J Mol Model (2009) 15:711–722
DOI 10.1007/s00894-008-0423-8

ORIGINAL PAPER

Accurate micro- and macro- gas phase basicities of hydroxyl-radical-modified pyrimidines estimated by advanced quantum chemistry methods

Piotr Cysewski · Katarzyna Kozłowska · Beata Szefler

Received: 16 August 2008 / Accepted: 18 November 2008 / Published online: 17 December 2008

Abstract The gas phase basicities (GPB) of 16 hydroxyl-radical-derived analogues of model pyrimidine nucleosides were analysed at B3LYP/aug-cc-pVDZ and G3MP2B3 levels of theory. All possible tautomeric equilibriums of neutral and protonated forms of pyrimidine analogues were taken into account. The oxidation of pyrimidines usually reduces the GPB values with respect to canonical values. The only exception was observed in the case of 5,6-dihydroxycytidine, which is more basic than model cytidine if both macro- and micro-scopic measures of basicities are used for comparison. In addition, 6-hydroxycytidine is characterised more by the basic character of the O_2 atom compared to the more basic centre of cytidine, despite the fact that the GPB values of both these compounds are almost identical. Although the B3LYP/aug-cc-pVDZ approach seems to be an accurate and robust method for GPB estimation, the microscopic protonation properties are much more sensitive to this method since the difference in energy between some tautomers is often less than 1 kcal/mol with method-dependent succession. In such cases, G3MP2B3 is recommended.

Keywords Cytidine · Gas phase basicities · Hydroxyl radical · Protonation · Thymine · Uridine

Introduction

Protolytic equilibriums play an important role in the chemistry and biochemistry of heterocyclic compounds. There are many important biochemical species of this type, including the bases of nucleic acids and their numerous analogues, amino acids, drugs and many others [1–3]. Susceptibility to protonation or deprotonation affects such phenomena as pairing via hydrogen bonds, aromatic stacking, or interactions with metal ions. Thus, basicity or acidity may have a significant impact on a plethora of biochemical reactions, e.g. interactions with receptors and enzymes, base recognition and repair, transcription and translation. One of the spectacular consequences of nucleobase protonation is potential mutagenic mispairing of complementary bases [1, 2]. For example, cytosine protonated at the O_2-oxygen is thought to be responsible for the stabilisation of a adenine–cytosine mispair that was observed in single crystals of oligonucleotide duplexes [3].

A special class of heterocyclic compounds has been identified in and isolated from DNA cellular hydrolysates [4, 5]. These products of nucleobase degradation may be imposed by exogenous factors such as, for example, the ionising activity of ultraviolet radiation, chemical carcinogens, or reactive oxygen forms. They may also be formed during normal, nonpathogenic cellular processes, e.g. by the activities of superoxide-generating enzymes. There are direct correlations between levels of modified DNA bases and carcinogenesis, cancer and many non-cancerous diseases [6] since the presence of oxidised base lesions in DNA affects replication and transcription processes [7]. Thus, a knowledge of the chemical properties of these species is of particular importance. Although hydroxyl-radical-derived analogues were analysed in our previous investigations [8–14], open questions still remain. The main purpose of this study was to estimate the gas phase basicities (GPB) along with microscopic protolytic properties of such pyrimidine derivatives as 5-hydroxycytidine (CA), cytidine glycol (CB), 5,6-dihydroxycytidine (CC),

P. Cysewski (✉) · K. Kozłowska · B. Szefler
Department of Physical Chemistry, Collegium Medicum,
Nicolaus Copernicus University,
Kurpińskiego 5,
85-950 Bydgoszcz, Poland
e-mail: piotr.cysewski@cm.umk.pl

6-hydroxy-5,6-dihydrocytidine (CD), 5-hydroxy-5,6-dihydrocytidine (CE), 5-hydroxy-methyluridine (TA), thymidine glycol (TB), 5-hydroxy-5,6-dihydrothymidine (TC), 5,6-dihydrothymidine (TD), 6-hydroxy-5,6-dihydrothymidine (TE), 5-formyluridine (TF), isodialuric acid (UA), 5,6-dihydroxyuridine (UB), 5-hydroxyuridine (UC), 5-hydroxy-5,6-dihydrouridine (UD) and uridine glycol (UE). Since populations of neutral and protonated forms of heterocyclic compounds may involve a number of potential tautomers, the selection of an accurate and cost effective method is not a trivial task. Thus, the second goal of this study was to analyse the accuracy of selected quantum chemistry methods applied to description of a training set of compounds for which experimental data are available [15]. The GPB is a macroscopic property related directly to Gibbs free energy of protonation reaction. However, in the case of heterocyclic compounds, many potential tautomers of both neutral and protonated forms may be involved in the equilibrium. The following scheme defines the microscopic quantity:

$$GPB_{ij} = -\Delta G_{ij} \tag{1}$$

$$A_i + H^+ \xrightarrow{\Delta G_{ij}} B_j^+ \tag{2}$$

where A_i denotes i-th tautomer of base A and B_j^+ represents the j-th tautomer of its protonated form. The Gibbs free energy and equilibrium constant of the protonation reaction are related to each other by the elementary formulas:

$$\Delta G_{ij} = G_{B_j} - G_{A_i} - G_{H^+} = 1.36 \cdot pK_{ij} \tag{3}$$

$$K_{ij} = \frac{[B_j^+]}{[A_i] \cdot [H^+]} = \frac{[B^+] \cdot b_j}{[A] \cdot a_i \cdot [H^+]} = K \cdot \frac{b_j}{a_i} \tag{4}$$

where

$$[A] = \sum_i [A_i],\ [B^+] = \sum_j \left[B_j^+\right],\ a_i = \frac{[A_i]}{[A]},\ b_j = \frac{[B_j^+]}{[B^+]} \tag{5}$$

Since accurate experimental data [15] for gas phase protonation of all five nucleic acid bases are available, it is possible to verify the quality of theoretical predictions. It is necessary to estimate only the Gibbs free energies of all potential reactants along with their populations (a_i, b_j) derived from Boltzmann probabilities. The Gibbs free energy of proton formation in the gas phase at 1 atm and 25°C is equal to −6.28 kcal/mol, and may be calculated via the Sackur-Tetrode equation [16]. The final definition of GPB may then be presented as follows [17]:

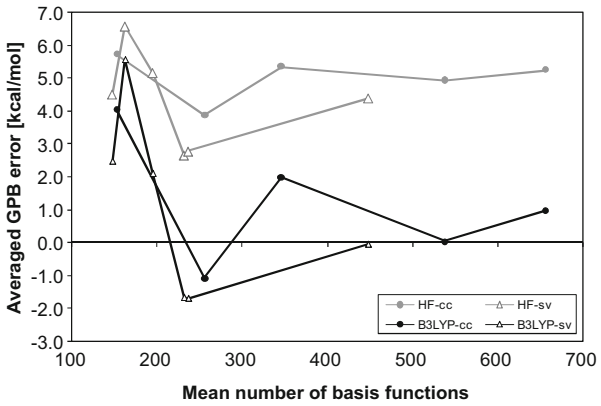
$$\Delta G = 1.36 \cdot pK = \Delta G_{ij} + 1.36 \cdot \log\left(\frac{b_j}{a_i}\right) = -GPB \tag{6}$$

In order to estimate the ΔG value, all potential tautomers were taken into account for both neutral and protonated forms of all species under investigation.

