Tautomerism of the Natural Thymidine Nucleoside and in the Antiviral Analogue D4T. Structure and influence of an Aqueous Environment Using MP2 and DFT Methods

ARTICLE *in* PHYSICAL CHEMISTRY CHEMICAL PHYSICS · JANUARY 2010

Impact Factor: 4.49 · DOI: 10.1039/b915566j · Source: PubMed

2 AUTHORS, INCLUDING:

Mauricio Alcolea Palafox
Complutense University of Madrid
129 PUBLICATIONS **1,440** CITATIONS

SEE PROFILE

Tautomerism of the natural thymidine nucleoside and the antiviral analogue D4T. Structure and influence of an aqueous environment using MP2 and DFT methods†

M. Alcolea Palafox* and N. Iza

Received 30th July 2009, Accepted 28th October 2009
First published as an Advance Article on the web 4th December 2009
DOI: 10.1039/b915566j

A comparative theoretical conformational analysis of the tautomerism of the four most stable conformers of the antiviral analogue D4T (stavudine) and natural thymidine (Thy) nucleosides was carried out, by using the B3LYP and MP2 quantum chemical methods. The calculated structure data and energy values of the tautomers were compared with the T1 keto form and with the results determined for the thymine molecule. For each conformer, only two stable enol forms, T3 and T5, were obtained when considering different positions of hydrogen around the base. Tautomer T3 always appears more stable than T5, and more stable in the natural nucleoside Thy than in thymine or in D4T.

The effect of water on the tautomers was estimated by using an explicit number of up to five water molecules to surround the nucleoside, and also by Tomasi's polarized continuum model (PCM). A total of about 200 clusters were optimised and the geometrical parameters and energies discussed. The water net differs in the three tautomers of D4T and Thy. Depending on the nature of the tautomers, cyclic, distributed water molecules, or clustered structures are formed. The deformation and interaction counterpoise (CP)-corrected energies between the nucleoside and water molecules were determined. The relative stabilities of all tautomers were established. The microhydrated environment stabilizes remarkably the enol forms more than the canonical keto one, although this one continues being the most stable. The changes in the intramolecular H-bondings and in the total atomic charges were discussed. Intramolecular H-bonds being stronger in D4T than in Thy could indicate a higher molecular flexibility for Thy.

1. Introduction

Uracil and thymine exist in the diketo form in the isolated state as well as in aqueous solution, but there is also experimental evidence of small amounts of tautomeric forms.[1–5] Generally, the keto form exists as the main form in the double helix.[6] The formation of specific Watson–Crick pairs is responsible for the maintenance of the genetic code and if uracil is replaced by its tautomer, it may lead to the introduction of a wrong genetic code. Tautomeric pairs A–T are significantly less stable than the Watson–Crick pair consisting of the canonical forms.[7] However, in the G–C pair, the energy differences are small and a very small but non-negligible amount of tautomeric pairs can appear in the DNA double helix. That is, some of the tautomeric forms may cause base mispairs,[8] which has been proven to be one of the origins of gene mutation.[4,9–12]

A large amount of work has been performed on the tautomerism of nucleic acid bases, using both theoretical[3b,13–16] and experimental[17–20] approaches. The process is reported to be intimately connected with the energetics of the chemical bonds. In the present paper, a detailed theoretical study of the tautomerism of nucleoside analogue D4T (2′,3′-didehydro-3′-deoxythymidine) and natural nucleoside Thy was carried out. Previously, a full theoretical conformational analysis[21a] and hydration[21b] of these molecules was performed.

Nucleoside analogues play a crucial role in the current treatment of cancer and viral infections.[22–24] Compounds such as D4T belong to the most effective alternative substrates of the reverse transcriptase enzyme of the HIV virus, named NRTIs (nucleoside reverse transcriptase inhibitors), antiviral agents that also have a wide range of other biological activities, such as antitumour and antibiotic agents.[25,26] NRTIs are applied alone or in combination with other medications used to treat HIV in clinical practice. These nucleoside analogues have no 2′ and 3′-hydroxyl groups; therefore they inhibit the further growth of a viral DNA chain.

Previous studies have reported that anti-HIV-1 activity depends upon the ribose conformer.[24] Differences in the conformation of the ribose ring lead to appreciable changes in the positions of the thymine ring and the C5′–OH group.[25] From our understanding, it would be interesting to analyse the tautomerism in the different conformational possibilities for D4T. To our knowledge these compounds have been extensively examined for antiviral properties, but very little structural and energetic information obtained from theoretical studies and on their tautomers are available.[27–31]

Departamento de Química-Física I, Facultad de Ciencias Químicas, Universidad Complutense, Ciudad Universitaria, Madrid-28040, Spain. E-mail: alcolea@quim.ucm.es

† Electronic supplementary information (ESI) available: Computational details. See DOI: 10.1039/b915566j

2. Calculations

The calculations were carried out by using the MP2 *ab initio* method, and also by using the B3LYP density functional methods (DFT). Both methods are implemented in the Gaussian 03 program package.[32] The UNIX version with the standard parameters of this package was run on the alpha computer of the Computational Center of University Complutense of Madrid, Spain. Several basis sets were used, starting from 6-31G** to 6-311 ++ G(3df,pd), but the 6-31G** represents a compromise between accuracy and computational cost, and thus it was the base set selected for all the calculations.

DFT methods provide a compromise between the desired chemical accuracy and the heavy demands put on computer time and power. Becke's three-parameter exchange functional (B3) in combination with the correlational functional of Lee, Yang and Parr (LYP), the B3LYP method, appears the most adequate. DFT methods, in particular B3LYP, have been satisfactorily used in many studies of tautomerism,[33b,c] drug design,[33] nucleosides[21] and uracil.[34] Moreover, this method was chosen because the data obtained appeared in general to be in good agreement with those obtained by other more computationally costly methods, such as MP2, and it predicts vibrational wavenumbers of DNA bases better than the HF and MP2 methods.[34–38] Where the computer memory requirements could be met, MP2 computations were also performed on all the structures to confirm or correct the B3LYP results.

Atomic charges were determined with the NBO (natural bond orbital)[39,40] procedure. Wavenumber calculations were performed in all the optimised clusters determined by B3LYP, to assess whether or not they show positive harmonic vibrations only, corresponding to real minima (true energy minima). Relative energies were obtained by including zero-point vibrational energies (ZPE). For the calculation of the ZPE, the wavenumbers were retained unscaled.

2.1 Hydration methods

To theoretically simulate the hydration effects, three procedures have been reported:[41] (1) empirical scaling of the quantum mechanical force constants of the isolated molecule,[42] (2) use of the continuum model[43] or (3) the discrete model (DM) by including a sufficient number of explicit water molecules.[44]

Tomasi's polarized continuum model (PCM)[45] was used, as it is implemented in Gaussian 03 by default (using the integral equation formalism model, IEF-PCM).[46] With this model, the number of conformers corresponding to real minima is remarkably reduced, compared to those found in the isolated state. This model is the most widely used today, mainly owing to its simplicity and the lower computational time required.

The DM provides a thorough description of the water molecule net and specific solute–solvent interactions. A select number of solvent configurations is included in the description of the system. Unfortunately, both the large number of solvent molecules required to mimic a dilute solution and the computational cost of the calculations[47] limit the applicability of this model. However, from a strictly theoretical point of view, it is the preferred method and it was used in the present paper, provided that the size of the considered cluster is not too large. Moreover, when it is used with several water molecules, since it gives an account of hydrogen bonds explicitly, the first hydration shell can be better reproduced than by using the continuum model. The methodology applied to the hydration in the present manuscript has been previously reported in detail.[48]

2.2 Interaction energies

The energies obtained from the hydration of D4T with explicit water molecules were corrected for basis set superposition error (BSSE) using the counterpoise (CP) procedure from Danilov *et al.*[49] The CP-corrected B-$(H_2O)_n$ formation energy is calculated as:

$$\Delta E^{\mathrm{CP}}_{\mathrm{BW}_n} = E^{(\mathrm{BW}_n)}(\mathrm{BW}_n) + E^{(\mathrm{def})}(\mathrm{BW}_n) \qquad (1)$$

where B ≡ D4T, W_n ≡ water *n*mer and $E^{(\mathrm{BW}_n)}(\mathrm{BW}_n)$ and $E^{(\mathrm{def})}(\mathrm{BW}_n)$ are defined as:

$$E^{(\mathrm{BW_n})}(\mathrm{BW}_n) = E^{(\mathrm{BW}_n)}_{\mathrm{BW}_n}(\mathrm{BW}_n) - E^{(\mathrm{BW}_n)}_{B}(\mathrm{BW}_n) - E^{(\mathrm{BW}_n)}_{W_1} - [\ldots] - E^{(\mathrm{BW}_n)}_{W_n}(BW_n)$$

$$E^{(\mathrm{def})}(\mathrm{BW}_n) = E^{(\mathrm{def})}_{B}(\mathrm{BW}_n) + E^{(\mathrm{def})}_{\mathrm{W_1}}(\mathrm{BW}_n) + [\ldots] + E^{(\mathrm{def})}_{\mathrm{W}_n}(\mathrm{BW}_n)$$

The deformation energy of monomer X (X ≡ D4T or W) is:

$$E^{(\mathrm{def})}_{\mathrm{X}}(\mathrm{BW}_n) = \mathrm{E}^{(\mathrm{X})}_{\mathrm{X}}(\mathrm{BW}_n) - E^{(\mathrm{X})}_{\mathrm{X}}(\mathrm{X}) \qquad (2)$$

Here the subscripts denote the molecular system and the superscripts indicate whether the calculation is done with the basis set of the nucleoside, (D4T), the basis set of a water molecule, (W), or the basis set of the entire system, ($\mathrm{D4TW}_n$). The contents of the baseline parentheses indicate whether the calculation is done at the optimized geometry of the entire system, (BW_n), or at the monomer optimized geometry (X).

