

Aldehyde-hydrate equilibrium in nucleobase 2-oxoethyl derivatives: An NMR, ESI-MS and theoretical study



Mariano J. Nigro^a, Juan I. Brardinelli^b, Elizabeth S. Lewkowicz^{a, b}, Adolfo M. Iribarren^a, Sergio L. Laurella^{c, *}

^a Laboratorio de Biotransformaciones y Biotransformaciones, Universidad Nacional de Quilmes, Roque Sáenz Peña 352, 1876 Bernal, Buenos Aires, Argentina

^b Plataforma de Servicios Biotecnológicos, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Sáenz Peña 352, 1876 Bernal, Buenos Aires, Argentina

^c Laboratorio LADECOR, División Química Orgánica, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, 47 y 115, 1900 La Plata, Buenos Aires, Argentina

ARTICLE INFO

Article history:

Received 20 December 2016

Received in revised form

1 May 2017

Accepted 2 May 2017

Available online 5 May 2017

Keywords:

Adenine

Thymine

Aldehyde

Hydrate

ABSTRACT

N-2-oxoethyl derivatives of nucleobases are useful starting materials for the preparation of potentially active nucleoside analogues. The ¹HNMR, ¹³CNMR, DEPT and ESI-MS spectra of adenine and thymine *N*-2-oxoethyl derivatives reveal that the different species in equilibrium exist mainly in two forms: aldehyde and hydrate. The NMR spectra show that the equilibrium is shifted towards the hydrate form in water-DMSO 2:1, giving equilibrium constants of 8.3 and 5.3 for adenine and thymine derivatives, respectively. ESI-MS experiments show the dependence of equilibrium shift on pH: in the case of the thymine derivative, the effect on the equilibrium is more important than in the case of the adenine derivative; this difference is explained considering different protonation sites in both structures. All assumptions are supported by theoretical calculations, which suggest the important role played by solvent in the stabilization of molecular structures and equilibrium shift. All aspects analyzed in this work are very important in order to understand the further reactivity of these nucleobase derivatives.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The aldol reaction is one of the most powerful strategies in synthetic organic chemistry to perform carbon–carbon bond formation and besides, the resulting aldol products are valuable for the synthesis of biologically important compounds. During this reaction, two simple carbonylic compounds are transformed into a more complex one and, as a consequence, two new stereogenic centers can be formed. Since the lack of selectivity of the classical aldol reaction constrains its application, transition metal catalysis, biocatalysis and organocatalysis have emerged as advantageous and sustainable catalytic alternatives [1].

Regarding biocatalysis, aldolases catalyze the reversible aldol addition of a nucleophilic donor, usually a ketone, onto an electrophilic aldehyde acceptor. These enzymes can usually accept a wide variety of aldehydes as acceptors, but the donor compound is often structurally invariable [2]. In addition, small organic amine-

catalyzed aldol addition reactions of different ketones with a variety of aldehydes have been performed [3]. By means of these strategies, different nucleobases containing aldehydes obtained by *N*-alkylation of the corresponding bases were employed to prepare acyclic nucleoside (AN) analogues [4,5]. This kind of compound shows relevant biological activities as antiviral agents; in particular, acyclic analogues of guanosine such as acyclovir, ganciclovir and penciclovir, and phosphonates such as adefovir and tenofovir were approved for therapies against herpes and retrovirus [6].

During our studies, we observed that the ¹³CNMR spectra of *N*-2-oxoethyl derivatives of nucleobases thymine and adenine, contain two sets of signals. It is known that many aldehydes, especially those containing electron-withdrawing substituents, are slowly transformed to gem-diols by water addition [7]. In fact, the stability of an aldehyde can be measured by how easily it becomes hydrated in aqueous solution [8]. Low molecular weight aliphatic aldehydes exist in equilibrium together with their hydrate forms, except for formaldehyde that is present mostly as its hydrate. Moreover, the equilibrium constant for aldehyde hydration can be modified by the presence of substituents. As an example, the substitution of acetaldehyde for acylamido groups enhances the

* Corresponding author.

E-mail address: sllaurella@quimica.unlp.edu.ar (S.L. Laurella).

hydration constant by one order of magnitude; these compounds are about 90% hydrated in dilute aqueous solution [9]. Proton magnetic resonance has great potential for direct measurements of the relative concentrations of both species. Signals for the hydrate and free aldehyde forms can be observed as separate species in the ^1H NMR spectrum, and the relative integration of the peaks provides the hydration equilibrium constant. The hydration constants of some aldehydes and ketones obtained by this strategy have been reported [10].

The aldehyde/hydrate equilibrium was shown to be temperature and solvent dependent. The thermodynamic values of the Van't Hoff plot showed a significant entropy term that nearly compensates the effect of enthalpy at room temperature: while bonding of two oxygen atoms to the same carbon is energetically more favorable than double bonding to a single oxygen, the entropy decreases because the number of product molecules is smaller than the number of reactant molecules [11], and this equilibrium is also pH-dependent. Regarding the compounds studied in this work, aldehyde-hydrate equilibrium was observed in pyridinecarboxaldehydes: neutral conditions favor the free aldehyde, showing a hydration constant significantly smaller than the unity, while in pyridinecarboxylic acids the hydrate is the major species [12].

As mentioned above, the NMR spectra of *N*-2-oxoethyl derivatives of nucleobases, thymine and adenine, showed two sets of signals. Further MS studies revealed the existence of both the aldehyde and the hydrate structures. However, we found some differences between purine and pyrimidine derivatives. Then, the main objective of this work was to analyze, explain and justify these differences.

2. Materials and methods

Synthesis of adenine- and thymine-type aldehydes. As described previously [4], a solution of adenine or thymine (1 equiv.) and K_2CO_3 (2 equiv.) in DMF (10 mL) was stirred at 90 °C in the presence of 2-bromo-1,1-dimethoxyethane (2 equiv.). After 24 h, the reaction was filtered and the solvent removed under reduced pressure. The crude mixture was purified by column chromatography in silica gel with $\text{MeOH}:\text{CH}_2\text{Cl}_2$ (adenine) or $\text{EtOAc}:\text{hexane}$ (thymine) as mobile phase. 9-(2,2-dimethoxyethyl)-adenine was obtained as an orangish yellow liquid in 61% isolated yield; and 1-(2,2-dimethoxyethyl)-thymine was obtained as a yellow liquid in 38% isolated yield. Aldehydes were obtained after hydrolysis of the corresponding acetals in HCl (1 N) at 90 °C for 1 h. In order to achieve the different pH values required for ESI-MS experiments, 1 mL imidazole buffer at the corresponding pH was then added to 1 mL of aldehyde reaction mixture.

NMR experiments were recorded on a Bruker Avance II 500 spectrophotometer (Madison, WI, USA) at 500 MHz for ^1H NMR experiments and 125.7 MHz for ^{13}C NMR experiments. Solutions were prepared by mixing 300 μL of reaction mixture (after hydrolysis with HCl), 200 μL of water and 250 μL of $\text{DMSO}-d_6$; resulting samples of concentration 50 mg/mL approximately.