Methods

The application of Eq. 6 requires full geometry optimisation followed by hessian calculations for all possible tautomers of all 16 pyrimidine derivatives mentioned above. However, selection of a method is not straightforward since different levels of theory sometimes lead to contradictory predictions. Thus, in the first step, the methods must be verified against experimental data. However, GBP measurement results are available only for canonical nucleobases, and these compounds were used as a training set for selection of the computation level. All calculations were performed using the Gaussin03 package [18]. Although the relative energies of neutral and protonated nucleic acid bases were the subject of numerous studies, and experimental values were consistently reproduced by ab initio calculations, some additional notes are provided at this initial stage. Currently, advanced ab initio post-Hartree-Fock (HF) methods have been applied successfully to the identification of populations of neutral and protonated forms of DNA bases [19]. Correlation corrections up to MP4 were taken into account and the resulting values of proton affinities were within 2% of the experimental data. However, such demanding computations cannot be performed for the extended set of free-radical-derived analogues of nucleic acid bases. Thus, an alternative level has to be chosen. All five DNA bases were then subjected to full gradient geometry optimisation using HF, B3LYP, MP2, G3MP2 and G3MP2B3 methods. Also, the role of the basis set expansions was also studied for a series of split valence (sv) and correlation-consistent (cc) basis sets.

The analysis of GPB error is presented in Figs. 1 and 2. The experimental GPB values [15] were considered as the reference data. In Fig. 1, the averaged differences between experimental and predicted values of GPB were plotted as a function of the number of basis functions for standard sv and cc basis sets. The correspondence to common acronyms is also given. The data provided unequivocally dismiss the HF level since the predicted GPB values are significantly outside the experimental uncertainty, which is about 2.0 kcal/mol. In contrast, the application of a density functional method such as B3LYP is quite reasonable,



sv	ADE	GUA	CYT	THY	URA
6-31G(d)	160	175	130	147	128
6-31G(d,p)	175	190	145	165	140
6-311G(d,p)	210	228	174	198	168
6-311+G(d,p)	250	272	206	234	200
6-311++G(d,p)	255	277	211	240	204
6-311++G(3df,3pd)	480	519	402	459	384
cc	**ADE**	**GUA**	**CYT**	**THY**	**URA**
cc-pVDZ	165	179	137	156	132
aug-cc-pVDZ	275	298	229	261	220
cc-pVTZ	370	400	310	354	296
aug-cc-pVTZ	575	621	483	552	460
cc-pVQZ	700	755	590	675	560

cc - correlation consistent basis set
sv - split valence basis set

Fig. 1 The averaged error of gas phase basicities (GPB) of five nucleic acid bases estimated by the Hartree-Fock (HF) and B3LYP methods as a function of basis sets expansion. The numbers of primitive basis functions are also provided together with their common acronyms. *ADE* Adenine, *GUA* guanine, *CYT* cytosine, *THY* thymine, *URA* uracil

especially if extended basis sets were used. Unfortunately, a non-monotonous correlation between the accuracy of GPB and the number of basis functions used was observed, irrespective of the type of basis set. Of the basis sets studied here, aug-cc-pVTZ (aTZ) was the most accurate. This is a quite extended set, requiring more than 700 basis functions for purines. The same averaged error was obtained for the 6-311++G(3df,3pd) sv basis set, which in the case of guanine calls for 519 primitive functions. Fortunately, for the modest basis set, aug-cc-pVDZ (aDZ), the predicted values of GPB are within 1.0 kcal/mol with respect to the experimental data. Thus, this particular level was selected as providing a reasonable balance between the cost of computation and GPB accuracy.

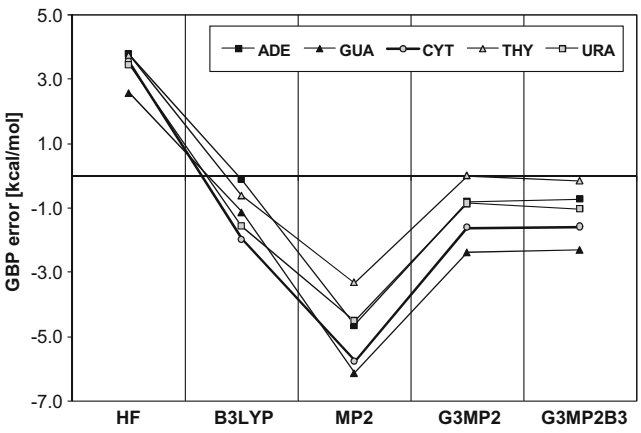


Fig. 2 The error of GPB of canonical purines and pyrimidines estimated by HF, B3LYP, and MP2 methods with an aDZ basis set and two composite methods: G3MP2 and G3MP2B3. The experimental values of GPB are equal to 218.1, 222.0, 221.7, 203.2 and 201.2 kcal/mol for adenine, guanine, cytosine, thymine and uracil, respectively [15]

Three additional methods were also checked (Fig. 2). The MP2/aDZ level of theory used along with two model chemistry methods provided a broader perspective on the error of GPB predictions. It is not surprising that the MP2 method is insufficient to characterise the protolytic properties due to the inherent errors in its calculations of normal vibration modes [20]. Compensation of this error by using different scaling parameters for different levels of theory was proposed more than decade ago by Scott and Radom [20]. Thermodynamic functions estimated without any scaling parameters, as in the MP2 method, are even more poorly predicted than those coming from just the HF/aDZ level. Model chemistry methods parameterised for a broad range of compounds are also very accurate in the case of calculations of protolytic properties of heterocyclic compounds.

The values of macroscopic quantities, such as proton affinities [19] or GBP [21], are known to depend mainly on the properties of the most probable tautomer. The inclusion in the analysis of the other, less stable, tautomers usually does not alter the basicity of DNA bases. However, the aim of this paper was not only to characterise the thermodynamic quantities but also to provide insight into the detailed structure of the neutral and protonated forms of the compounds analysed. Thus, additional inspection seems to be essential. Fortunately, there are valuable and precise experiments providing important details. Micro-basicity, as a measure of the susceptibility to protonation of individual proton-acceptor centres, was analysed here based on B3LYP/aDZ and G3MP2B3 methods; the resulting Boltzmann probabilities for all five canonical nucleic acid bases are presented in Figs. 3 and 4.

The acid–base equilibriums characterising purine bases have been the subject of numerous experimental and