The CP corrected interaction energy between the base (B ≡ D4T) and the water *n*mer (W_n) is computed according to:

$$\Delta E^{\mathrm{CP}}_{\mathrm{B\text{-}(W}_n)} = E^{(\mathrm{BW}_n)}_{\mathrm{BW}_n}(\mathrm{BW}_n) - E^{(\mathrm{BW}_n)}_{\mathrm{B}}(\mathrm{BW}_n) - E^{(\mathrm{BW}_n)}_{(\mathrm{W}_n)}(\mathrm{BW}_n) + E^{(\mathrm{def})}_{\mathrm{B}}(\mathrm{BW}_n) \qquad (3)$$

3. Results and discussion

3.1 Thymine molecule in an isolated state

An initial analysis of the most stable tautomers of the nucleobase thymine was carried out. The optimum geometry was obtained at the B3LYP/6-31 ++ G(3df,pd) and MP2/6-31G** levels and the relative energy values are listed in Table 1. This molecule may exist in various tautomeric forms, differing from each other by the position of the proton, which may be bound to either the ring nitrogen or oxygen atoms. From all the possible combinations, the six tautomers of Table 1 are the most stable and studied tautomers.[1,7,50]

The *keto* form, tautomer T1, always appears as the most stable one, while the *enol* forms, tautomers T2–T4 and T5–T6 are less stable (*ca.* 9–13 kcal mol^{-1} and *ca.* 17–24 kcal mol^{-1},

Table 1 Relative energies (in kcal mol^{-1}) calculated for the six most stable tautomers of the thymine molecule in the isolated state

Tautomer	T1	T2	T3	T4	T5	T6
B3LYP/6-311++G(3df,pd)+ZPVE	0[a]	10.89	12.48	13.45	18.16	21.99
MP2/6-31G**+ZPVE	0[b]	10.47	13.14	11.72	18.62	23.42
MP2/6-31G**+ZPVE[c]	0	10.47	13.14	12.74	18.62	24.21
RI-MP2+CCSD(T)[c]	0	9.28	11.43	9.46	17.07	21.47

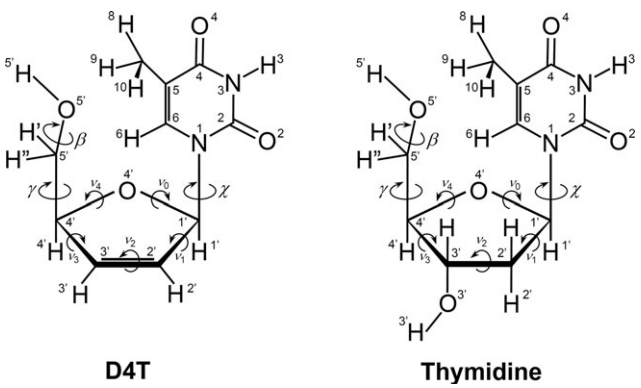
[a] Total energy (+ZPE) −454.192995 AU (atomic units). [b] −452.739098 AU. [c] Ref. 1.

respectively). Tautomers T3 and T4 are very close in energy, and their relative energies are similar to those calculated for the uracil molecule.[1,7,50,51] The relative energy values calculated appear to be in accordance to those reported at the sophisticated RI-MP2+CCSD(T) level,[1] with a difference from MP2 in tautomers T3 and T5 of *ca.* 1.7 kcal mol^{-1}. It is noted that our MP2 energies in tautomers T4 and T6 differ by 1 kcal mol^{-1} to those reported at the same level by Rejnek *et al.*,[1] although our energies are always lower, *i.e.*, more stable. Thus, the tautomeric study on the D4T molecule can be carried out at the MP2/6-31G** and B3LYP/6-311++G(3df,pd) levels with accuracy.

3.2 D4T in an isolated state

The most important angles, as well as the labeling of the atoms, are defined in Scheme 1 for the D4T molecule. An important structural characteristic in the furanose ring is the presence of a double bond, a feature that renders it nearly planar and imparts a high degree of rigidity to the structure. However, there are many conformational possibilities depending on the endocyclic and exocyclic torsional angles of the furanose ring.

The plane of the base is almost perpendicular to that of the furanose ring, and the base has two principal orientations, which are *anti* or *syn* to the glycosylic bond. These orientations are determined by the glycosylic torsional angle, χ (O4′–C1′–N1–C2). Generally, in the progress from the *syn* to the *anti* form, the furanose ring conformer is not affected. X-Ray studies[52,53] of the nucleosides have shown that the χ angle is in *anti* orientation and their geometries are stabilized by the formation of self-associated species. The preference for the *anti* conformer is necessary for biological activity.[54] Thus, all our calculations with the hydration were carried out in this *anti* conformer.



Scheme 1 Definition of the exocyclic and endocyclic torsional angles in D4T and thymidine molecules.

The exocyclic torsional angle, γ (C3′–C4′–C5′–O5′), describes the orientation of the 5′-OH hydroxyl group relative to the furanose ring. This ring is twisted out-of-plane in order to minimize nonbonding interactions between their substituents and those located in the north (N) [C2′-*exo*/C3′-*endo*] and south (S) [C2′-*endo*/C3′-*exo*] conformers.

Finally, the exocyclic torsional angle, β (C4′–C5′–O5′–H5′), describes the orientation of the hydroxyl hydrogen H5′ relative to the furanose ring. It is of particular interest since the phosphorylation of the –CH_2–OH group is required to carry out the biochemical activity of the nucleoside. Three conformers appear[21a] by rotation of this angle.

In the tautomerisation process, proton migration causes expected sizeable variations in the pyrimidine ring structure and the exocyclic CO bond lengths. The proton transfer proceeds from isolated T1 *via* a transition state (Tts) and it has been reported to be difficult, with a high barrier[55] of 178.68 kJ mol^{-1} in the uracil molecule. Only two tautomers are possible in D4T, namely T3 and T5, following the notation used for uracil.[56]

3.2.1 Conformers and energetics in D4T. Our previous work,[21a] analysing the potential energy surface (PES) of D4T by rotation of the glycoxylic bond (χ), and the γ and β exocyclic torsional angles, establishes that the more stable conformers have values around $\chi \approx 60^\circ$, -120° and 180°. The relative energies of these conformers have been calculated at the B3LYP/6-31G**, B3LYP/6-311++G(3df,pd) and MP2/6-31G** levels. Tautomers T3 and T5 were determined in the four more stable conformers (Fig. 1). The global minimum, conformer I, corresponds to the *anti-gg-gg* form with respect to χ, γ and β torsional angles, respectively. The values of these angles at the MP2 level are $\chi = -103.6^\circ$, $\beta = 63.9^\circ$ and $\gamma = 60.6^\circ$. For this conformer, the relative energies of the T3 and T5 tautomers in relation to the T1 keto form in the isolated state and with the PCM model are collected in Table 2. T3 always appears more stable than T5, *ca.* 8 kcal mol^{-1}. The stability of T3 in D4T is similar to that found in thymine (Table 1). However, T5 appears *ca.* 2 kcal mol^{-1} less stable than thymine. By symmetry, another similar conformer but with β *ca.* -60°, conformer Ia, was determined. Because of the similarities with conformer I, it was omitted in the discussion.

The second most optimum conformer considered, denoted as conformer II, corresponds to the *anti-gg-gt* form, and it is in accordance with that reported experimentally for crystals.[57] By MP2, it has values of $\chi = -122.3^\circ$, $\beta = 165.0^\circ$ and $\gamma = 43.7^\circ$. Both conformers I and II appear with the O4′ atom in the *endo* position. Its difference in energy in the T1 tautomer is very small (1.61 kcal mol^{-1} at the MP2 level).