The typical ^1H NMR spectral conditions were as follows: spectral width 3201 Hz, acquisition time 4.09 s and 8–16 scans per spectrum, digital resolution 0.39 Hz per point. Deuterium from the solvent was used as the lock and TMS as the internal standard. Water was suppressed from ^1H NMR spectra using presaturation method.

The spectral conditions for ^{13}C NMR spectra were the following: spectral width 10559 Hz, acquisition times 1.303 s and 512–1000 scans per spectrum, digital resolution was 1.29 Hz per point. The distortionless enhancement by polarization transfer (DEPT) spectra were measured under conditions that allow signal integration, as described in previous studies by Vlahov et al. [13].

EI-MS spectra were recorded on a ThermoFinnigan LCQ Advantage Max spectrometer (San Jose, CA, USA) by direct injection at room temperature (4 kV ion spray voltage, 350 °C capillary temperature), in positive mode, after solid phase extraction (SPE). The SPEs were performed using Phenomenex C18 silica gel cartridges (Torrance, CA, USA) and eluted with a mixture of $\text{MeOH}:\text{NH}_4\text{OH}$ 50 mM pH 10.8 (50/50).

Theoretical calculations were carried out following several steps. The molecules under study were first subjected to geometry optimizations using the DFT [14,15]. To this end, the B3LYP hybrid exchange–correlation functional [16,17] together with the 6-31G(d,p) basis set as implemented in the Gaussian 03 package [18] were used. All geometrical parameters were optimized without constraints. The minimum energy structures obtained from this first optimization were reoptimized using 6-311G++(d,p) basis. Thermodynamic and NBO analyses were carried out on these reoptimized structures at the same level of theory. Dielectric solvent effects were taken into account using the SCRF-PCM version of the polarization continuum model, PCM [19–21].

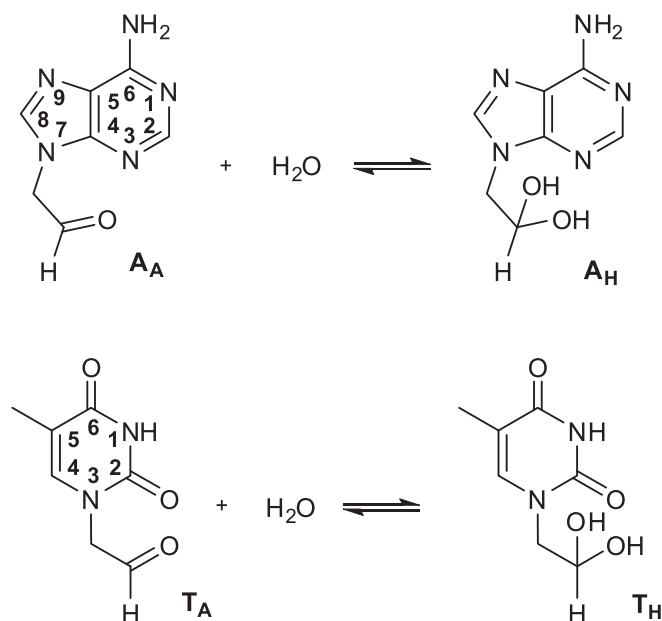
Isotropic chemical shifts for carbon atoms were also calculated. In this case, the isotropic magnetic shielding tensor was obtained at the B3LYP/6-31G(d,p) level of theory. The reported shifts are relative to tetramethylsilane (TMS). The absolute isotropic shieldings of TMS were also calculated using the B3LYP/6-31G(d,p) model.

3. Results and discussion

Scheme 1 shows the equilibrium between aldehyde and hydrate forms of the two *N*-2-oxoethyl nucleobase derivatives studied in this work: 2-(6-amino-9H-purin-9-yl)acetaldehyde (compound A) and 2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetaldehyde (compound T).

The hydrate and aldehyde forms of the adenine derivative will hereafter be named as A_H and A_A , and the hydrate and aldehyde forms of the thymine derivative as T_H and T_A . The heterocyclic atoms in A_A and T_A have been numbered for the sake of clarity.

The occurrence of these equilibria is evidenced in their ^1H NMR, ^{13}C NMR, DEPT and ESI-MS spectra from solutions of compounds A and T in water and DMSO-water.



Scheme 1.

3.1. Theoretical calculations

In principle, different conformations are expected depending on the mutual orientations of the nucleobase (adenine or thymine) and the 2-oxoethyl (in the aldehyde form) or 2,2-dihydroxyethyl groups (in the hydrate form). Thus, the computed potential energy curves around the C–N–C–C and N–C–C–O dihedral angles involved in these conformations were explored, and the structures of minimum energy drawn from these curves were reoptimized using the 6-311++G(d,p) basis set. Fig. 1 shows the structures obtained for the hydrate forms at this point.

At this first optimization step, it can be seen that both N3 in A_H and C=O2 in T_H are quite near OH groups (1.85 Å in A_H and 1.90 Å in T_H), so they are likely to interact with them.

In order to estimate the energy involved in the internal OH–Base hydrogen bond (and so understand the factors that cause a different equilibrium shift in adenine and thymine derivatives), we evaluated the donor→acceptor interactions using the natural bond orbital (NBO) population analysis [22] at the B3LYP/6-311++G(d,p) level of approximation. It was interesting to observe that a remote interaction of the type lp(N3)→σ*(O–H) occurs in the A_H form, while T_H appears to have a lp(O8)→σ*(O–H) interaction. This interaction is a clear indication that an O–H⋯N intramolecular hydrogen bond is established in the case of the hydrate of the adenine derivative A_H. In fact, for conventional O–H⋯N hydrogen bonds, a partial charge transfer from a lone pair electron formally located at the nitrogen atom toward the O–H antibonding orbital is expected to occur [23]. It has been shown that such hyperconjugative interactions can operate between remote orbitals, determining the conformational behavior of simple molecules including 1-acyl thioureas [24], thiazolopyrrolidin-2-ones [25] and hydrazones [26], as well as the backbone conformation of peptides [27].

The stabilization energies, E_{i,j}(2), associated with this hyperconjugative interaction are 16.76 and 12.18 kcal/mol for A_H and T_H, respectively, at the B3LYP/6-311++G(d,p) level of approximation. This difference explains (at least at first) the greater concentration of hydrate in the adenine derivative (see section 3.2).

Thermodynamic parameters corresponding to both aldehyde and hydrate equilibria proposed in Scheme 1 were also estimated from frequency calculations. All structures optimized in vacuum were further reoptimized in DMSO and water as continuum

dielectric solvents (PCM method). Theoretical enthalpies and entropies were calculated at this level of theory. The values obtained at this step (shown in Table 1) were not satisfactory, since ΔG⁰ values were ranged between +7 and +8 kcal/mol for both compounds, contrary to the equilibrium constants observed in the NMR spectra.

Table 2 resumes dipole moments of all molecules involved and their stabilization when they are surrounded by a continuum solvent (for sake of simplicity, water was chosen as solvent for these calculations). It can be seen that even when hydrate molecules are more stabilized than aldehyde molecules, the stabilization differences do not compensate the stabilization of water molecules in a polarized media (−9.82 kcal mol^{−1}).