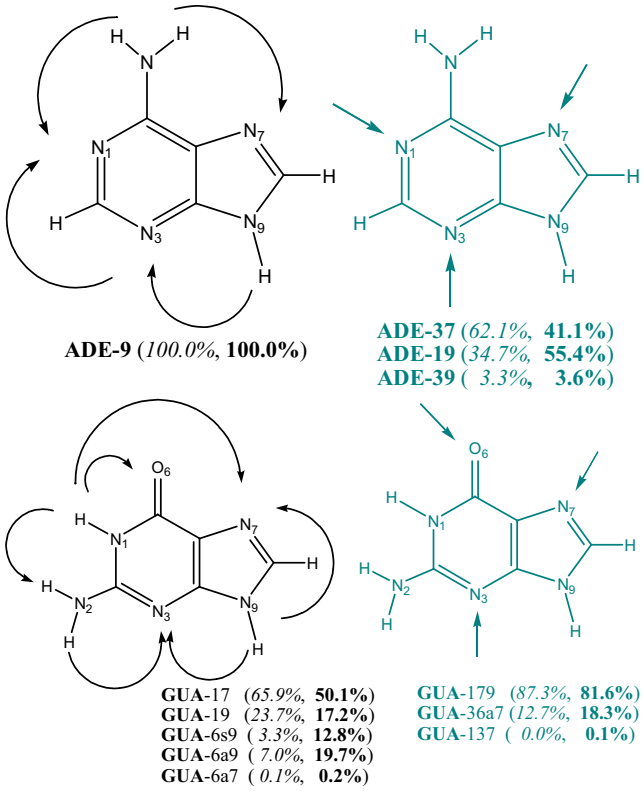
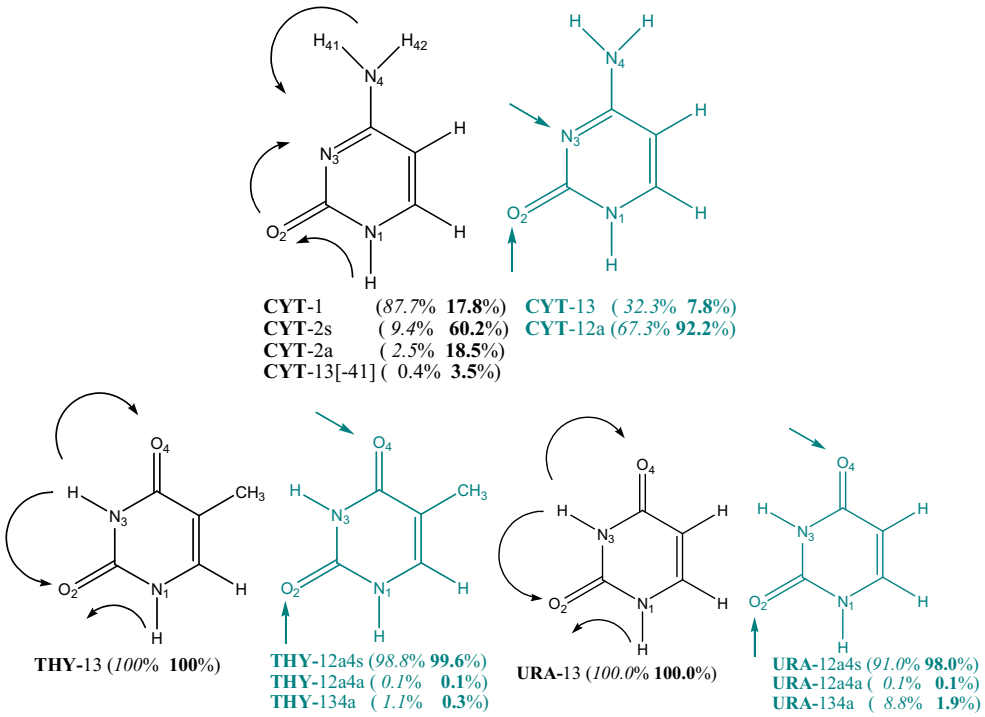


Fig. 3 The structures of neutral and protonated forms of canonical purines in the gas phase. Numbers in *bold* correspond to the Boltzmann populations estimated by the G3MP2B3 method, while those in *italics* indicate the results of B3LYP/aDZ calculations. Symbols of tautomeric forms describe the centres with attached protons followed by purine symbol. *s* and *a* represent the *syn* and *anti* conformations of side groups with respect to the N_1 atom, respectively

theoretical studies in the gas and condensed phases [21–28]. Several instrumental measurements, e.g. ultraviolet photoelectron spectroscopy [22], microwave spectroscopy [23] and R2PI and IR-UV double resonance spectra [24], have proved that the neutral adenine molecule is mainly in amino tautomeric form with hydrogen attached to the N_9 atom [H(N_9)-adenine]; H(N_7)-adenine was found to be less stable. On the other hand, under different reaction conditions adenine may be protonated in the gas phase at one of the basic nitrogen atoms, with varying composition of the tautomers in the mixture [25]. A recent study [26] demonstrated that protonation at the nitrogen in position N_1 of the neutral H(N_9)-adenine tautomer yields the most stable protonated adenine. The tautomers formed by protonation of H(N_7)- and H(N_9)-adenine at the N_3 position are slightly higher in energy. A previously reported theoretical analysis [26–28] of adenine tautomerism led to the same conclusions. The results of our calculations (Fig. 3) also confirm that, irrespective of the method applied, of the 12 possible unique tautomeric forms of adenine, the most abundant is the amino isomer with proton attached to the N_9 atom (ADE-9). In contrast, the protonation of adenine is not so consistently described by both applied methods, since B3LYP/aDZ calculations suggest that the most stable protonated adenine tautomer corresponds to a form with the proton attached to the N_3 and N_7 atoms (ADE-37), while G3MP2B3 computations suggest protonation of N_1 and N_9 centres (ADE-19). Since the energy difference between these two tautomers is very low (<0.4 kcal/mol), the order of adenine protonated forms

Fig. 4 The structures of neutral and protonated forms of canonical pyrimidines. Notation as in Fig. 3



does not affect the value of GPB. The predictions of G3MP2B3 methods agree with observed populations in the gas phase.

The guanine molecule has complex tautomeric properties, and may exist in several low-energy tautomeric forms [29, 30]. There is consensus that guanine has four low-energetic tautomers in the gas phase [19, 31–35]. However, under laser-desorption conditions, less favourable tautomeric species can be also populated to a minor extent. The most stable tautomer of guanine corresponds to an amino-keto tautomer with a hydrogen atom attached to the N_1 and N_7 centres. The protonation of guanine is thought to take place at the N_7 atom followed by the O_6 atom. Protonation at the N_3 site is highly unfavourable compared to the H(N_9)- and H(N_7)-species [33–38]. The micro-basicities of guanine are presented in Fig. 3. Both applied methods describe consistently the same succession of tautomers of neutral and protonated forms. The keto-amino tautomer of guanine with a proton attached to the N_7 atom is the most stable (GUA-17). This is in good accord with previously reported calculations and experimental observations [16–26]. The protonation of guanine leads to the structure GUA-179, which results from proton attachment to the N_9 centre of the GUA-17 tautomer. Both applied methods predict that more than 80% of this isomer is present in the population of protonated guanine.

Cytosine has been the subject of several experimental [39–43] and theoretical [19, 44–47] studies in the gas phase. Despite such interest, the experimental composition of tautomers in the gas phase is not precisely known. Both enol-amino and keto-amino forms have been identified in matrix isolation infrared and MW spectroscopy experiments [41, 42]. A higher concentration of the canonical form was suggested. This conclusion was additionally confirmed by thermo-chemical analysis [43]. Although there are no experimental data on cytosine protonation in the gas phase, computational studies [19, 44] strongly support the notion that the N_3 atom is the most basic in cytosine. The second stable tautomer of protonated cytosine is the one with a proton attached to the O_2 atom, with orientation of the resulting hydroxyl group towards the N_3 centre. The same conclusions may be drawn from Fig. 4, where the results of both methods applied here are presented. Although there are differences in the percentage of tautomer populations estimated based on B3LYP/aDZ and G3MP2B3 levels, the succession of tautomers is consistently described. Despite such differences, the values of GPB of cytosine are very similar and differences are within experimental uncertainty.

Thymine and uracil have similar chemical structures, and as a result their acid–base properties are also analogous. NMR [48–50], UV [51–54], IR, Raman [55, 56] and microwave spectroscopic [57] studies have provided evidence that these neutral pyrimidines exist in the gas phase predominantly in the 2,4-diketo form. This conclusion is consistent with previous reports [19, 44–47] and with the B3LYP/aDZ or G3MP2B3 results presented in Fig. 4. Experimental information on the composition of protonated thymine or uracil in the gas phase is not available to date. However, several computational studies have been published, all of which consistently suggest that the protonation of thymine and uracil takes place at oxygen atoms rather than at nitrogen centres of the heterocyclic rings [19, 28, 58–60]. Unfortunately, there is a lack of consensus about the sequence of potential tautomeric forms. Podolyan et al. [19] suggested that the most stable form corresponds to an isomer with a proton attached to the O_4 atom. However, the results presented in Fig. 4 suggest that the forms protonated at both oxygen atoms, THY-12a4s and URA-12a4s, are the most stable among all possible protonated species. These tautomers result from the migration of a hydrogen atom from the N_3 centre toward the O_2 atom of O_4-protonated thymine or uracil. Both applied methods lead to the same conclusion. This is the only discrepancy between previously reported basicities of nucleic acid bases. It is worth mentioning that the 2,4-dihydroxy tautomer has been suggested as the most stable protonated form of protonated uracil [60]. This supports the reliability of our calculations.