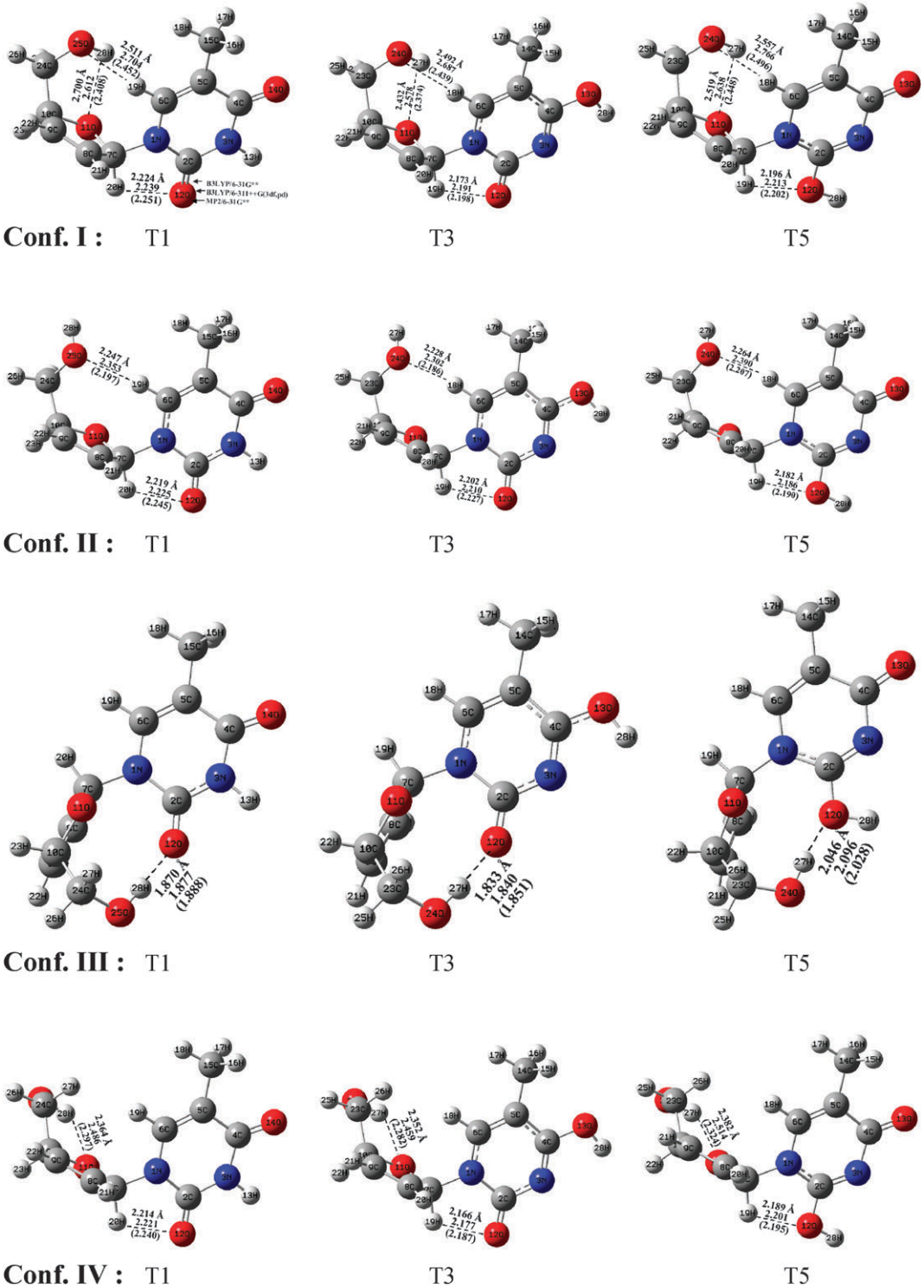


Fig. 1 "*keto*" (T1) and "*enol*" forms (T3 and T5) of the conformers studied in the present work: I, II, III, and IV of the D4T molecule. The calculated H-bonds at the B3LYP/6-31G**, B3LYP/6-311++G(3df,pd) and MP2/6-31G** levels are shown in the Figure.

In tautomer T3, a lower difference in energy between both conformers was determined (1.36 kcal mol^{-1}). This fact leads to tautomer T3 in conformer I being the most stable one of all the T3 tautomers, and analogously, tautomer T5 in conformer I is the most stable one of all the T5 tautomers, at all levels of calculation used.

Conformer III (*syn-gg-gg*) appears stabilized by an intramolecular H-bond between O2 and H5′. The selection of conformer IV (*anti-gt-tg*) follows the criterion of better energy by MP2 rather than by B3LYP.

The ketonic form, T1, is always the most stable one at all the levels calculated. Also, tautomer T3 is always more stable than

Table 2 Relative energies (in kcal mol^{-1}) calculated for the three tautomers of the four most stable conformers of the D4T molecule

	Conformer I			Conformer II			Conformer III			Conformer IV		
Level of theory	T1	T3	T5	T1	T3	T5	T1	T3	T5	T1	T3	T5
Isolated state:												
B3LYP/6-31G**	0	12.68	20.33	0	12.50	20.64	0	11.97	21.59	0	12.64	20.34
B3LYP/6-311++G(3df,pd)	0	12.76	19.59	0	12.51	20.00	0	11.87	21.19	0	12.64	19.69
MP2/6-31G**	0	13.25	20.16	0	13.00	20.43	0	12.44	21.21	0	13.24	20.16
Water solution (PCM):												
B3LYP/6-31G**	0[a]	10.05[a]	15.50[a]	0	9.84	15.56	0	9.36	16.35	0[a]	9.79[a]	15.31[a]
B3LYP/6-311++G(2d,p)	0[a]	9.79[a]	14.18[a]	0[a]	9.85[a]	14.44[a]	0	9.18	15.05	0[a]	9.79[a]	14.18[a]

[a] Saddle point.

T5 and their relative energy values are close to those of thymine (Table 1). By MP2, the calculated relative energy of tautomer T3 is slightly higher, *ca.* 0.5 kcal mol^{-1}, than by B3LYP. A similar result was determined for T5 between MP2 and B3LYP/6-311++G(3df,pd). By MP2, T3 and T5 of conformer I are the most stable tautomers. B3LYP/6-311++G(3df,pd) confirms this fact with the exception of T3 in conformer III. By B3LYP, the increase of the basis set has little effect, especially in T3.

3.2.2 Thymine moiety in D4T. The thymine ring in D4T is always planar, in agreement with the calculated and experimental results reported in the thymine molecule.[1,58] The tautomerism does not change the planarity of the ring. Thus, the O=C2–N3–C4 and C2–N3–C4=O torsional angles are, in general, close to 180°, in accordance to previous results on uracil[51] and thymine molecules. The lengths of the C–N and C–C single bonds are intermediate between those of the corresponding aromatic and saturated bonds. Hence, there is some aromatic character in the ring structure. The C4=O bond length appears slightly longer than the C2=O bond, due to the weak interaction between O4 and the hydrogens of the methyl group.

The greatest changes of geometrical parameters in the tautomerisation processes are those relating to the O4=C4–N3–C2=O2 moiety, whose bond lengths and angles differ significantly. If the proton moves to oxygen atom O4 of tautomer T3, the C4–O bond is remarkably lengthened, while the neighbors N3–C4 and C4–C5 are shortened. The N3–C4–C5 angle is increased by *ca.* 11°, while the neighboring C2–N3–C4 and C4–C5=C6 bonds decrease. An analogous effect appears when the proton moves to oxygen atom O2 of tautomer T5, with a longer C2–O bond, shorter N1–C2 and C2–N3 bonds, increase of the N1–C2–N3 angle, and decrease of the N1–C6=C5 and C2–N3–C4 angles. The calculated values of the three theoretical levels used are similar. The tautomerisation does not significantly change the torsional angles involved in the methyl group.

3.2.3 Furanose moiety in D4T. The value of the glycosylic angle, χ, differs remarkably in T1 between conformers I and II by MP2 (−103.6° and −122.3°, respectively). This discrepancy is due to a stronger O5′···H6 intramolecular H-bond in conformer II. Remarkable differences in χ also appear between conformers III and IV. The increment in conformer II is perhaps due to the strength of O2···H1′ and O5′···H6. By contrast, in T5, the value of χ is slightly lower than in T1 due to a weakness of O5′···H6 with a value longer, by MP2, by 0.05 Å in conformer I and 0.01 Å in conformer II. The orientation of the furanose ring with respect to the base also appears defined by the N1–C1′–C2′=C3′ torsional angle, with a value of 123.0° at the MP2 level in both conformers I and II. This value remains almost unchanged in both conformers and tautomers.

In conformers I and II, the furanose ring appears almost planar in T1. Comparing with the natural nucleoside Thy, the double bond C2′=C3′ in D4T greatly affects the location of the C5′-hydroxyl group relative to the position of the base. The reduced ring puckering in D4T produces a nearness of the O5′ and H6 atoms, which is slightly shortened in T3 and slightly lengthened in T5. In conformer III, these endocyclic angles are almost twice as large as in conformers I, II and IV, due to its stronger H5′···O2.

In T1, the γ torsional angle is *ca.* 60° in conformer I and 45° in conformer II. A slight increase in T3 and decrease in T5 is observed in both conformers. In T3, the value of β is slightly reduced in conformer I, but it increases in conformer II, due to the O5′···H6 intramolecular H-bond. In T5, the changes are smaller in β, as well as in χ and γ angles.