The disagreement between theoretical and experimental data suggests that there are factors that are not being considered. The solvents DMSO (which is a hydrogen bond acceptor) and water (which is a hydrogen bond donor and acceptor) may be playing a key role in the equilibrium. It is important to remark at this point that the PCM model does not consider specific solute-solvent interactions (only mutual solute-solvent electrostatic polarization is taken into account in PCM calculations), as has been pointed out when studying similar equilibria [28]. In order to solve this inconsistency, an equilibrium involving internal hydrogen bonds in the hydrate form and an OH-solvent interaction is proposed (see Scheme 2, Sv can be either H₂O or DMSO). Therefore, explicit solvent molecules were added to the calculations (for simplicity, calculations were carried out considering only water molecules as solvent).

Water molecules were placed near the solute, and a geometry optimization calculation at B3LYP/6-31G(d,p) level in vacuum was carried out. From all studied configurations of the hydrate-water system, the most stable was chosen for thermodynamic calculations, as is shown in Fig. 2. It can be seen that water may interact with both hydroxyls of the hydrate molecule.

Then, with the optimized geometry of the solute-solvent complex, both calculations of single point and the PCM method were used at B3LYP/6-311++G(d,p) level of theory. The results are shown in Table 3.

These ΔG⁰ values clearly support the equilibrium constants observed in NMR spectra (see section 3.2) and the assumption that these compounds establish internal hydrogen bonds in the hydrate forms and that the solvents (DMSO and water) play an important

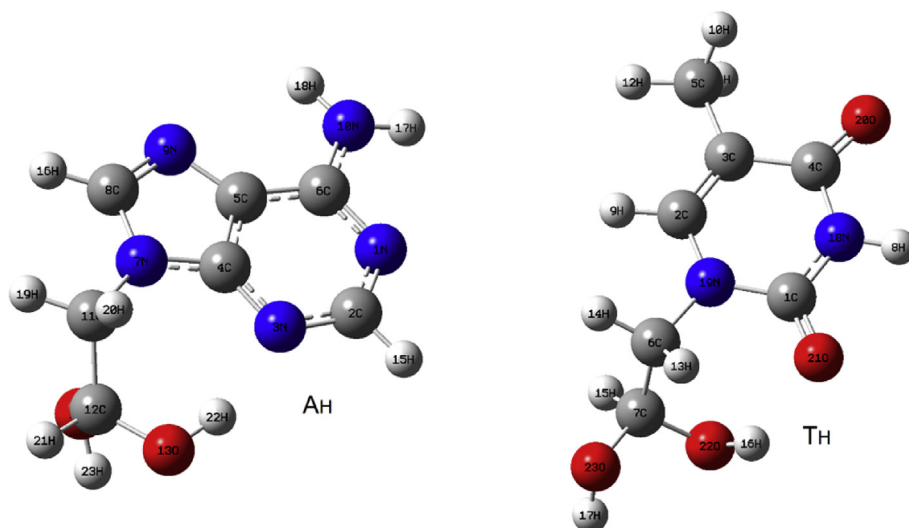


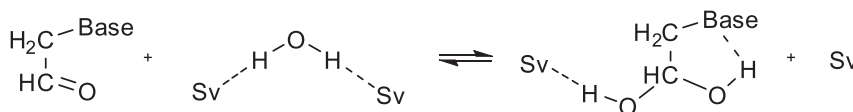
Fig. 1. Optimized structures at B3LYP/6-311++G(d,p) theory level for the hydrate forms of compounds A and T.

Table 1
Aldehyde-hydrate equilibrium thermodynamic parameters for adenine and thymine derivatives.

		Compound	
		Adenine derivative (T)	Thymine derivative (A)
Vacuum	ΔH (kcal/mol)	-8.9	-7.3
	ΔS (cal/molK)	-44.2	-42.7
	ΔG^0 (kcal/mol)	+4.3	+5.4
DMSO (PCM model)	ΔH (kcal/mol)	-5.3	-4.6
	ΔS (cal/molK)	-42.7	-39.0
	ΔG^0 (kcal/mol)	+7.4	+7.0
H ₂ O (PCM model)	ΔH (kcal/mol)	-6.0	-4.7
	ΔS (cal/molK)	-49.4	-42.8
	ΔG^0 (kcal/mol)	+8.7	+8.1

Table 2
Dipole moment in vacuum and in solution and (polarized solute)-solvent stabilization in water (PCM) for adenine and thymine hydrates and aldehydes.

Compound	Dipole moment (Debye)		(polarized solute)-solvent stabilization PCM, in kcal mol ⁻¹
	Vacuum	Water	
Adenine hydrate (A _H)	2.00	3.27	-23.45
Adenine aldehyde (A _A)	2.26	3.73	-19.79
Thymine hydrate (T _H)	3.99	6.91	-29.97
Thymine aldehyde (T _A)	2.50	4.34	-22.40
Water	2.15	2.56	-9.82



Scheme 2.

role in the equilibrium shift, resulting in very high concentrations of the hydrate forms (A_H and T_H) in these equilibria.

The ΔG^0 values calculated by using the PCM model and explicit solvent molecules (H₂O in this case) would lead to equilibrium constants *K* of 9.0 for compound A and 5.4 for compound T, which are very close to the experimental values and support the

assumptions made regarding the solvation model.

3.2. NMR evidence of aldehyde-hydrate equilibrium

Figs. 3 and 4 show the ¹³CNMR, DEPT-90 and DEPT-135 spectra of compounds A and T in water: DMSO-*d*₆ 2:1 at 25 °C, respectively. Each spectrum is the result of the overlapping of the aldehyde and hydrate spectra, since the equilibrium is sufficiently slow (at least at this temperature) to produce separate signals.

The signal assignment was performed by means of theoretical calculations. Shielding tensors corresponding to each carbon were calculated using the B3LYP functional at a 6-311G++(2d,p) level of theory over structures optimized using the 6-31G(d) basis set. Then, each shielding tensor can be transformed into a chemical shift either by subtracting the shielding tensor corresponding to TMS (at the same level of theory) or using scaling factors [29]. The last method was chosen since it has proven more accurate for chemical shift estimation. According to this method, shielding tensors (σ) and chemical shifts (δ) are linearly related by the linear equation $\delta = (I - \sigma)/S$, where *I* and *S* values are experimental and they are listed in literature [30]. Under the experimental and calculation conditions given in this work, their values are *I* = 182.0245 and *S* = 1.0290.