In the conclusion of this initial stage, it is worth mentioning that the B3LYP/aDZ approach seems to provide a rational balance between cost and accuracy. Although this method is precise enough to estimate GPB of native DNA bases, the microscopic protonation features are much more sensitive to the method applied since the difference in energies between some tautomers is often less than 1 kcal/mol, with method-dependent succession. The correct sequence of neutral and cationic forms may, however, be obtained by using one of the model chemistry approaches, e.g. G3MP2B3. In cases where the B3LYP/aug-cc-pvdz and G3MP2B3 methods lead to contradictory predictions of the order of neutral or protonated tautomers, the latter is suggested to be used in the interpretation of microscopic protonation properties. Such disagreements between succession of tautomeric forms predicted by these two methods were noted for purines but were not observed for pyrimidines. If only macroscopic properties are required, then the B3LYP/aug-cc-pvdz level is sufficient since it provides GPB values with 1.0 kcal/mol accuracy with respect to experimental data.

The GPB of 16 pyrimidine analogs known as by-products of oxidative stress were then estimated based on B3LYP/aug-cc-pvdz and G3MP2B3 methods. Model nucleosides, which, instead of the whole 2′-deoxy-ribose ring, comprised only a methylmethoxy group (-CH_2-O-CH_3) mimicking the N-glycosidic bond, were taken into account. Such a model was previously documented as reliable [11, 12] ,as it

provides all important features of nucleosides but with a significant reduction in size of the studied systems. Furthermore, the impact of many sugar-related conformations is omitted. Puckering of the five-membered ring, side group rotation, and syn-anti N-glycosidic bond conformations are irrelevant for micro- and macro-basicities of the systems analysed. The results of thermodynamic calculations were critically checked against the imaginary frequencies, ensuring that the geometries obtained correspond to global minima in potential hyperspace. Tight criteria were imposed for gradient minimisation. In addition, the ultra fine option was used for definition of the grid mesh in density functional theory (DFT) computations.

Results and discussion

GPB of model cytosine analogs

At least five known cytidine analogues imposed by reactive oxygen forms are known. The following products have been identified in chromatin hydrolysates [1–4]: 5-hydroxycytidine (CA), cytidine glycol (CB), 5,6-dihydroxycytidine (CC), 6-hydroxy-5,6-dihydrocytidine (CD) and 5-hydroxy-5,6-dihydrocytidine (CE) (structures presented in Fig. 5). All these analogues posses complex tautomeric equilibriums in both their neutral [10] and protonated forms. The first of the model cytidine derivatives studied here, 5-hydroxycytidine, can exist in its neutral form as a mixture of keto-imino and keto-amino tautomers. Both applied methods, B3LYP/aDZ and G3MP2B3, lead to the same predictions, but the latter slightly favours the most probable isomer (CA2-1a). This agrees with the common expectation [10] that the non-polar environment promotes the less polar species. The dipole moment of the most abundant keto-imino tautomer (CA2-1a) is much smaller than that of the keto-amino isomer (CA1). The corresponding values are equal to 1.76D and 6.32D, respectively. Interestingly, the protonation of 5-hydroxycytidine leads only to amino forms, and the predominant isomer of cationic 5-hydroxycytidine, CA1-O2b, adopts the enol-amino structure. Again, a slight increase in the percentage of the population of the most probable isomer is observed if the composite method is used.

From the values of GPB presented in Table 1, it can be concluded that modification of 5-hydroxycytidine has an insignificant impact on the macroscopic measure of basicity. However, the microscopic properties, describing basicities of particular proton–acceptor centres, are altered compared to model cytidine. The canonical cytidine model takes its basic character from the N_3 and O_2 centres, and both these sites posses almost identical basicities. In 5-hydroxycytidine the same atoms are involved in the protonation process, and the N_4 centre of the imino tautomers may also be susceptible to proton attachment. The most basic character exhibits an O_2 centre, and both nitrogen atoms are slightly less basic (see Table 1). Thus, although the GPB of C and CA are almost the same, the micro-basicities are affected by oxidation of cytidine at the C_5-centre, and there is an observed increase in the basicity of the O_2 atom for CA compared to canonical cytidine.

The next analogue studied in this paper is cytidine glycol (CB). CB may have quite an intricate structure since it posses two chirality centres. Among all possible stereo isomers and tautomers, CB2(5ax,6ax) was identified as the most probable form of neutral CB. This imino-keto tautomer has both chirality centres set to axial conformation. Although both of the applied methods point to this neutral form as the most probable form, the model chemistry method suggests a much higher percentage of this isomer in the total population. The products of CB protonation can also adopt numerous tautomers and diastereoisomers but two predominate in the entire population (Fig. 5). The form denoted as CB3(5ax,6ax) was found to be the most probable product of protonation, and may be treated either as a protonated keto-imino tautomer at the N_4 centre or as a product of proton attachment to the N_3 nitrogen atom of the amino-keto form. CB is less basic than canonical cytidine by about 5 kcal/mol. The micro-basicities provided in Table 1 suggest that the N_3 atom is the most basic one in the case of CB, but that it is still less basic than the O_2 centre of cytidine. Thus, both macro- and micro-basicities of CB are lower than those of cytidine.

The third analogue of cytidine studied is its 5,6-dihydroxy derivative (CC). Although the keto-amino tautomer CC1(bd) was identified as the most stable among all potential isomers, the keto-imino tautomer CC3(1,ac) may also occur in the population of neutral forms. Thus, there are three potential basicity centres and, as shown in Table 1, the N_3 atom is responsible for the most basic character of this derivative. Interestingly, CC is more basic compared to cytidine both from the macroscopic and microscopic point of view, and even the least basic atom (N_4) is as basic as the O_2 centre of cytidine.

The fourth product of cytidine oxidation analysed here, 6-hydroxy-5,6-dihydrocytidine (CD), has similar tautomeric properties as the above-described cytidine derivatives. Again, keto-amino and keto-imino tautomers occur in the population of neutral species, and the latter predominates over all other structures. The axial conformation of the C_6 atom is energetically favourable over the equatorial one. This conclusion is consistently drawn based on both applied computational procedures. Although the macroscopic basicity of CD is about 3 kcal/mol smaller compared to that of cytidine, the most basic sites of CD and C are equally susceptible to proton attachment.