3.2.4 Intramolecular H-bonds in D4T. These were discussed following[21] the classification of Desiraju *et al.*[59] Thus, in T1 of conformer I, three H-bonds stabilize the structure as the global minimum: O2···H1′, O4′···H5′, and O5′···H6 (Fig. 1). In T3, these three H-bonds are slightly strengthened, with a shortening of *ca.* 0.04 Å. However, in T5 they are weakened.

Conformer II is only stabilized by two H-bonds: O2···H1′ and O5′···H6. This lower number of H-bonds could be responsible for the lower stability of conformer II compared to conformer I. The three H-bonds appear strengthened in T3. In T5, O2···H1′ is strengthened and O5′···H6 is weakened.

Conformer III is stabilized by a new and strong O2′···H5′ bond, which appears strengthened in T3. By contrast, this H-bond is weakened in T5. Conformer IV appears stabilized by two H-bonds, which are strengthened in T3. In T5, O2···H1′ is strengthened and O5′···H6 is weakened.

3.2.5 Charges in D4T. In general, MP2 predicts a larger negative charge on the atoms than B3LYP/6-31G**,

which predicts a larger negative charge than B3LYP/6-311++G(3df,pd). It can be pointed out that the largest negative charge is on the O5′ atom, and this charge remains almost unaffected in T3 and T5 (values of the natural NBO atomic charges on the atoms can be supplied on request).

In T1, the atom with the next largest negative charge by MP2 corresponds to N3. The charge on the O2 atom is slightly decreased in T3, and increased in negative value in T5. O4 and O4′ have similar charges in T1, but by contrast, the charge on O4 is increased in T3 and decreased in T5. That is, the negative charge on the oxygen involved in the tautomerisation increases by *ca.* 0.05 e^- by withdrawing negative charge from the bonded hydrogen. The charge on O4′ remains unaffected in the tautomers.

The largest positive charge is on C2 and C4 atoms, in concordance with the large negative charge on the O2 and O4 atoms, respectively. The O2 oxygen has a larger negative charge than O4. The hydrogen atoms with the largest positive charge, that is, the most reactive, are H5′ in T1, and H4 and H2 of the hydroxyl group in T3 and T5, respectively.

The energy gap between the HOMO and LUMO is large in T1 (0.199 Hartree by B3LYP/6-31G**), similar to that reported in uracil[55] (0.206 by B3LYP/6-31+G*). By MP2, the energy gap is much greater (0.446). In T3, the energies of the HOMO and LUMO are slightly lower than in T1, and the energy gap decreases slightly to 0.428. In T5, the energy gap is 0.451.

3.2.6 Dipole moment (μ) in D4T. The prediction of accurate dipole moments is a very important issue, because its magnitude is strongly related to the tautomeric stability in polar environments. The values of μ lie in a wide range (4.3–8.7 D). Thus, μ decreases in MP2 in the order T5 > T3 > T1. The values of μ appear slightly overestimated by MP2 (*ca.* 0.5 D) and slightly underestimated by B3LYP/6-31G**, as compared to that calculated at the B3LYP/6-311++G(3df,pd) level. Conformers I and IV have the lowest values, while conformer II has the highest one. Experimental data have not been reported for D4T, and the values in uracil are 4.13 D in dioxane solution[60] and 3.87 D in the vapour phase.[61]

3.3 Thymidine in the isolated state

3.3.1 Conformers and energetics. Scheme 1 defines the most important exocyclic and endocyclic torsional angles of this nucleoside. Thy has more stable conformers than D4T because of the higher number of degrees of freedom of the system. At all levels used, the global minimum corresponds to conformer II, while in D4T, this is conformer I. At the MP2 level, the difference of energy between both conformers is only 1.30 kcal mol^{-1}. Thus, we only discuss the results of these conformers. The tautomers in Thy (Table 3) appear more stable than those in D4T (Table 2).

Due to the C2′–C3′ single bond, the furanose ring is significantly puckered, influenced by the relative position of their substituents. The puckering is much higher in conformer I than conformer II, perhaps due to intramolecular H-bonding being stronger in conformer I. Tautomers T3 and T5 do not significantly change the puckering in conformer I and only slightly in conformer II. (Selected bond lengths and angles of the optimized conformers and tautomers can be supplied on request.)

The furanose ring appears more rotated ($\chi = -115.5^\circ$ in conformer I and -128.8° in conformer II) in Thy by MP2 than in D4T. The higher χ in Thy is due to a slight weakness of the O2···H1′ H-bond. In T3, the value of χ is remarkably increased up to -173.5° by MP2 in conformer I and up to -146.1° in conformer II, *vs.* -106.4° and -138.7°, respectively in D4T.

3.3.2 Intramolecular H-bonds in Thy. Three weak H-bondings appear in Thy, with lower strengths than those analogous in D4T, in general. This weakness of the H-bond leads to a more flexible and opened structure than D4T. The O2···H1′ is, in general, the strongest one. In conformer I, it appears slightly more shortened in Thy than in D4T, and has similar C1′–H1′···O2 angles.

The second H-bond, O5′···H6, is slightly weaker in Thy than in D4T as consequence of its higher χ angle. Although this H-bond, as well as O2···H1′, stabilizes conformer II, the stabilization is lower in Thy than in D4T, because of the remarkably large χ angle. In T3, this H-bond disappears in conformer I and it is weakened in conformer II. By contrast, in T5, the values are similar to those in T1, and also weaker than in D4T.

The third H-bond, O4′···H5′, can only appear in conformer I. The value of the O4′···H5′–O5′ angle, 99.2°, indicates a very weak H-bond that slightly stabilizes the conformer. A similar angle was determined in T3 and T5.

As in D4T and as a consequence of the large χ angle in Thy, O4′···H6 is not a H-bond in T1 of conformer I. However, in T3, the strength of this H-bond remarkably increases, with a value by MP2 of 2.198 Å and an O4′···H6–C6 angle of 102.6°. Neither H4′···O3′ nor H2′′···O3′ are considered H-bonds, revealing that intramolecular H-bonds are missing with the O3′–H3′ group. An analogous result is found in T3 and T5.

3.3.3 Charges in Thy. In general, they are similar to D4T. The main difference is a remarkable increase in the negative charge on the C2′ atom (-0.494 e^- in conformer I of Thy *vs.* -0.276 e^- conformer I of D4T). This increase is due to the C2′ atom withdrawing negative charge on the neighbor C3′ and C1′ atoms. Conformer I and II appear with similar charges. It is only noticeable that conformer II has a slightly lower negative charge on O4′ and C2′ and a higher positive charge on C4′ and C1′. (Figures with the values are available on request).

In conformer I both hydroxyl groups show almost the same charge. In conformer II the negative value in O5′ is slightly higher, due to O5′···H6 H-bond. In T3 and T5 the values remain almost unchanged. The new tautomeric hydroxyl oxygen atoms that appear in O4 (in T3) and O2 (in T5) have always a lower negative charge and their hydroxyl hydrogen atoms a higher positive value than in O5′–H and O3′–H, that is, they are more reactive.

3.3.4 Dipole moment (μ in Thy). In conformer I, its value in Thy is lower than in D4T, *ca.* 0.7 D, but in conformer II, they are similar in Thy and D4T. Conformer II has a remarkably

higher μ than conformer I, that is, it should be better stabilized in polar environments. Tautomer T5 always has a higher μ than T3. The lowest μ (*ca.* 3–3.6 D) corresponds to T3 of conformer I. (Tables with the corresponding values are available on request).

3.4 Tautomerism with water molecules

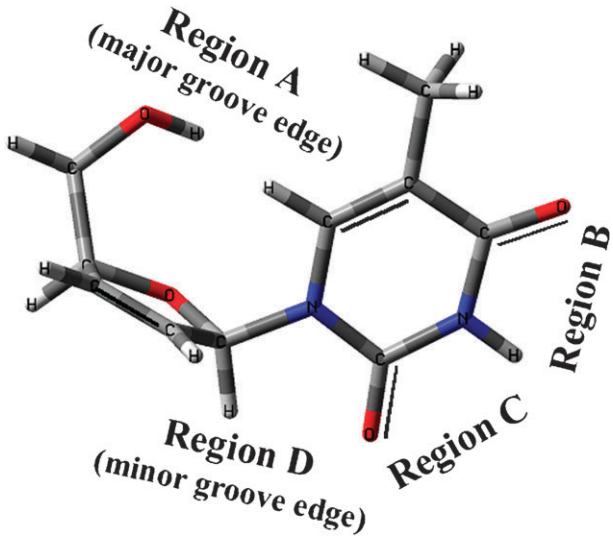
Many structural features that are necessary for the biological functions of nucleic acids depend on their interactions with surrounding water. Experimental results in the gas phase reported in uracils,[62] thymines and nucleic acid bases reveal an important function of water in protecting the genetic code from photodamage.[63–66] X-Ray and neutron diffraction studies[67] have shown that the water molecule is a versatile connector, which can serve as both a H-bond donor and acceptor. It is known that bulk water stabilizes unusual tautomers with very large dipole moments, which are extremely disfavored energetically (by *ca.* 20 kcal mol^{-1}), or reduces the gap between the global minimum (T1 in our case) and the local minima.