The only aromatic peaks that can be differentiated from the hydrate and aldehyde forms are those corresponding to the CH carbons adjacent to the substituted N atom in each nucleobase (146.7 and 146.6 ppm in C8 of compound A, and 145.0 and 145.1 in C4 of compound T). Moreover, both compounds give observable CH₂ signals (separated by approximately 1.4 ppm from one another), and the CH(OH)₂ of the hydrate forms A_H and T_H appear as peaks at 88.7 and 88.60 ppm, respectively. Aldehyde carbons (which usually show very low intensities in ¹³CNMR) are not

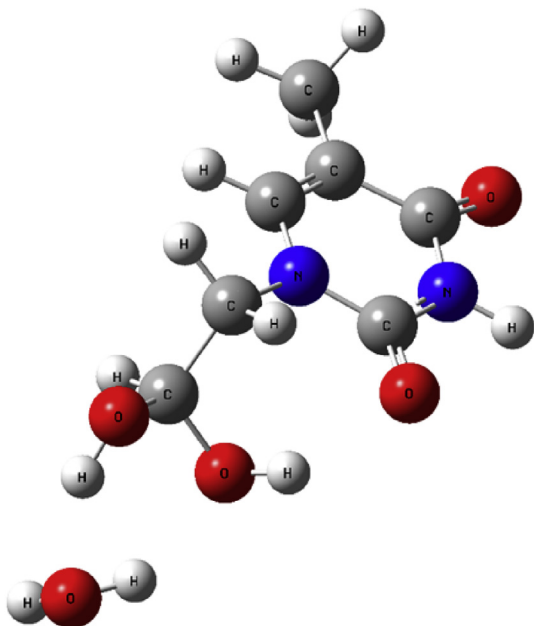
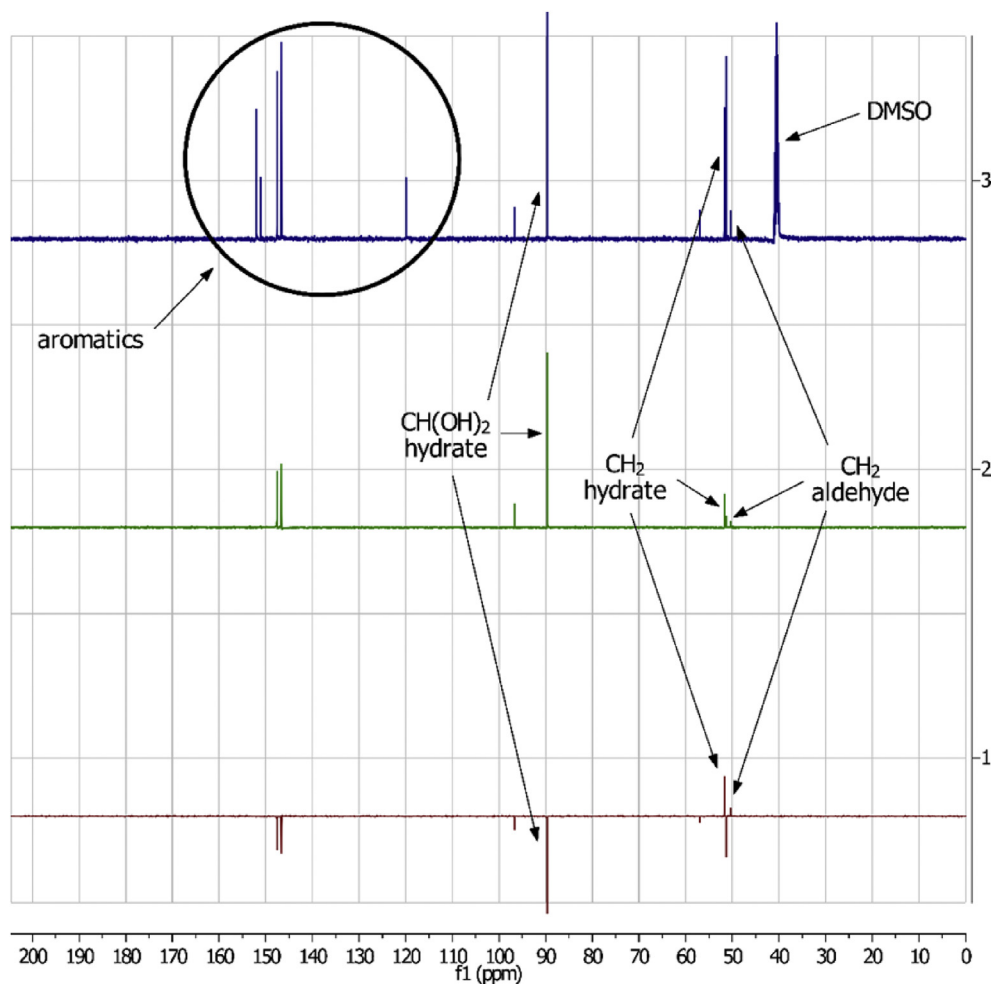


Fig. 2. Optimized structure for T_H-H₂O complex at B3LYP/6-31G(d,p) theory level.

Table 3Thermodynamic parameters for adenine and thymine derivatives according to the equilibrium proposed in Scheme 2 in vacuum and in continuum H₂O solvent.

Compound	Vacuum			H ₂ O (PCM model)		
	ΔH (kcal/mol)	ΔS (cal/molK)	ΔG^0 (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/molK)	ΔG^0 (kcal/mol)
Adenine derivative (A)	-4.5	-19.1	+1.2	-4.3	-10.2	-1.3
Thymine derivative (T)	-3.1	-16.5	+1.8	-3.5	-8.5	-1.0

**Fig. 3.** ¹³CNMR (3), DEPT-90 (2) and DEPT-135 (1) spectra of compound A in water:DMSO-*d*₆ 2:1 at 25 °C.

observable in any of the two spectra, which is supported by the assumption that aldehyde forms are present as minor components in both equilibria.

The experimental and calculated δ values are listed in Table 4.

Even though ¹³CNMR signals cannot be integrated in order to compare the relative quantity of carbons (due to different relaxation times and NOE effect), DEPT spectra can be integrated under certain experimental conditions given in literature [13,31] that allow signal integration and quantitative studies. Then, the equilibrium constant can be obtained from the integration of the CH₂ signals (which appear clearly separated from one another in the DEPT-135 spectrum) of the aldehyde and hydrate form. Integration of the selected signals allows us to define an operational equilibrium constant:

$$K = \frac{[\text{hydrate}]}{[\text{aldehyde}]} = \frac{\text{CH}_2 \text{ integration in hydrate form}}{\text{CH}_2 \text{ integration in aldehyde form}}$$

This ratio gives K constants of 8.3 in A and 5.3 in T, which indicate that in both equilibria the hydrate form is the main species in equilibrium, having a higher concentration in the case of the adenine derivative.

The ¹HNMR spectra of A and T (Figs. 5 and 6) also show the presence of aldehyde and hydrate forms, being the last one more abundant.