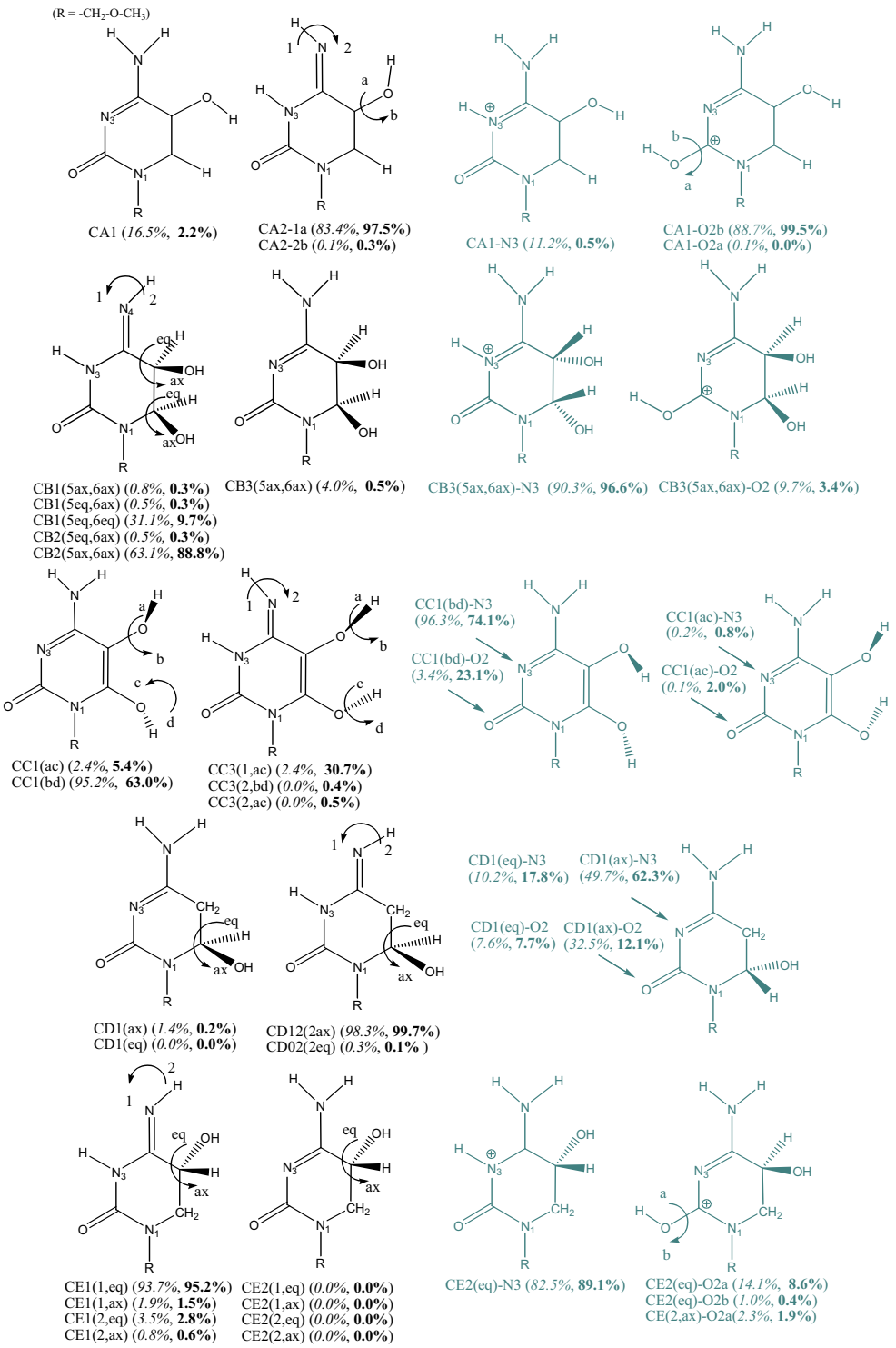


Fig. 5 The structures of neutral and protonated forms of hydroxyl radical modified model cytidine: *CA* 5-hydroxy-cytidine; *CB* cytidine glycol; *CC* 5,6-dihydroxycytidine; *CD* 6-hydroxy-5,6-dihydrocytidine; *CE* 5-hydroxy-5,6-dihydrocytidine. Notation as in Fig. 3

The final cytidine analogue studied here, 5-hydroxy-5, 6-dihydrocytidine (CE), has very simple tautomeric features. The only expected tautomer of the neutral form exists as a keto-imino isomer. Since the N_3 centre of this tautomer is capped by a hydrogen atom, this site cannot exhibit its basic nature directly, and only two centres are available for protonation, namely the O_2 and N_4 centres. There is a strong difference between the basicities of these two atoms and although the former has a much higher affinity for protonation, the resulting cationic form CE2(eq) is energetically disfavoured. Thus, the overall basic character of CE comes from the basicity of the N_4 centre. This derivative is thought to be the least basic among all the cytidines studied here, including canonical cytidine.

Table 1 Micro- and macro- gas phase basicities (GPB) (in kcal/mol) of canonical and hydroxyl-radical-modified model pyrimidines. *CA* 5-Hydroxycytidine; *CB* cytidine glycol; *CC* 5,6-dihydroxycytidine; *CD* 6-hydroxy-5,6-dihydrocytidine; *CE* 5-hydroxy-5,6-dihydrocytidine; *C* cytidine; *TA* 5-hydroxy-methyluridine; *TB* thymidine glycol; *TC* 5-hydroxy-5,6-dihydrothymidine; *TD* 5,6-dihydrothymidine; *TE* 6-hydroxy-5,6-dihydrothymidine; *TF* 5-formyluridine; *T* thymidine; *UA* isodialuric acid; *UB* 5,6-dihydroxyuridine; *UC* 5-hydroxyuridine; *UD* 5-hydroxy-5,6-dihydrouridine; *UE* uridine glycol; *U* uracil

	Macroscopic basicities	Microscopic basicities
CA	224.8 (225.0)	O_2(227.2) > N_4(221.9) ≈ N_3(222.1)
CB	220.9 (220.2)	N_3(223.3) > O_2(221.3) > N_4(220.2)
CC	227.4 (226.6)	N_3(228.1) > O_2(226.0) > N_4(224.4)
CD	222.0 (221.2)	N_3(224.7) > O_2(223.7) > N_4(220.9)
CE	220.0 (219.0)	O_2(223.2) > N_4(219.0)
C	224.6 (224.9)	O_2(224.6) ≈ N_3(224.2)
TA	204.5 (205.1)	O_4(203.6) > O_2(198.9)
TB	201.4 (201.8)	O_2(201.7) > O_4(194.0)
TC	204.5 (205.0)	O_2(205.0) > O_4(193.1)
TD	204.7 (205.1)	O_2(204.3) > O_4(197.2)
TE	201.7 (201.7)	O_2(201.8) > O_4(196.1)
TF	206.3 (206.8)	O_4(206.6) ≈ O_5(206.0) > O_2(187.0)
T	206.6 (207.2)	O_2(207.2) > O_4(205.0)
UA	194.3 (194.1)	O_2(195.5) > O_4(189.6)
UB	202.4 (202.5)	O_4(202.4) > O_2(200.6)
UC	200.7 (201.1)	O_4(200.7) ≈ O_2(200.0)
UD	201.8 (202.9)	O_2(201.4) > O_4(195.8)
UE	199.9 (200.5)	O_2(200.5) > O_4(192.4)
U	204.7 (205.1)	O_2(205.1) > O_4(203.8)

GPB of model thymidine analogues

Six different products of thymine degradation by reactive oxygen forms have been identified [1–4]: 5-hydroxy-methyluridine (TA), thymidine glycol (TB), 5-hydroxy-5,6-dihydrothymidine (TC), 5,6-dihydrothymidine (TD), 6-hydroxy-5,6-dihydrothymidine (TE) and 5-formyluridine (TF) (structures presented in Fig. 6). The tautomeric properties of all these compounds are much simpler than their cytidine analogues, since all derivatives adopt a 2,4-diketo form [10]. Although some of these compounds can exist as a mixture of different isomers, their structures result from either chirality conformations or side group rotation rather than from tautomeric equilibriums. The basic character of canonical thymidine is about 19 kcal/mol weaker than that of cytidine. Consequently, all thymidine analogues studied here are extremely weak bases, and in all cases oxidation reduces the basic nature of the resulting compounds even further. For example, the only stable tautomer of 5-hydroxy-methyluridine (TA) adopts a diketo form, and can be protonated only at O_2 and O_4 centres (Fig. 6). Protonation of TA can lead either to 2-keto-4-enol or 2,4-dienol cations, with the population of the latter being most abundant. Thymidine itself has the same proton acceptor centres and the O_2 site is about 2 kcal/mol more basic than the O_4 atom. However, in 5-hydroxy-methyluridine the order of the micro-basicities is reversed. Nevertheless, both macro- and micro-basicities are lower for TA compared to canonical thymidine.