Theoretical results on the interaction between the nucleic acid bases with the water molecules have also been reported.[68,69] In particular, a consistent number of studies have focused on the microhydration of uracil, which is structurally the simplest base,[1,55,70] and thymine.[1,71,72] Despite these efforts, the effect of hydration on the charges and geometrical parameters, in particular in the tautomer forms, remains unclear. This understanding is crucial not only in terms of revealing the true origin of the activity of nucleic acid bases, but also in reconciling the differences between results from the isolated state and those from the condensed phase.[62] In the present work, we go further with the hydration of the tautomers of D4T and Thy nucleosides for the first time.

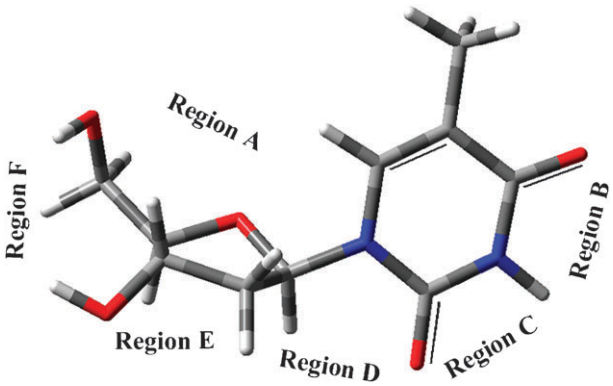
Water in the tautomerisation process provides a range of possible new H-bonded arrangements for the environmental water molecules. The relative stability of the tautomeric forms of uracil and thymine and their interactions with water molecules have been reported.[1,5,55,73,74] However, there have been no similar studies on D4T and Thy. Experimental work in methyl-substituted uracils and thymines have shown that when more than four water molecules are attached, the photophysical properties of these hydrated clusters should rapidly approach that in the condensed phase.[62] Thus, we have simulated the hydration in D4T and Thy with up to five water molecules.

Water can interact favourably in four regions of D4T, labelled as positions A, B, C and D in Scheme 2. This notation is in accordance with that followed on the thymine molecule.[72] Region A is also called "major groove edge", while region D is called "minor groove edge". In addition to the four regions of hydration in D4T, other two regions appear in Thy, one between H3′ and H1′ (region E) and the another one between H3′ and H5′ (region F) (Scheme 3).

3.4.1 D4T and their tautomers under hydration. In the next sections, the results obtained with the successive addition of explicit water molecules were described, and the most stable clusters with up to five waters were determined. The addition of further water molecules changes little the bond lengths of the nucleoside (calculated geometry parameters for D4T in the isolated state and in water solution are available on request). In general, the changes observed in the lengths are about 0.01–0.02 Å, due to an increase in the number of D4T-water H-bonds. The methyl group is barely affected by the water molecules.



Scheme 2 The four regions of hydration in D4T.



Scheme 3 The six regions of hydration in Thy.

The presence of water is first considered within a simple model with only one molecule, and the order of filling up to five water molecules is described in a previous work.[21b] In the notation used, the number 1 or 2 refers to the number of intermolecular H-bonds of the water molecule, and the letters A–D refer to its position. Water molecules are thus denoted as W_{A2} to W_{D2}. In the bottom of Fig. 2 and 3 is listed the total energy of this molecule at the B3LYP/6-31G** level.

3.4.1.1 Monohydration: D4T-H2O.

3.4.1.1.1 Clusters and energetics. Eight stable positions appear in the T1 form.[21b] Water interactions do not modify the stability order of the tautomers in the isolated state of either conformers. Thus, T1 remains the most stable tautomer, with a similar relative stability in T3 and T5. All cluster structures are characterized by a planar thymine ring.

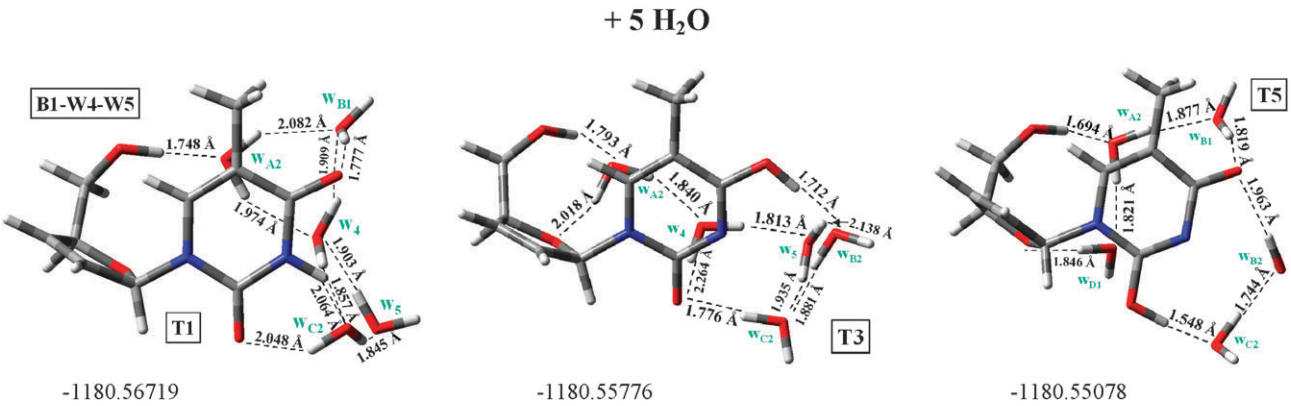


Fig. 2 Optimum clusters with five water molecules in the hydration of conformer I in D4T. The B3LYP energy was calculated with the 6-31G** basis set.

The stability order of the four main positions in T1 of conformer I is:

A2		B2		C2		D2	
0	>	0.01	>	0.72	>	4.05	(kcal by B3LYP)
0		1.92		2.42		4.69	(kcal by MP2)

Both methods have the same stability order. In T3, the stability order is: B2 > A2 > D2 > C2, while in T5 this is C2 > A2 > B2 > D2. This is due to the higher reactivity (higher charge) of N3 and H4 in T3 (or H2 in T5) of the hydroxyl group than O4′ and H5′ of the monohydration in A2.

In T1 of conformer II, the stability order differs:

B2		C2		D2		C1	
0	>	0.92	>	5.27	>	4.40	(kcal by B3LYP)
0		0.61		2.93		4.12	(kcal by MP2)

This is due to A2 not being stable, due to the low barrier height corresponding to the β torsional angle. Thus, conformer II changes to conformer I.[21b] In conformer I, the two H-bonds of the water molecule in A2 forces the β torsional angle to a value of 88.7°, when in the isolated state, it is 64.9°, impeding the stability of conformer II with a β angle of 164.4° in the isolated state.