In Fig. 5 (¹HNMR spectrum of A at 25 °C), the triplet at 5.37 ppm and the doublet at 4.35 ppm correspond to the CH₂–CH(OH)₂ system in the hydrate A_H. Aromatic protons attached to C2 and C8 in A_H show two signals at 8.37 and 8.44 ppm, respectively. Next to the last one, at 8.36 ppm, a small signal is observed and it corresponds to the proton attached to C8 in the aldehyde form A_A. The integral ratio of the signals at 8.37 ppm (hydrate A_H) and 8.36 ppm (aldehyde A_A) renders a K value of 8.5, which is in accordance with the value of 8.3 obtained from the DEPT-135 spectrum of A. Aldehyde proton and CH₂ protons from the A_A form appear as low

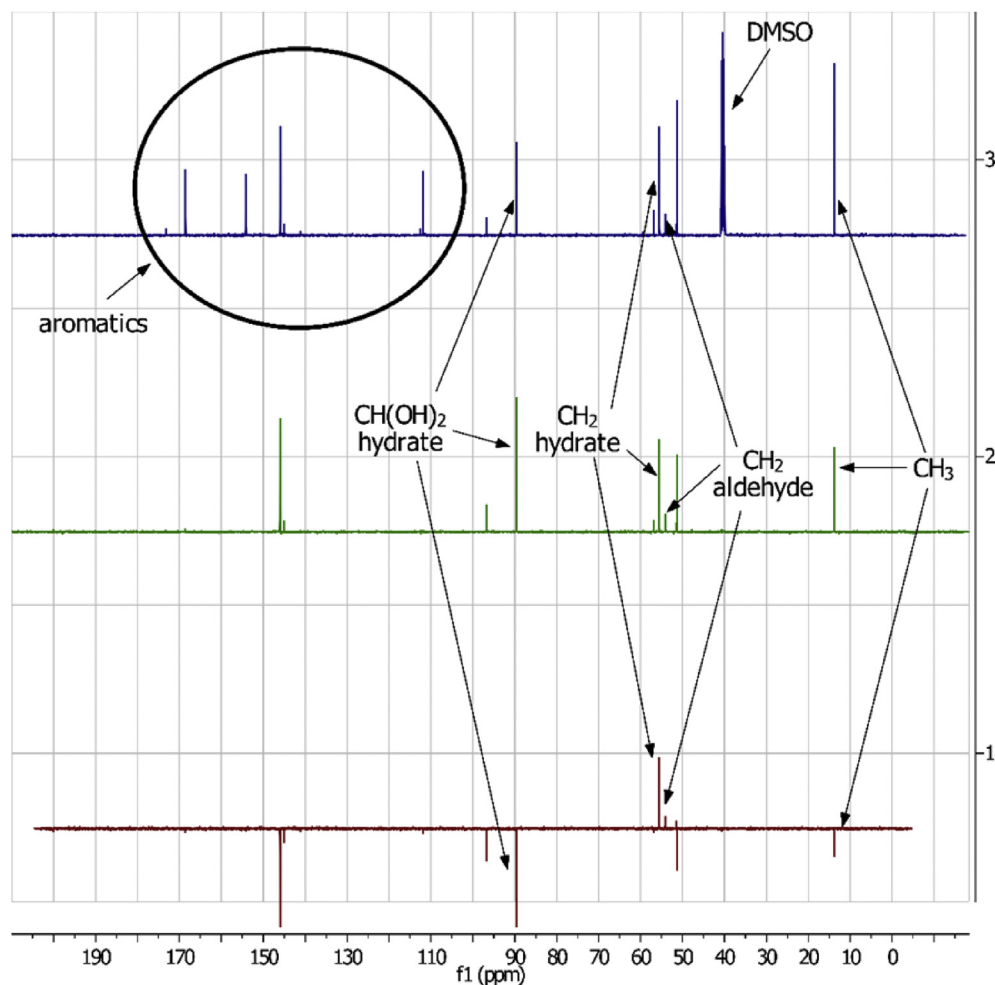


Fig. 4. ^{13}C NMR (3), DEPT-90 (2) and DEPT-135 (1) spectra of compound T in water:DMSO- d_6 2:1 at 25 °C.

intensity triplets at 9.75 and 5.01 ppm respectively.

Analogously, in Fig. 6 (^1H NMR spectrum of T at 25 °C), the triplet at 5.17 ppm and the doublet at 3.74 ppm correspond to the $\text{CH}_2\text{--CH}(\text{OH})_2$ system in the hydrate T_{H} . The aromatic proton attached to C4 in T_{H} show a signal at 7.44, next to which, at 7.43 ppm, a small signal is observed and it corresponds to the proton attached to C4 in the aldehyde form T_{A} . The integral ratio of

the signals at 7.44 ppm (hydrate T_{H}) and 7.43 ppm (aldehyde T_{A}) renders a K value of 5.2, which is in accordance with the value of 5.3 obtained from the DEPT-135 spectrum of T. Aldehyde proton and CH_2 protons from the T_{A} form appear as low intensity triplets at 9.60 and 4.81 ppm respectively. The methyl protons appear as separated signals in the spectrum (1.83 ppm for T_{H} and 1.85 ppm for T_{A}).

Table 4

Experimental and calculated chemical shifts (in ppm) for A_{A} , A_{H} , T_{A} and T_{H} molecules.^a

Type of C	Compound A				Compound T					
	A_{A} form		A_{H} form		T_{A} form		T_{H} form			
	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.		
CH_2	49.3	51.1	50.7	51.8	53.2	51.8	54.6	55.9		
CHO	n.o. ^b	201.3	—	—	n.o.	202.6	—	—		
$\text{CH}(\text{OH})_2$	—	—	88.7	89.9	—	—	88.6	91.3		
Aromatics	C5	118.9	117.6	119.0	118.6	C5	111.1	112.7	111.0	111.4
	C2	145.7	138.7	145.7	140.6	C4	145.0	139.2	145.1	140.5
	C8	146.7	148.1	146.6	147.8	C2	153.2	148.8	153.3	151.4
	C4	150.1	151.7	150.1	149.6	C6	167.6	160.4	167.6	160.3
	C6	151.1	151.8	151.1	152.0					
CH_3	—	—	—	—	12.9	12.4	12.9	12.4		

^a The signal at 50 ppm corresponds to methanol, the by-product of the deprotection reaction.

^b n.o.: not observable.

3.3. ESI-MS evidence of aldehyde-hydrate equilibrium

Figs. 7 and 8 show the ESI spectrum (positive mode) obtained for A and T. It are two important to remarks to be made at this point. First, even when the sample is exposed to high vaporization temperature (350 °C) in vacuum, it does not decompose (or if it does, the decomposition ratio is negligible) because the vaporization process is very fast.

Second, vaporization is so fast that it can be assumed that the equilibrium does not shift when passing from solution to gas phase. In other words: the molecules that arrive to the detector follow the same proportion that they had in solution (before vaporization).

In Fig. 7, the peak at m/z 178 corresponds to the protonated aldehyde of the adenine derivative ($A_{\text{A}} + \text{H}$)⁺, while the peak at m/z 196 is given by its protonated hydrate form ($A_{\text{H}} + \text{H}$)⁺. Likewise, m/z 169 and m/z 187 peaks in Fig. 8 are due to the presence of protonated aldehyde and hydrate forms of the thymine derivative ($T + \text{H}$)⁺.