The neutral form of the second derivative, thymidine glycol (TB), can exist as a mixture of different 2,4-diketo tautomers with different conformations of two chirality centres. Among all four possible structures, the one characterised by 5-equatorial and 6-axial conformation, TB(5eq,6ax), was identified as the most favourable by both applied methods. The succession of O_2 and O_4 basicities is the same as for non-modified model thymidine. Similar conclusions may be drawn for next three derivatives, namely 5-hydroxy-5,6-dihydrothymidine (TC), 5,6-dihydrothymidine (TD) and 6-hydroxy-5,6-dihydrothymidine (TE). In all these cases, the 5-equatorial conformations are energetically more stable than the 5-axial conformations for neutral forms. The only exception is 6-hydroxy-5,6-dihydrothymidine, for which TE(5eq,6eq) and TE(5ax,6eq) are almost equally probable. However, the equatorial conformation of the C_6 site is profoundly favoured over the axial one. Irrespective of the distereoisomer, protonation leads to 2-enol-4-keto forms of TC, TD and TE. In the case of TC, protonation at the N_3 site leads to promotion of equatorial conformation of the C_5 atom. This is the only case among all the derivatives studied for which protonation is accomplished with changes in centre chirality.

The last thymidine analogue has interesting protonation abilities. In the neutral form, free rotation of the formyl group bound to the C_5 carbon atom is possible. However, after protonation, the site group is fixed in syn orientation with respect to the O_4 atom. This allows for hydrogen attachment in two ways: on either O_4 or O_5 basic centres. In whichever case, a very strong and short intra-molecular hydrogen bond is formed (see Fig. 6). The basic character of the O_4 site is slightly stronger than that of the O_5 centre and much stronger compared to the O_2 oxygen atom. The basicity of 5-formyluridine is almost the same as canonical thymidine if macroscopic GPB is taken into account.

GPB of model uridine analogues

The final set of hydroxyl radical analogues comprised the following uridine derivatives: isodialuric acid (UA), 5,6-dihydroxyuridine (UB), 5-hydroxyuridine (UC), 5-hydroxy-5,6-dihydrouridine (UD) and uridine glycol (UE). As presented in Fig. 7, all neutral forms have simple tautomeric properties. The only stable tautomer adopts a 2,4-diketo form. As one may expect, there are great formal similarities between thymidine and uridine since the same centres define the basic nature of these compounds. However, model uridine is even less basic than thymidine, in terms of both macro- and microscopic quantities. All five uridine derivatives analysed are less basic than non-modified uridine. The

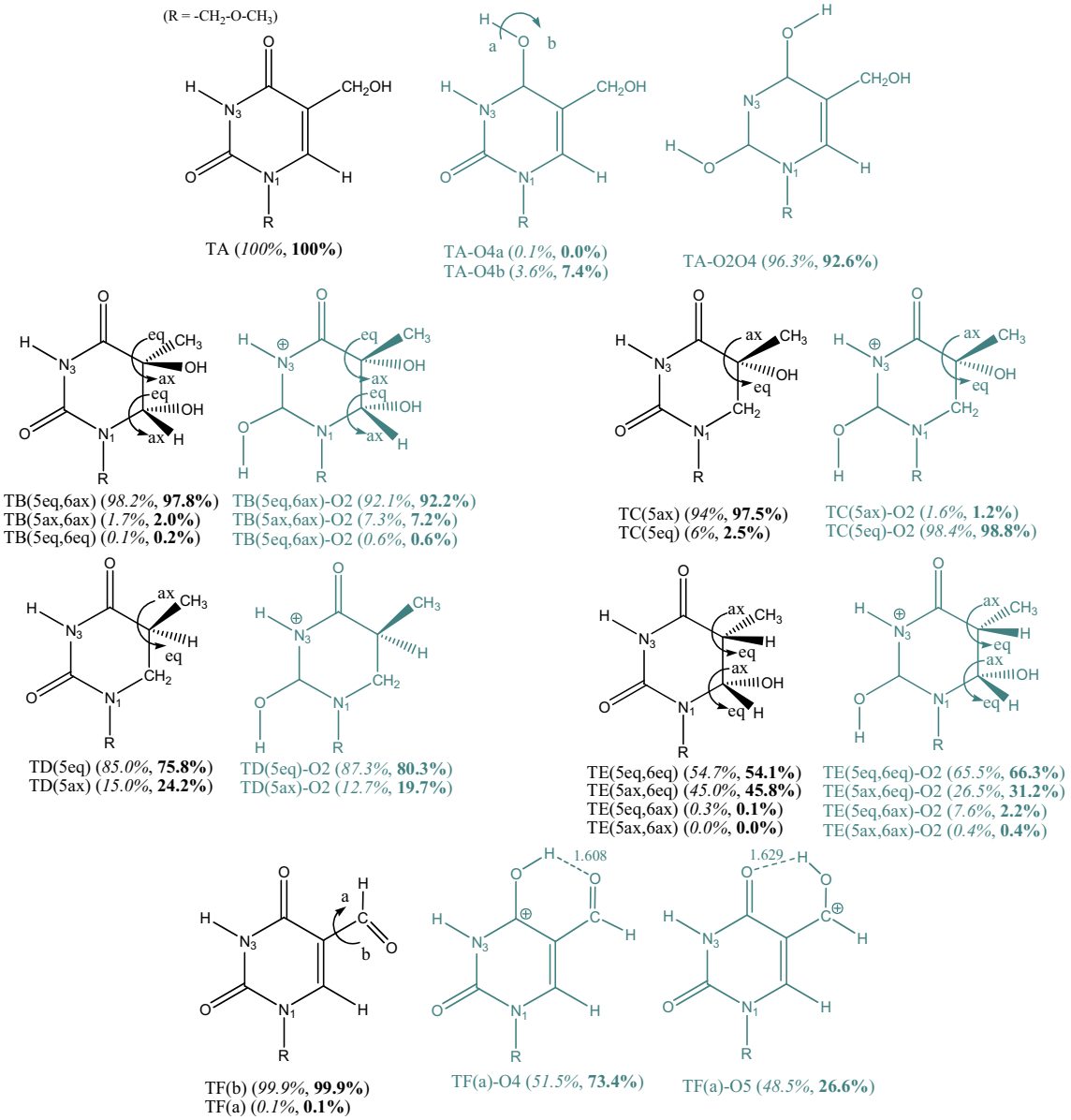


Fig. 6 The structures of neutral and protonated forms of hydroxyl radical modified model thymidine: *TA* 5-hydroxy-methyluridine; *TB* thymidine glycol; *TC* 5-hydroxy-5,6-dihydrothymidine; *TD* 5,6-dihydro-thymidine; *TE* 6-hydroxy-5,6-dihydrothymidine; *TF* 5-formyluridine. Notation as in Fig. 3

smallest values of GPB were predicted for isodialuric acid at both B3LYP/aDZ and G3MP2B3 levels. The protonation of this derivative leads to a mixture of 2-enol-4-keto, 2-keto-4-enol and 2,4-dienol isomers. The former, UA(ax), is the most abundant form of isodialuric acid, and has axial conformation of the C_6 atom. The total percentage of all enol forms does not exceed 1.6% as predicted by the G3MP2B3 method. The O_2 centre of UA is the least basic of all primary protonated sites of all the compounds studied here.

The next two uridine derivatives, 5,6-dihydroxyuridine (UB) and 5-hydroxy uridine (UC), are characterised by reversed basicities of O_4 and O_2 centres as compared to canonical uridine. However, the difference between the basicities of these two sites is modest, especially for UC. The most stable protonated form of 5,6-dihydroxyuridine corresponds to 2-keto-4-enol isomer (UB-O4b), while for 5-hydroxy uridine, two isomers, 2-keto-4-enol (UC-O4) and 2,4-dienol (UC-O2O4) occur with comparable percentage in the total population.