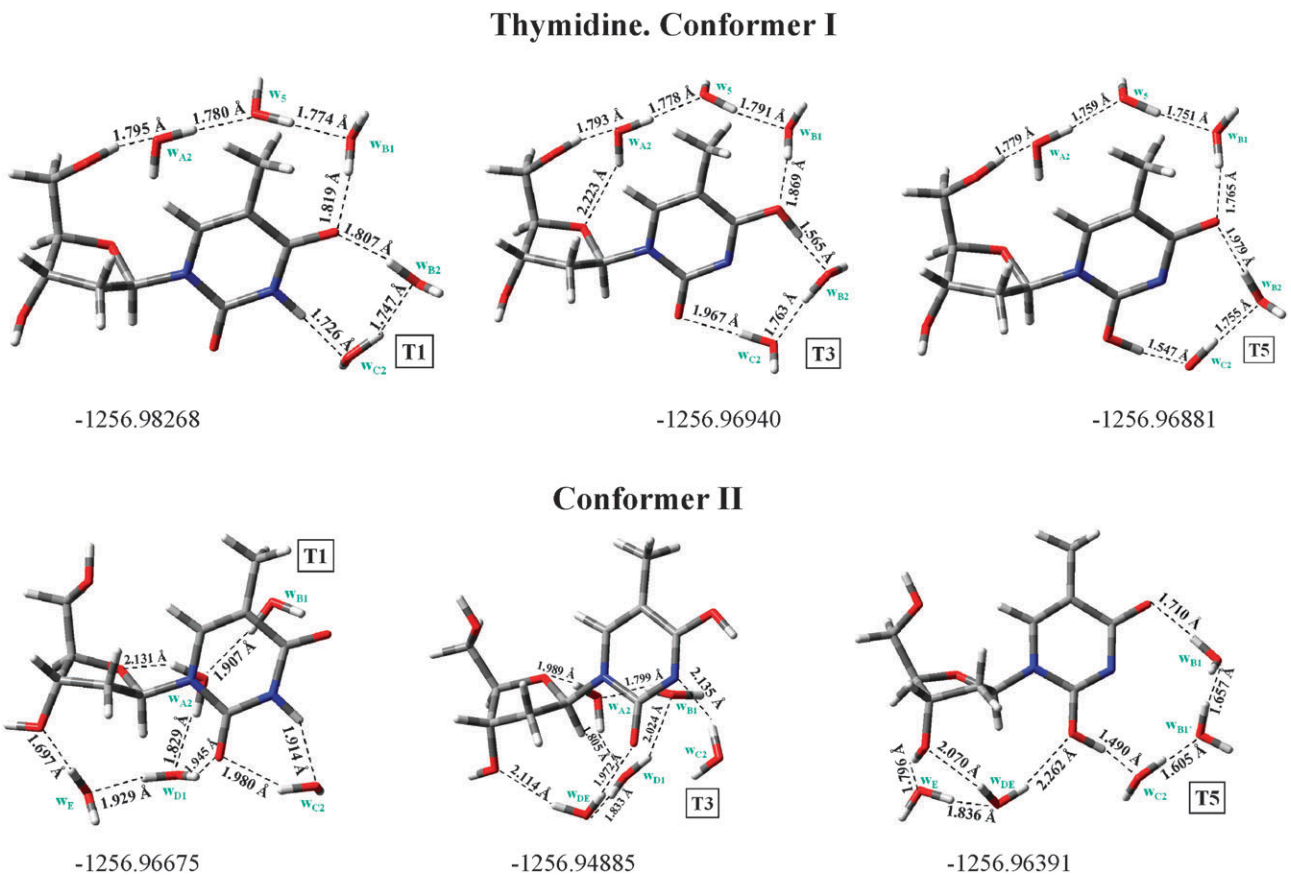


Fig. 3 Optimum clusters with five water molecules in the hydration of tautomers T1, T3 and T5 of conformer I and II of Thy.

CP-Energies: In T1, the water molecule in positions A2 and B2 represents the maximum CP complex formation energy (ΔE^{CP}_{D4TW1}) and the CP corrected interaction energy between D4T and the water molecule, ΔE^{CP}_{D4T-W1}, and thus these positions are the most stable ones. Small differences between MP2 and B3LYP are observed, as well as between conformer I and II. In T3, the value of ΔE^{CP} in position B2 is much higher (*ca.* 4 kcal mol^{-1}) than in T1, and thus this position is the most stable one. In T5, an additional increase in this energy (*ca.* 2 kcal mol^{-1}) is observed in position C2, the most stable one. These features indicate that although the stability order of the tautomers is T1 > T3 > T5, the strength order of the H-bonds in the water molecules is T5 (in C2) > T3 (in B2) > T1 (A2).

The deformation energy is much higher in D4T than in the water molecule, W_1. The maximum value for D4T, E^{def}_{D4T}, appears in position A2, but by contrast, it is very small for the water molecule, E^{def}_{W1}. The exception is the calculated value in C2 of T5. It is due to the short H2$\cdots$O$_w$ of *ca.* 1.65 Å. The highest deformation of water appears in the B2 and C2 clusters, due to the water forming two strong H-bonds of similar strengths in them, while in A2, the O4$'\cdots$H$_w$ bond is slightly weak.

Because conformer II is not stable with more than eight water molecules,[21b] only the results of conformer I are discussed in the next sections.

3.4.1.1.2 Effects of the monohydration. The main effect is a lengthening of the N–H and C=O (or O–H in T3 and T5) bonds involved in intermolecular H-bonds (by *ca.* 0.01–0.03 Å). By contrast, the neighbours, N–C (or N–C and C–C of cluster B2), shrank by 0.005–0.01 Å. This resembles the behaviour of a typical amide bond in water (enhanced amide resonance due to solute–solvent interaction).[75] As expected, the C–H bonds are not sensitive to hydration, and the C$_{methyl}$–H has a very small lengthening. In T1, the O2, O4 and N3 atoms involved in H-bonds increase its negative charge by 0.04 e$^-$ and 0.06 e$^-$, respectively, while H3 is more positive by *ca.* 0.08 e$^-$. A similar behaviour appears in T3 and T5, with an increase in the positive charge on the hydroxyl hydrogen and on the N3 atom.

3.4.1.2 D4T-$(H_2O)_1$ to -$(H_2O)_5$. The methodology of the addition of the second water molecule in the most stable cluster calculated in the monohydration, A2 (in tautomer T1) or B2 (in T3) or C2 (in T5), was mainly followed, although other combinations were also determined.[21b] Successively, the same methodology has been applied for incorporating the rest of water molecules. In the pentahydrated cluster, Fig. 2,3, the following is observed with the hydration:

(i) the clusters with water in the dimer form appear more stable than with water in the monomer form. Because water–water H-bonds are preferred over water–D4T, the water molecules appear more strongly H-bonded among themselves and weaker with D4T with the progression of the hydration.

(ii) The hydration stabilizes T3 and T5 more than T1, although this one continues being the most stable. This feature has also been reported in uracil and thymine.[1] Conformer I appears more stabilized than conformer II, and this stabilization increases with the successive addition of water molecules.

(iii) The optimum distribution of the water molecules, binding to the most polar groups, is, in general, in accordance to the previous hydrated structure found in the uracil molecule, although the distribution differs in the three tautomers.

(iv) In T1, the water molecule W_{A2} is the most H-bonded to D4T, with a value of the H-bond length in the different clusters ranging from 1.71 to 1.86 Å. In T3, the water molecule W_{B2}, and W_{C2} in T5, appear more strongly H-bonded to D4T than in T1, with values ranging from 1.57–1.84 Å in T3 and 1.55–1.79 Å in T5. The deformation energy is much higher in W_4 than in the other water molecules, and it increases with successive hydration.

(v) The intramolecular H-bond O2$\cdots$H1$'$ is slightly lengthened and the C1$'$-H1$'\cdots$O2 angle is slightly reduced by *ca.* 1–3°. Further hydration up to 13 water molecules[21b] breaks this weak H-bond, due to the rotation of the thymine and furanose rings. The O5$'\cdots$H6 is also weakened, with a slightly shorter length and a smaller O5$'\cdots$H6–C6 angle.

(vi) With the hydration, the geometric parameters change a little more in T3 than in T5. The water molecules slightly affect the planarity of the pyrimidine ring, with changes in their torsional angles of less than 4°, in general. However, the χ torsional angle changes remarkably, in the range 90–115° in T1 and T5 by B3LYP, and 77–115° in T3. Also β changes remarkably (89–97° in T1, 64–88° in T3 and 64–113° in T5). Minor differences appear in γ (55–60° in T1, 60–65° in T3 and 51–61° in T5). These features show the great flexibility of D4T. A noticeable lengthening of the C=O, N–H and O–H bonds, shortening of the C–O bonds, and an increase of the *ipso* ring angles involved in H-bonding are also observed.

(vii) The simulation of the hydration in T1 with five explicit water molecules appears adequate, because the further increase of up to 13 water molecules[21b] in general only shows slight differences in the charges. As an exception, a greater enhancement of the charges appear on O4$'$, H1$'$, H$''$ and O2, and a decrease appear on C6 and O4.

(viii) In T1, the negative charge on O2 is always higher than on O4, so in principle, O2 will be more susceptible to attack by a proton than O4. However, hydration with few water molecules increases the negative charge more on O4 than O2, and thus with five water molecules, the negative charge on O4 is higher than O2. This fact, together with a slightly shorter distance between H3 and O4 than between H3 and O2 in the pentahydrated cluster, indicates an easier tautomerisation to T3 than to T5.

3.4.1.3 PCM.

3.4.1.3.1 Energetics. Table 2 shows the energy of tautomers in the four most optimum conformers. With PCM, conformers I and IV are saddle points while conformers II and III are the only ones that correspond to real minima. This fact is observed by both the 6-31G** and 6-311++G(2d,p) basis set, although with the latter, conformer II is also a saddle point. The increase of the basis set stabilizes T5 by *ca.* 1 kcal mol^{-1}.

PCM stabilizes the enol forms more than the canonical keto one (*ca.* 3 kcal mol^{-1} in T3 and *ca.* 5 kcal mol^{-1} in T5). However, this stabilization is much higher by the discrete model (DM) (*ca.* 7 kcal mol^{-1} in T3 and *ca.* 10 kcal mol^{-1} in T5). There is another difference between the two models,

i.e. conformer II is not stable[21b] by DM and it changes to conformer I, whereas by PCM conformer II is the most stable form. According to PCM, the external polar cavity attracts H5′ of conformer I to the more open structure of conformer II. However, this model fails when water molecules can be introduced in the holes inside of the molecule, such as W_{A2}, where closer structures are favoured, such as conformer I. Thus, we can conclude that conformer I is the only form present in the first hydration shell.

By PCM, the energies of the HOMO and LUMO remain almost unchanged. However, by DM, a slight increase in the negative energies of the HOMO and LUMO is observed, with values of −0.058 and −0.259, respectively. The energy gap is similar (0.202 by DM *vs.* 0.200 by PCM, and 0.199 in the isolated state).