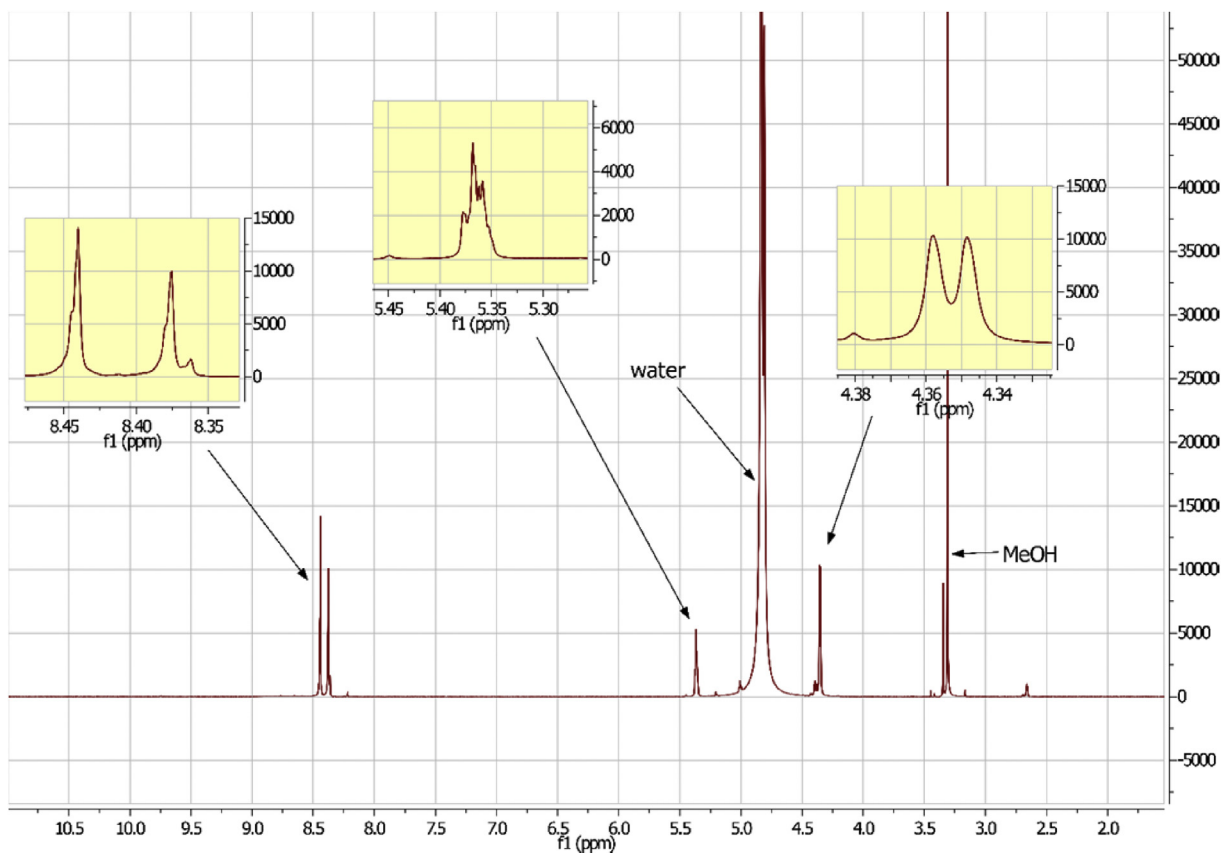


Fig. 5. ¹H NMR spectrum of compound A in water:DMSO-*d*₆ 2:1 at 25 °C.

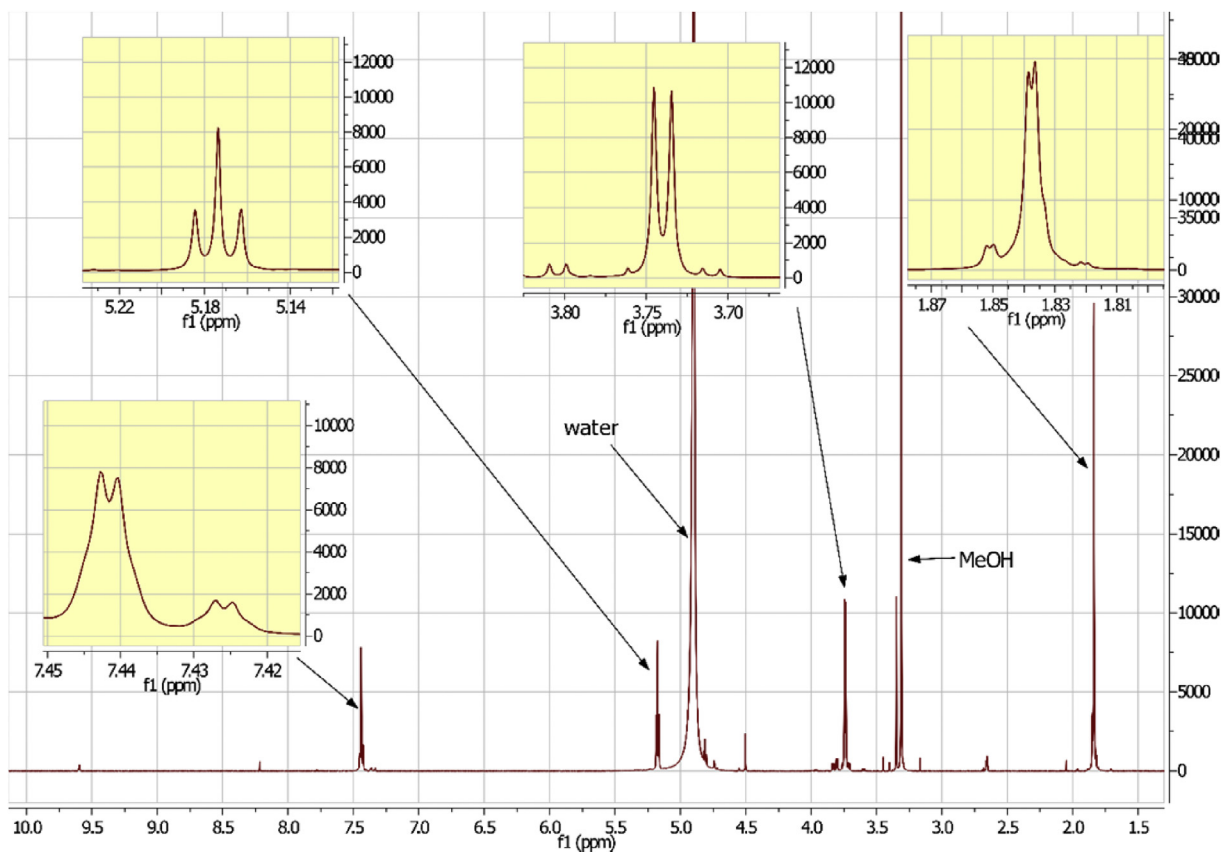


Fig. 6. ¹H NMR spectrum of compound T in water:DMSO-*d*₆ 2:1 at 25 °C.

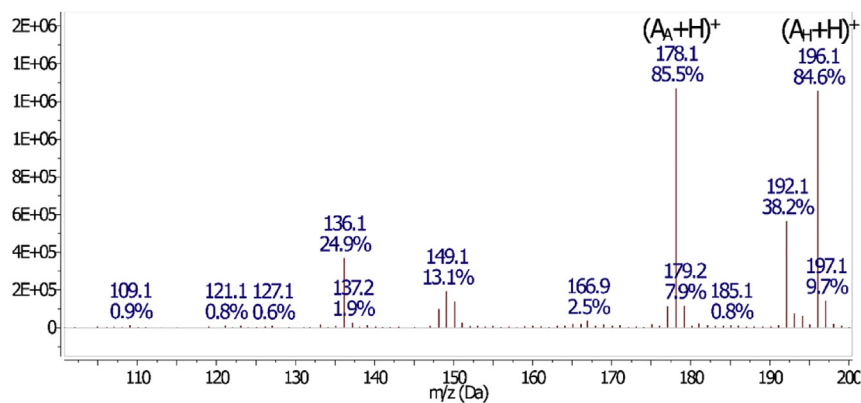


Fig. 7. ESI-MS spectrum of compound A obtained at pH 7.0.