The last two derivatives studied were 5-hydroxy-5,6-dihydrouridine (UD) and uridine glycol (UE). These two compounds are characterised by very similar tautomeric and prototropic properties. Although UD is deficient in one C_6 chirality centre, neutral forms both of UD and UE adopt 2,4-diketo tautomeric form. Their protonated species exist predominantly as 2-enol-4-keto cations. UE is slightly

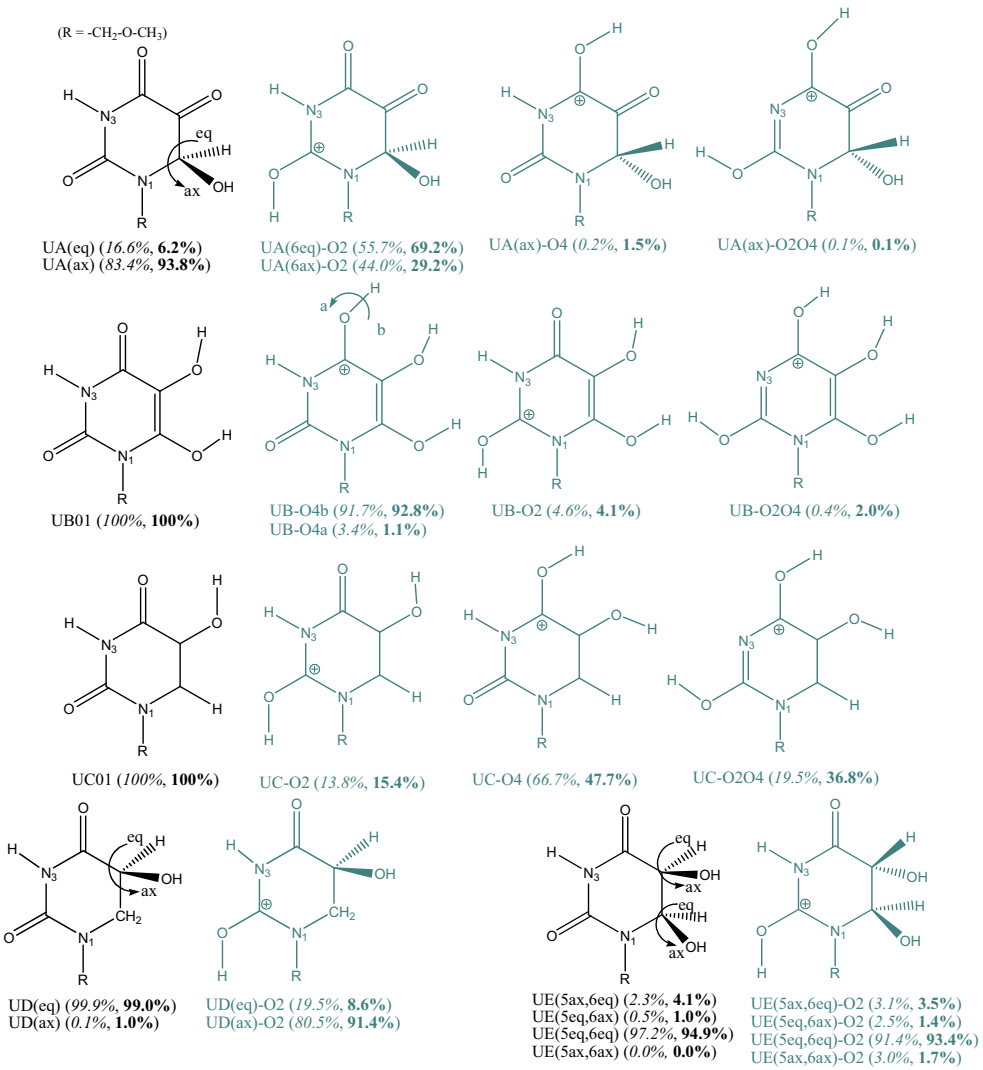


Fig. 7 The structures of neutral and protonated forms of hydroxyl radical modified model uridine: *UA* isodialuric acid; *UB* 5,6-dihydroxyuridine; *UC* 5-hydroxyuridine; *UD* 5-hydroxy-5,6-dihydrouridine; *UE* uridine glycol. Notation as in Fig. 3

less basic than UD with respect both to GPB values and microscopic measure of basicity. Interestingly, protonation of the UD derivative imposes an alteration of C_5 chirality analogous to that of the TC derivative.

Conclusions

The data presented above suggest unequivocally that both methods applied here, B3LYP/aDZ and G3MP2B3, are accurate enough for reliable prediction of macro- and micro-scopic GPB of pyridines and their reactive-oxygen-derived analogues. The consistent characteristics of the prototropic features of heterocyclic compounds requires taking into account many possible tautomers in both neutral and cationic forms. The similar characteristics of the purine analogues provided a warning [21] that these two methods may sometimes lead to contradictory descriptions of the succession of particular tautomers. Correlation with available experimental data suggests that the predictions of G3-like methods are more reliable than those of B3LYP/aDZ. Fortunately, in the case of pyrimidines, both applied levels of theory were always in agreement and provided the same order of neutral and protonated forms for all the hydroxyl radical analogues studied here. However, slight differences in percentages of populations were noted in some cases. Fortunately, this did not lead to changes in macroscopic basicities since such values are related mainly to the properties of the dominant tautomer [19]. Thus, the much less expensive level B3LYP/aDZ is recommended for characterisation of equilibriums involving heterocyclic compounds similar to those studied in this project. However, for the quantitative description of micro-basicities, model chemistries methods are recommended.

Among the three classes of derivatives studied in this paper, only cytidine and its derivatives have complex and intricate tautomeric properties. On the contrary, the thymidine and uridine derivatives are present mostly as one tautomer in their neutral forms. Usually, the tautomers found to be the most stable neutral species are characterised

by the lowest dipole moment values. The macroscopic measure of basicity, GPB, usually comes from dominant tautomers (i.e. energetically most favourable) and need not be associated with the most basic centres characterising micro-basicities. According to the results presented, cytidine and all its analogues are much stronger bases than thymine and uridine and their derivatives. The oxidation of pyrimidines usually leads to a decrease in basic character, with only two exceptions: 5,6-dihydroxycytidine is more basic then model cytidine if both macro- and micro-scopic measures of basicities are used for comparison. Besides, 5-hydroxycytidine is characterised more by the basic character of the O_2 atom than by the basic centre of cytidine; however, the macroscopic basicities GPB are almost the same for CA and C. Formyluridine is almost as basic as model thymidine. Other hydroxyl-radical-modified pyrimidines are less basic then their canonical equivalents.

Acknowledgement The results were partly obtained based on computational grants from PCSS (Poznań Supercomputing and Networking Centre, Poland). The allocation of computing time is greatly appreciated.

