. Struct. Dyn. 13 (1996) 827.
[44] P.L. Huyskens, J. Am. Chem. Soc. 99 (1977) 2578.
[45] H. Kleeberg, in: P.L. Huyskens, W.A.P. Luck, Th. Zeegers-Huyskens (Eds.), An Introduction to Modern Methods and Results, Springer, Heidelberg, 1991, pp. 251.
[46] L. Gorb, J. Leszczynski, J. Am. Chem. Soc. 120 (1998) 5024.
[47] J. Florian, J. Leszczynski, Int. J. Quant. Chem. Quant. Biol. Symp. 22 (1995) 207.
[48] M.T. Nguyen, N. Leroux, Th. Zeegers-Huyskens, J. Chem. Soc. Faraday Trans. 93 (1997) 33.
[49] N. Leroux, M. Goethals, Th. Zeegers-Huyskens, Vibr. Spectrosc. 9 (1995) 235.
[50] J. Lauransan, P. Pineau, J. Lascombe, J. Chim. Phys. 63 (1966) 635.
[51] J. Lauransan, P. Pineau, J. Chim. Phys. 65 (1968) 1937.
[52] R.J. McClure, B.M. Craven, Acta Crystallogr. B 29 (1973) 1234.
[53] S.F. Boys, F. Bernardi, Mol. Phys. 19 (1970) 553.
[54] J.E. Del, Bene, J. Phys. Chem. 87 (1983) 367.
[55] V.A. Voityuk, A.A. Bliznuk, J. Mol. Struct. 205 (1990) 119.
[56] M. Mautner, J. Am. Chem. Soc. 106 (1984) 1257.
[57] T.B. Mc Mohan, P. Kebarle, J. Am. Chem. Soc. 98 (1976) 3399.
[58] J.B. Cumming, P. Kebarle, Can. J. Chem. 56 (1978) 1.
[59] M.S. Dewar, K.M. Dieter, J. Am. Chem. Soc. 108 (1986) 8075.
[60] Th. Zeegers-Huyskens, J. Mol. Struct. 177 (1988) 125.
[61] Th. Zeegers-Huyskens, P. Huyskens, in: P.L. Huyskens, W.A.P. Luck, Th. Zeegers-Huyskens (Eds.), Intermolecular Forces. An Introduction to Modern Methods and Results, Springer, Berlin, 1991, pp. 24.
[62] W.R. Davidson, J. Sunner, P. Kebarle, J. Am. Chem. Soc. 101 (1979) 1675.
[63] C.V. Speller, M. Meotner, J. Phys. Chem. 89 (1985) 5217.
[64] Th. Zeegers-Huyskens, Chem. Phys. Lett. 129 (1986) 172.
[65] P.J. Desmeules, L.C. Allen, J. Chem. Phys. 72 (1980) 4731.
[66] M.T. Nguyen, N. Leroux, Th. Zeegers-Huyskens, J. Mol. Struct. 404 (1997) 75.
[67] J. Smets, L. Adamowicz, G. Maes, J. Phys. Chem. 100 (1996) 6434.
[68] A.K. Chandra, M.T. Nguyen, Th. Zeegers-Huyskens, in preparation.
[69] K. Hiraoka, H. Takimoto, S. Yamabe, J. Phys. Chem. 90 (1986) 5910.
[70] Th. Zeegers-Huyskens, in preparation.