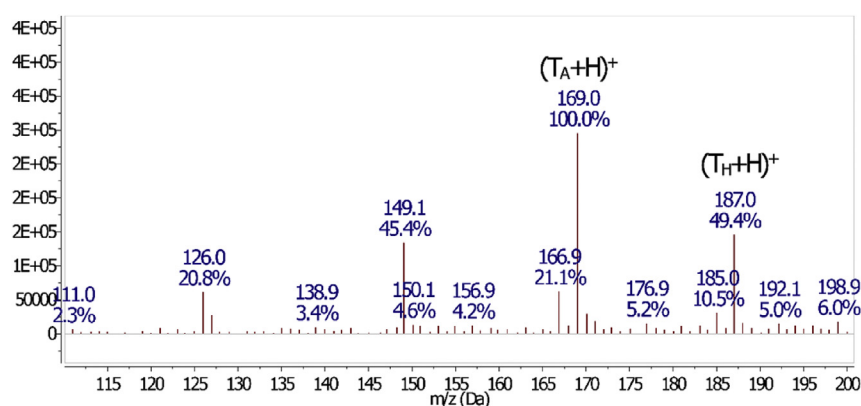


Fig. 8. ESI-MS spectrum of compound T obtained at pH 7.0.

Table 5

Hydrate-aldehyde abundance ratio for compounds A and T at different pH values.

Compound	Ratio	pH			
		1.0	7.0	7.5	8.0
Adenine derivative (A)	$Ab_{AH}/Ab_{AA} (m/z196)/(m/z178)$	0.49	1.0	1.1	1.1
Thymine derivative (T)	$Ab_{TH}/Ab_{TA} (m/z187)/(m/z169)$	0.046	0.49	0.57	0.78

If Ab_{XA} and Ab_{XH} (where X can be either A or T) are the abundances of the MS peaks corresponding to the protonated aldehyde and hydrate forms, respectively, then the ratio Ab_{XH}/Ab_{XA} will allow us to compare the equilibria of both A and T compounds.

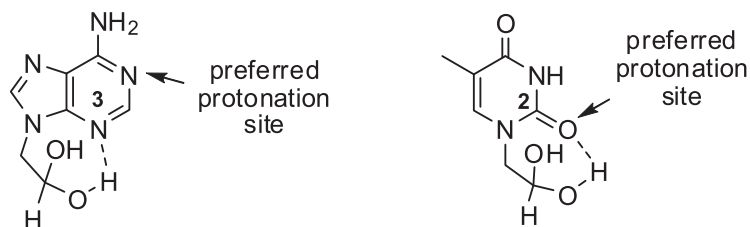
It is important to stress that this ratio is *not a real equilibrium constant* (since the abundances depend on many factors), but it permits the comparison between both substances, as has been done with other families of compounds [32,33].

The Ab_{XH}/Ab_{XA} ratios evaluated at several pH values for both compounds are listed in Table 5.

Two main observations can be drawn from Table 5:

- The Ab_{AH}/Ab_{AA} ratio is larger than Ab_{TH}/Ab_{TA} at all pH values between 1 and 8.
- The Ab_{XH}/Ab_{XA} ratio increases with pH, i.e., the equilibrium is shifted towards the aldehyde form when $[H^+]$ increases.

These two facts can be explained supposing that A_H and T_H are stabilized via internal H bonding, between OH and N3 in A_H and between OH and C=O 2 in T_H (Scheme 3).



Scheme 3.

If such interaction exists, an increase in $[H^+]$ may weaken these interactions, which seem to be greater in A_H than in T_H . In fact, the protonations of compounds A and T are quite different. When the pH lowers, compound A is mainly protonated on the heterocyclic N1 and compound T is protonated on the ureic C=O in position 2 [34]. In the case of compound A, the internal hydrogen bond would not be much affected by protonation, while in compound T protonation would break such bond.

The assumption about the existence of internal hydrogen bonds in these compounds (and the role they play in stabilization and equilibrium shifts) is supported by theoretical calculations in the previous section (NBO analysis).

4. Conclusions

The analysis of 1H NMR, ^{13}C NMR, DEPT and ESI-MS spectra of the two *N*-2-oxoethyl derivatives under study lets us conclude that they exist mainly as two different forms, aldehyde and hydrate. The different NMR spectra show that the equilibrium is shifted towards the hydrate form in water-DMSO 2:1, giving equilibrium constants of 8.3 and 5.3 for adenine and thymine derivatives, respectively. ESI-MS experiments show the dependence of equilibrium shift on pH in aqueous solutions: in the case of thymine derivative, the equilibrium shows a greater shift than in the case of adenine derivative, which is explained considering different protonation sites. All assumptions are supported by theoretical calculations, which suggest the important role played by solvent in the stabilization of molecular structures and equilibrium shift. All aspects analyzed in this work are very important in order to understand the further reactivity of these nucleobase derivatives in aqueous solutions.

Acknowledgements

The authors acknowledge Universidad Nacional de Quilmes, Facultad de Ciencias Exactas (Universidad Nacional de La Plata) and Agencia Nacional de Promoción Científica y Tecnológica, República Argentina, for financial support. AMI and ESL are members of the Scientific Researcher Career of Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Argentina.