3.4.1.3.2 Geometries. Comparing the calculated bond lengths and angles by PCM with those by DM with five explicit water molecules, the following can be noted:

(i) the results by PCM are closer to the isolated state geometry than those by DM. The largest differences between both models appear in T1 and T3, while they are small in T5. This is due to a stronger H-bond net by DM in T1 and T3 than in T5. These differences also appear, in general, in atoms H-bonded to water molecules (*ca.* 0.02 Å in the bond lengths and *ca.* 2° in the angles).

(ii) In T1, the O2···H1′ H-bond is weakened with the hydration, and the weakening is greater by PCM than by DM.

(iii) Another difference between the two models appear in the weak H-bond O5′···H6, which remains almost unchanged by PCM, in clear disagreement with DM (this H-bond disappears).

(iv) The dipole moment, μ, calculated by PCM is *ca.* 2–3 D higher than in the isolated state. By contrast, by DM, it is remarkably lower in T1 and in T5, and higher in T3.

For these features, we can conclude that the PCM model does not reproduce the hydration pattern of D4T well.

3.4.1.4. Charges. The main effect observed with the hydration, by both the PCM and DM models and in all the tautomers, is a general enhancement of the positive and negative charges on the atoms, that is, an increase of their reactivity. The effect is very small on the atoms of the furanose ring, and much higher on the pyrimidine ring atoms. By the PCM and DM models, the atoms with strong H-bonds with water have a high enhancement of their charges, especially the oxygen atoms. O5′ continues to be the atom with the highest negative charge in all the tautomers, and it increases again with the hydration. The highest positive charge corresponds to C2, but it changes very little with the hydration.

3.4.2 Thymidine hydration.

3.4.2.1 Energetics and geometries. The most stable monohydrated cluster in T1 of conformer II is that with the water molecule in the position D–E and with H3′···O_w and O2···H_w H-bonds. This cluster is only 0.93 kcal mol^{-1} more stable than A2 of conformer I. In tautomer T3 of conformer II, the minimum corresponds to B2 and in T5 to C2, in accordance with the high reactivity of O4–H and O2–H groups. The successive hydrated clusters were determined following the same methodology as that in D4T. For simplicity, in Fig. 3 we only show the final optimum pentahydrated clusters for conformers I and II and in the three tautomers. It can be noted that the different distribution of the water molecules depends on the conformer and tautomer.

The hydration of conformer I in Thy shows the water molecules distributed in the regions A, B and C around the methyl group. In conformer II, the water molecules are clustered in the B, C, D and E regions, in general, with more Thy-water H-bonds and far from the methyl group. By contrast, conformer I in D4T shows clustered water molecules in the A, B, and C regions around the methyl group (Fig. 2).

In most of the clusters, water accepts a proton from the nucleoside hydroxyl group and donates its proton to another water molecule. In conformer II, the anticooperative water molecule W_{DE} acts as bidonor. A linear trimer of water molecules appears strongly H-bonded in the three tautomers of conformer I and in T5 of conformer II, with H-bond lengths of 1.75–1.79 Å in conformer I and with 1.60–1.66 Å in conformer II.

A much greater puckering of the ribose ring in Thy than in D4T leads to a larger O5′···O4 intramolecular distance in Thy than in D4T, and therefore a wider major groove edge. Thus, in Thy, three water molecules are required to connect H5′ with O4, while in D4T only two are required to connect with O2.

Conformer I has a similar χ angle in the three tautomers, and close to that calculated in the isolated state. In T3, due to the strengthening of O2···H1′ with hydration, the high value of χ of −169.6° in the isolated state by B3LYP drops to −119.5°. In conformer II, the χ angle is much greater in T3 and T5 due to the distribution of the water molecules. Under hydration, the β torsional angle is remarkably rotated in conformer I. However, in conformer II, it changes very little with the hydration, due to the fact that not one of the five first water molecules appear to be H-bonded to O5′–H.

The O3′–H group appears remarkably rotated in conformer II. It is due to water molecules being more strongly H-bonded to O3′–H than to O5′–H, leading to the high value of this torsional angle, which could facilitate the group's reactivity. PCM fails on this point and does not show changes in this angle. By PCM, T3 and T5 of conformer I are not stable and they change to conformer II (Table 3). This is due to the fact

Table 3 Relative energies (in kcal mol^{-1}) calculated for the three tautomers of the two most stable conformers of thymidine nucleoside

	Conformer I			Conformer II		
Level of theory	T1	T3	T5	T1	T3	T5
Isolated state:						
B3LYP/6-31G** (+ZPE)	0[a]	11.52	20.14	0[e]	12.36	20.29
B3LYP/6-311++G(3df,pd)	0[b]	12.31	19.63	0[f]	12.46	19.67
MP2/6-31G**	0[c]	11.64	19.98	0[g]	12.95	20.10
Water solution (PCM):						
B3LYP/6-31G** (+ZPE)	0[d]	(n)	(n)	0[h]	9.87	15.32

[a] −874.90391 AU. [b] −875.45806 AU. [c] −872.67179 AU. [d] −874.90391 AU. [e] −874.90537 AU. [f] −875.45845 AU. [g] −872.67387 AU. [h] −874.94653 AU.(n) not stable.

that this model favours a more open structure of conformer II than conformer I. T1 of conformer I is a stable minimum, in contrast to the saddle point determined in D4T.

In the energies of the HOMO and LUMO, the effect is almost null by PCM. The energy gap is similar (0.202 by DM *vs.* 0.200 by PCM, and 0.199 in the isolated state).

3.4.2.2 Intramolecular H-bonds. O2···H1′ is slightly weaker in T1 and T5 of conformer I with five water molecules than in the isolated state. By contrast, in T3, it is strengthened. In conformer II, this H-bond is further strengthened with five water molecules. This is due to W_E and W_{D1} or W_{DE} (in T5) shortening the distance between O2 and O3′, and thus shortening O2···H1′. In T3, it is slightly weaker.

O5′···H6 is remarkably strengthened in conformer I. In conformer II, this H-bond is also strengthened, but by contrast, it is remarkably weaker in T3 and in T5. O4′···H6 does not appear in the hydrated tautomers. Only a weak H-bond is observed in T3 of conformer II.

3.4.2.3 Charges. As in D4T, thymine and cytosine,[47] the main effect of the hydration is an enhancement of the negative charges on the oxygen and nitrogen atoms and an increase of the positive charge on H3. The charges obtained in T1 in the cluster with five water molecules appear close to those calculated with 13 water molecules,[21b] with very small differences. Although the calculated values by PCM are similar to those determined in D4T, the results differ a little by DM due to the different distribution of the water molecules. The main difference is a larger negative charge on O4 and a slightly smaller one on O2.

Because of the different distribution of the water molecules in conformer I than in conformer II, the charges on O2 and O4 appear reversed in T1. In conformer II, the negative charge on O2 always appears higher than O4. Comparing hydrated tautomers T1 → T3, an increase of the charge on O4 and C6, and a decrease on N3 and C4 are noted. Analogously, in T1 → T5, an increase on O4 and O2 and a decrease on C2, N3 and C4 is observed.

3.4.2.4 Intramolecular distances. This is of interest due to a possible relation with the acceptor sites in the enzyme. The main effect of the hydration is a remarkable shrinking of the structure, with shorter O5′···O2 and O5′···O4 distances. This shrinking increases with the further addition of water molecules. The structure of conformer I in Thy appears slightly more open than in D4T, with longer O5′···O2 and O5′···O4 distances. Both B3LYP and MP2 methods agree on these data. In O5′···O2, the change is much greater in T1 → T3 than in T1 → T5 tautomerisation. In O5′···O4′, the variations are very small in spite of the fact that, in the isolated state, the furanose ring differs with that of D4T. However, the tautomerisation in D4T only modifies this distance by *ca.* 0.02 Å, while in Thy the changes in T1 → T3 are greater (0.08 Å). With the hydration, the changes are reversed. In Thy, the O2···O4 distance in T1 and T5 of conformer I is exactly the same as that in D4T and it remains almost unchanged in the tautomers.

4. Summary and conclusions

In the present work, we have studied the tautomerisation of D4T and Thy and the influence of water on the stability of the tautomers. A good agreement is obtained, whenever available, with analogous theoretical studies, supporting the quality of our results derived from computations. The most important findings of this study are the following:

(1) D4T and Thy nucleoside equilibrium geometries were calculated for the first time at the B3LYP and MP2 quantum chemical levels. More than 200 cluster-optimized structures were determined.

(2) Conformer I of D4T was the global minimum in the isolated state and under hydration at all levels of computation. However, conformer II of Thy is the most stable one, evidenced by the high χ angle that strengths the O5′···H6 H-bond. In Thy, the hydration increases the difference in energy of both conformers.