References

1. Saenger W (1988) Principles of nucleic acid structures. Springer, Berlin
2. Sinden RR (1994) DNA structure and function. Academic, San Diego
3. Hunter WN, Brown T, Anand NN, Kennard O (1986) Nature 320:552–555. doi:10.1038/320552a0
4. Dizdaroglu M (1993) Chemistry of free radical damage to DNA and nucleoproteins. In: Halliwell B, Aruoma OI (eds) DNA and free radicals, Ellis Horwood, New York, pp 19–39
5. Evans MD, Griffiths HR, Lunec J (1997) Reactive oxygen species and their cytotoxic mechanisms. In: Chipman JK (ed) Mechanisms of cell toxicity. JAI, London
6. Cooke MS, Evans MD, Dizdaroglu M, Lunec J (2003) FASEB J 17:1195–1214. doi:10.1096/fj.02-0752rev
7. Cooke MS, Oliński R, Evans MD (2006) Clin Chim Acta 365:30–49. doi:10.1016/j.cca.2005.09.009
8. Cysewski P (2005) Z Phys Chem 219:213–234. doi:10.1524/zpch.219.2.213.57306
9. Rogstad KN, Jang YH, Sowers LC, Goddard WA III (2003) Chem Res Toxicol 16:1455–1462. doi:10.1021/tx034068e
10. Cysewski P (1998) Z Naturforschung 53C:1027–1036
11. Cysewski P, Bednarek D, Kozłowska K (2003) Phys Chem Chem Phys 5:4899–4904. doi:10.1039/b307744f
12. Cysewski P, Bira D, Białkowski K (2004) J Mol Struct THEOCHEM 678:77–81. doi:10.1016/j.theochem.2004.03.010
13. Cysewski P (1999) J Mol Struct THEOCHEM 466:49–58. doi:10.1016/S0166-1280(98)00337-6
14. Cysewski P, Jeziorek D (1998) J Mol Struct THEOCHEM 430:219–229
15. Afeefy HY, Liebman JF, Stein SE (2005) In: Linstrom PJ, Mallard WG (eds) Neutral thermochemical data in NIST Chemistry WebBook, NIST Standard Reference Database Number 69. http://webbook.nist.gov National Institute of Standards and Technology, Gaithersburg
16. Noggle JH (1996) Physical chemistry, 3rd edn. Harper Collins, New York
17. Jang YH, Goddard WA III, Noyes KT, Sowers LC, Hwang S, Chung DS (2002) Chem Res Toxicol 15:1023–1035. doi:10.1021/tx010146r
18. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02, Gaussian, Inc, Wallingford CT
19. Podolyan Y, Gorb L, Leszczynski J (2000) J Phys Chem A 104:7346–7352. doi:10.1021/jp000740u
20. Scott AP, Radom L (1996) J Phys Chem 100:16502–16513. doi:10.1021/jp960976r
21. Cysewski P, Kozłowska K (2008) Pol J Chem 12: (in press)
22. Lin J, Yu C, Peng S, Akiyama I, Li K, Lee Li Kao, LeBreton PR (1980) J Am Chem Soc 102:4627–4631. doi:10.1021/ja00534a010
23. Brown RD, Godfrey PD, McNaughton D, Pierlot AP (1989) Chem Phys Lett 156:61–63. doi:10.1016/0009-2614(89)87081-2
24. Plutzer C, Nir E, de Vries MS, Kleinermanns K (2001) Phys Chem Chem Phys 3:5466–5469. doi:10.1039/b107997b
25. Chen X, Syrstad EA, Gerbaux P, Nguyen MT, Turecek F (2004) J Phys Chem A 108:9283–9293. doi:10.1021/jp046575q
26. Marian C, Nolting D, Weinkauf R (2005) Phys Chem Chem Phys 7:3306–3316. doi:10.1039/b507422c
27. Guerra CF, Bickelhaupt FM, Saha S, Wang F (2006) J Phys Chem A 110:4012–4020. doi:10.1021/jp057275r
28. Chandra AK, Nguyen MT, Uchimaru T, Zeegers-Huyskens T (1999) J Phys Chem A 103:8853–8860. doi:10.1021/jp990647+
29. Tureček F, Chen X (2005) J Am Soc Mass Spectrom 16:1713–1726. doi:10.1016/j.jasms.2005.06.010
30. Marian CM (2007) J Phys Chem A 111:1545–1553. doi:10.1021/jp068620v
31. Christensen JJ, Rytting JH, Izatt RM (1967) J Phys Chem 71:2700. doi:10.1021/j100867a047
32. Piacenza M, Grimme S (2004) J Comput Chem 25:83–99. doi:10.1002/jcc.10365
33. Hanus M, Ryjaček F, Kabelač M, Kubar T, Bogdan TV, Trygubenko SA, Hobza P (2003) J Am Chem Soc 125:7678–7688. doi:10.1021/ja034245y
34. Langer H, Doltsinis NL (2003) J Chem Phys 118:5400. doi:10.1063/1.1555121
35. Shukla MK, Leszczynski J (2005) J Phys Chem A 109:7775–7780. doi:10.1021/jp052340i
36. Choi MY, Miller RE (2006) J Am Chem Soc 128:7320–7328. doi:10.1021/ja0607411
37. Mons M, Dimicoll I, Piuzzi F, Tardivel B, Elhanine M (2002) J Phys Chem A 106:5088–5094. doi:10.1021/jp0139742
38. Crews B, Abo-Riziq A, Grace L, Callahan M, Kabelac M, Hobza P, de Vries MS (2005) Phys Chem Chem Phys 7:3015–3020. doi:10.1039/b506107e
39. Nir E, Muller M, Grace L, de Vries MS (2002) Chem Phys Lett 355:59–64. doi:10.1016/S0009-2614(02)00180-X

40. Dreyfus M, Bensaude O, Dodin G, Dubois JE (1976) J Am Chem Soc 98:6338–6349. doi:10.1021/ja00436a045
41. Szczesniak M, Szczepaniak K, Kwiatkowski JS, KuBulat K, Person WB (1988) J Am Chem Soc 110:8319–8330. doi:10.1021/ja00233a006
42. Brown RD, Peter D, Godfrey PD, McNaughton D, Pierlot AP (1989) J Am Chem Soc 111:2308–2310. doi:10.1021/ja00188a058
43. Sambrano JR, de Souza AR, Queralt JJ, Andres J (2000) Chem Phys Lett 317:437–443. doi:10.1016/S0009-2614(99)01394-9
44. Leszczynski J (1999) In: Schleyer PR (ed) Encyclopedia of computational chemistry, vol. 5. Wiley, New York
45. Colominas C, Luque FJ, Orozco M (1996) J Am Chem Soc 118:6811–6821. doi:10.1021/ja954293l
46. Bonaccorsi R, Scrocco E, Tomasi J, Pullman A (1975) Theor Chim Acta 36:339–350. doi:10.1007/BF00549697
47. Yang Z, Rodgers MT (2004) Phys Chem Chem Phys 6:2749–2757. doi:10.1039/b315089e
48. Ts'o POP (1974), Basic principles in nucleic acids chemistry, Ts'o POP (ed) Academic, New York
49. Topal MD, Fresco JR (1976) Nature 263:285–289. doi:10.1038/263285a0
50. Ruterjans H, Kaun E, Hall WE, Limbach HH (1982) Nucleic Acids Res 10:7027–7039. doi:10.1093/nar/10.21.7027
51. Morsy MA, Al-Somali AM, Suwaiyan A (1999) J Phys Chem B 103:11205–11210. doi:10.1021/jp990858e
52. Schiedt J, Weinkauf R, Neumark DM, Schlag EW (1998) Chem Phys 239:511–524. doi:10.1016/S0301-0104(98)00361-9
53. Alyoubi AO, Hilal RH (1995) Biophys Chem 55:231–237. doi:10.1016/0301-4622(95)00002-F
54. Kubota M, Kobayashi T (1996) J Electron Spectrosc Relat Phenom 82:61. doi:10.1016/S0368-2048(96)03047-2
55. Chin S, Scott I, Szczepaniak K, Person WB (1984) J Am Chem Soc 106:3415–3422. doi:10.1021/ja00324a006
56. Ivanov AY, Plokhotnichenko AM, Radchenko ED, Sheina GG, Blagoi YP (1995) J Mol Struct THEOCHEM 372:91–100. doi:10.1016/S0166-1280(05)80001-6
57. Brown RD, Peter D, Godfrey PD, McNaughton D, Pierlot AP (1988) J Am Chem Soc 110:2329–2330. doi:10.1021/ja00215a069
58. Russo N, Toscano M, Grand A, Jolibois F (1998) J Comput Chem 19:989–1000. doi:10.1002/(SICI)1096-987X(19980715)19:9<989::AID-JCC1>3.0.CO;2-F
59. Kryachko ES, Nguyen MT, Zeegers-Huyskens T (2001) J Phys Chem A 105:1288–1295. doi:10.1021/jp001031j
60. Wolkena JK, Tureček F (2000) J Am Soc Mass Spectrom 11:1065–1071. doi:10.1016/S1044-0305(00)00176-8