References

- [1] C.A. Busacca, D.R. Frandrick, J.J. Song, C.H. Senanayake, The growing impact of catalysis in the pharmaceutical industry, *Adv. Synth. Catal.* 353 (2011) 1825–1864.
- [2] P. Clapés, W.-D. Fessner, G.A. Sprenger, A.K. Samland, Recent progress in stereoselective synthesis with aldolases, *Curr. Opin. Chem. Biol.* 14 (2010) 154–167.
- [3] V. Bisai, A. Bisai, S. Singh, Enantioselective Organo-catalytic aldol reaction using small organic molecules, *Tetrahedron* 68 (2012) 4541–4580.
- [4] M. Palazzolo, M. Nigro, A.M. Iribarren, E. Lewkowicz, A chemoenzymatic route to prepare acyclic nucleoside analogues, *Eur. J. Org. Chem.* 5 (2016) 921–924.
- [5] M. Palazzolo, M. Pérez-Sánchez, A.M. Iribarren, E. Lewkowicz, P. Domínguez de María, Organocatalytic synthesis of novel purine and pyrimidine acyclic nucleosides, *Tetrahedron Lett.* 53 (2012) 6797–6800.
- [6] E. De Clercq, H.J. Field, Antiviral prodrugs - the development of successful prodrug strategies for antiviral chemotherapy, *Br. J. Pharmacol.* 147 (2006) 1–11.
- [7] R.P. Bell, The reversible hydration of carbonyl compounds, *Adv. Phys. Org. Chem.* 4 (1966) 1–29.
- [8] Z.-Q. Xu, Y.-L. Qiu, S. Chokekijchai, H. Mitsuya, J. Zemlicka, Unsaturated acyclic analogs of 2'-deoxyadenosine and thymidine containing fluorine: synthesis and biological activity, *J. Med. Chem.* 38 (1995) 875–882.
- [9] R.A. McClelland, P. Sukhai, K.M. Engell, P.E. Sorensen, Hydration equilibria of 9-acridinecarboxaldehyde, *Can. J. Chem.* 72 (1994) 2333–2338.
- [10] M.T. Doel, A.S. Jones, N. Taylor, An approach to the synthesis of peptide analogues of oligonucleotides (nucleopeptides), *Tetrahedron Lett.* 10 (1969) 2285–2288.
- [11] P. Charisiadis, C.G. Tsiafoulis, A.G. Tzakos, P. Gerathanassis, Dynamic changes in composition of extracts of natural products as monitored by in situ NMR, *Magn. Reson. Chem.* 52 (2014) 764–768.
- [12] S. Huang, A.K. Miller, W. Wu, Substantial formation of hydrates and hemiacetals from pyridinium ketones, *Tetrahedron Lett.* 50 (2009) 6584–6585.
- [13] G. Vlahov, C. Schiavone, N. Simone, Quantitative ^{13}C NMR method using the DEPT pulse sequence for the determination of the geographical origin (DOP) of olive oils, *Magn. Reson. Chem.* 39 (2001) 689–695.
- [14] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* 140 (1965) 1133–1138.
- [15] R.G. Parr, W. Yang, *Density-functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
- [16] D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* 98 (1993) 5648–5652.
- [17] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* 37 (1988) 785–789.
- [18] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.B. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K. Kudin, J. Burant, J. Millam, S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. Petersson, H. Nakatsuji, M. Hada, M. Ehara, R. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. Knox, H. Hratchian, J. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. Stratmann, A. Yazyev, R. Austin, C. Cammi, J.W. Pomelli, P.Y. Ochterski, K. Ayala, G.A. Morokuma, P. Voth, J.J. Salvador, V.G. Dannenberg, S. Zakrzewski, A.D. Dapprich, M.C. Daniels, O. Strain, D.K. Farkas, A.D. Malick, K. Rabuck, J.B. Raghavachari, J.V. Foresman, Q. Ortiz, A.G. Cui, S. Baboul, J. Clifford, B.B. Cioslowski, G. Stefanov, A. Liu, P. Liashenko, I. Piskorz, R.L. Komaromi, D.J. Martin, T. Fox, M.A. Keith, C.Y. Al-Laham, A. Peng, M. Nanayakkara, P.M.W. Challacombe, B. Gill, W. Johnson, M.W. Chen, C. Wong, R. Gonzalez, J.A. Pople, Gaussian 03, Revision B.03, Gaussian, Inc., Pittsburgh PA, 2003.
- [19] M.T. Cancs, B. Mennucci, J. Tomasi, A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics, *J. Chem. Phys.* 107 (1997) 3032–3041.
- [20] M. Cossi, V. Barone, B. Mennucci, J. Tomasi, Ab initio study of ionic solutions by a polarizable continuum dielectric model, *Chem. Phys. Lett.* 286 (1998) 253–260.
- [21] B. Mennucci, J. Tomasi, Continuum solvation models: a new approach to the problem of solute's charge distribution and cavity boundaries, *J. Chem. Phys.* 106 (1997) 5151–5158.
- [22] E. Reed, L.A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, *Chem. Rev.* 88 (1988) 899–926.
- [23] V. Alabugin, M. Manoharan, S. Peabody, F. Weinhold, Electronic basis of improper hydrogen bonding: a subtle balance of hyperconjugation and rehybridization, *J. Am. Chem. Soc.* 125 (2003) 5973–5987.
- [24] A. Saeed, A. Khurshid, J.P. Jasinski, C.G. Pozzi, A.C. Fantoni, M.F. Erben, Competing intramolecular N–H...O=C hydrogen bonds and extended intermolecular network in 1-(4-chlorobenzoyl)-3-(2-methyl-4-oxopentane-2-yl) thiourea analyzed by experimental and theoretical methods, *Chem. Phys.* 431–432 (2014) 39–46.
- [25] S.L. Laurella, M.F. Erben, Comment on "Structural and vibrational studies on 1-(5-Methyl- [1,3,4] thiadiazol-2-yl)-pyrrolidin-2-ol" [*Spectrochimica Acta Part A*, 152 (2016) 252–261], The importance of intramolecular OH...N hydrogen bonding in the conformational properties of thiadiazol-pyrrolidin-2-ol bearing species, *Spectrochim. Acta A* 164 (2016) 40–42.
- [26] A. Saeed, M. Ifzan Arshad, M. Bolte, A.C. Fantoni, Z.Y. Delgado Espinoza, M.F. Erben, On the roles of close shell interactions in the structure of acyl-substituted hydrazones: an experimental and theoretical approach, *Spectrochim. Acta A* 157 (2016) 138–145.
- [27] J.J. Koivisto, E.T.T. Kumpulainen, A.M.P. Koskinen, Conformational ensembles of flexible β -turn mimetics in DMSO-*d*₆, *Org. Biomol. Chem.* 8 (2010) 2103–2116.
- [28] D.L. Ruiz, A.G. Albesa, A. Ponzinibbio, P.E. Allegretti, M.M. Schiavoni, Solvent effects on tautomeric equilibria in β -ketonitriles: NMR and theoretical studies, *J. Phys. Org. Chem.* 23 (2010) 985–994.
- [29] M.W. Lodewyk, M.R. Siebert, D.J. Tantillo, Computational prediction of 1H and ^{13}C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry, *Chem. Rev.* 112 (2012) 1839–1862.
- [30] M.W. Lodewyk, C. Soldi, P.B. Jones, M.M. Olmstead, J.R. Larrucea, J.T. Shaw, D.J. Tantillo, The correct structure of aquatolide - experimental validation of a theoretically-predicted structural revision, *J. Am. Chem. Soc.* 134 (2012) 18550–18553.
- [31] Z. Xia, L.G. Akim, D.S. Argyropoulos, Quantitative ^{13}C NMR analysis of lignins with internal standards, *J. Agric. Food Chem.* 49 (2001) 3573–3578.
- [32] V.B. Di Marco, G.G. Bombi, Electrospray mass spectrometry (ESI-MS) in the study of metal-ligand solution equilibria, *Mass. Spectrom. Rev.* 25 (2006) 347–379.
- [33] F. Rosu, V. Gabelica, C. Houssier, E. De Pauw, Determination of affinity, stoichiometry and sequence selectivity of minor groove binder complexes with double-stranded oligodeoxynucleotides by electrospray ionization mass spectrometry, *Nucleic Acids Res.* 30 (2002) e82.
- [34] V. Verdolino, R. Cammi, B.H. Munk, H.B. Schlegel, Calculation of pK_a values of nucleobases and the guanine oxidation products guanidinohydantoin and spiroiminodihydantoin using density functional theory and a polarizable continuum model, *J. Phys. Chem. B* 112 (2008) 16860–16873.