(3) Two tautomeric forms, T3 and T5, were determined and optimised in the four most stable conformers of D4T and Thy. The stability order is similar in both molecules. In D4T, T3 and T5 of conformer I are the most stable tautomers of all the conformers, while in Thy they correspond to conformer II.

(4) The T1 → T3 tautomerism appears thermodynamically more favourable in the natural nucleoside Thy than in the analogue D4T and in thymine. In conformer I, the energy difference is 1.6 kcal mol^{-1} at the MP2 level. By contrast, T1 → T5 tautomerisation appears slightly more favourable in thymine (*ca.* 1.4 kcal mol^{-1}) than in Thy, and this one is *ca.* 0.2 kcal mol^{-1} more stable than in D4T. In conformer II, the differences are much more reduced (only 0.05 kcal mol^{-1} in T1 → T3 and only 0.3 kcal mol^{-1} in T1 → T5 at the MP2 level).

(5) The stronger H-bonds between water molecules than between a water molecule and the nucleoside favour the appearance of a water net. This net differs in the three tautomers of D4T and Thy. Depending on the nature of the tautomers, cyclic (distributed water molecules) or clustered structures are formed in the hydration. In most of the cyclic structures, water accepts the OH proton of the tautomer and donates its proton to another water molecule, while in clustered structures, water–water interactions dominate.

(6) The structure of conformers I and II in the three tautomers appears stabilized by stronger intramolecular H-bonds in D4T than in Thy. This feature could indicate a higher flexibility and reactivity for Thy than D4T. With the hydration, all the intramolecular H-bonds are weakened.

(7) The structure of Thy appears more open than in D4T. The tautomerisation does not significantly change the puckering of the furanose ring. The T1 → T3 tautomerisation changes the geometry of the molecule, and the positions and distances of the oxygen atoms more in Thy than in D4T. The changes are greater in conformer II, but very small in T1 → T5.

(8) O4–H (in T3) and O2–H (in T5) tautomeric hydroxyl groups appear more reactive, with a higher charge on the hydrogen atom than the O5′–H and O3′–H non-tautomeric groups.

(9) Conformer II has the highest μ in D4T and Thy, that is, this conformer is the more active form in polar environments. Tautomer T5 has, in general, a higher μ than T3.

(10) In both D4T and Thy molecules, the hydrated clusters with tautomers T3 and T5 are less stable than with T1 and thus this keto form is the dominant one under hydration. However, the water molecules stabilize the enol forms more than the keto ones. The stabilization is *ca.* twice higher in T5 than in T3. Although in T1 the water molecules in the B and C positions can protect from tautomerisation, they reduce the energy gap between the keto and the enol tautomers, and increase the reactivity of O2 and O4.

(11) In the monohydration, the water molecule appears more strongly H-bonded in T5 (in C2) than in T3 (in B2), which is stronger than in T1 (in A2). This stronger H-bonding of the water molecules is responsible for the reduction of the energy gap between the keto and the enol tautomers. The effect of the PCM model on the geometry of D4T and Thy is much lower than by DM with five explicit water molecules.

(12) In D4T, the specific interaction with one water molecule reduces the energy gaps between T1 and T3, and between T1 and T5. The addition of a second water molecule has a relatively very small effect. The third water molecule produces a remarkable further reduction, in T3 and in T5. With the addition of the fourth water molecule, the energy gap becomes half that of the isolated state in both T3 and T5 tautomers. With a further increase in the number of water molecules, the tautomerisation is reinforced.

Acknowledgements

The authors are grateful to the UCM of Spain for financial support through UCM-BSCH GR58/08 grant numbers 921628 and CTQ 2006-14933/BQU, and the Computational Center at Universidad Complutense of Madrid for running the Gaussian program package.

References

1 J. Rejnek, M. Hanus, M. Kabelác, F. Ryjácek and P. Hobza, *Phys. Chem. Chem. Phys.*, 2005, **7**, 2006.
2 M. A. Morsy, A. M. Al-Somali and A. Suwaiyan, *J. Phys. Chem. B*, 1999, **103**, 11205.
3 (*a*) A. Les and L. Adamowicz, *J. Phys. Chem.*, 1989, **93**, 7078; (*b*) A. Les and L. Adamowicz, *J. Phys. Chem.*, 1990, **94**, 7021; (*c*) S. Millefiori and A. Alparone, *Chem. Phys.*, 2004, **303**(1–2), 27.
4 (*a*) M. D. Topal and J. R. Fresco, *Nature*, 1976, **263**, 285; (*b*) X. Hu, H. Li, W. Liang and S. Han, *J. Phys. Chem. B*, 2005, **109**(12), 5935.
5 E. S. Kryachko, M. T. Nguyen and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 2001, **105**, 1288.
6 J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 737.
7 M. Piacenza and S. Grimme, *J. Comput. Chem.*, 2004, **25**(1), 83.
8 E. S. Kryachko and J. R. Sabin, *Int. J. Quantum Chem.*, 2003, **91**(6), 695.
9 S. R. Holbrook, C. Cheong and S. H. Kim, *Nature*, 1991, **353**, 579.
10 K. Shi, M. Wahl and M. Sundaralingam, *Nucleic Acids Res.*, 1999, **27**, 2196.
11 V. I. Danilov, V. M. Anisimov, N. Kurita and D. Hovorun, *Chem. Phys. Lett.*, 2005, **412**, 285.
12 M. D. Topal and J. R. Fresco, *Nature*, 1976, **263**, 285.
13 (*a*) D. A. Estrin, L. Paglieri and G. Corongiu, *J. Phys. Chem.*, 1994, **98**, 5653; (*b*) J. Rejnek and P. Hobza, *J. Phys. Chem. B*, 2007, **111**(3), 641.
14 (*a*) T.-K. Ha and H. H. Gunthard, *THEOCHEM*, 1992, **276**, 209; (*b*) J. Leszczynski, *J. Phys. Chem.*, 1992, **96**, 1649; (*c*) T. M. El-Gogary and A. M. El-Nahas, *THEOCHEM*, 2008, **851**(1–3), 54.
15 (*a*) J. W. Boughton and P. Pulay, *Int. J. Quantum Chem.*, 1993, **47**, 49; (*b*) R. Czerminski, K. Szczepaniak, W. B. Person and J. S. Kwiatkowski, *J. Mol. Struct.*, 1990, **237**, 151.
16 (*a*) I. R. Gould, N. A. Burton, R. J. Hall and I. H. Hillier, *THEOCHEM*, 1995, **331**, 147; (*b*) T.-K. Ha, H.-J. Keller, R. Gunde and H.-H. Gunthard, *J. Phys. Chem. A*, 1999, **103**(33), 6612.
17 (*a*) O. C. Desfrançois, H. Abdoul-Carime and J. P. Schermann, *J. Chem. Phys.*, 1996, **104**, 7792; (*b*) R. A. Bachorz, J. Rak and M. Gutowski, *Phys. Chem. Chem. Phys.*, 2005, **7**(10), 2116; (*c*) G. N. Ten and V. I. Baranov, *J. Appl. Spectrosc.*, 2004, **71**(6), 767.
18 J. H. Hendricks, S. A. Lyapustina, H. L. Clercq and K. H. Bowen, *J. Chem. Phys.*, 1998, **108**, 8.
19 (*a*) M. Chahinian, H. B. Seba and B. Ancian, *Chem. Phys. Lett.*, 1998, **285**, 337; (*b*) M. K. Shukla and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**(37), 8642.
20 (*a*) M.-P. Gaigeot, N. Leulliot, M. Ghomi, H. Jobic, C. Coulombeau and O. Bouloussa, *Chem. Phys.*, 2000, **261**, 217; (*b*) P. Beak and J. M. White, *J. Am. Chem. Soc.*, 1982, **104**, 7073; (*c*) X. Li, K. H. Bowen, M. Haranczyk, R. A. Bachorz, K. Mazurkiewicz, J. Rak and M. Gutowski, *J. Chem. Phys.*, 2007, **127**(17), 174309.
21 (*a*) M. Alcolea Palafox, N. Iza, J. Talaya and M. de la Fuente, *J. Phys. Chem. B*, in preparation; (*b*) M. Alcolea Palafox, N. Iza, M. de la Fuente and R. Navarro, *J. Phys. Chem. B*, 2009, **113**(8), 2458; (*c*) Y. P. Yurenko, R. O. Zhurakivsky, S. P. Samijlenko, M. Ghomi and D. M. Hovorum, *Chem. Phys. Lett.*, 2007, **447**, 140; (*d*) Y. P. Yurenko, R. O. Zhurakivsky, M. Ghomi, S. P. Samijlenko and D. M. Hovorum, *J. Phys. Chem. B*, 2008, **112**, 1240.
22 (*a*) E. De Clercq, *J. Clin. Virol.*, 2001, **22**, 73; (*b*) E. De Clercq, *Biochim. Biophys. Acta, Mol. Basis Dis.*, 2002, **1587**, 258; (*c*) E. De Clercq, *Med. Chem. Res.*, 2004, **13**(6–7), 439.
23 (*a*) Y. P. Yurenko, R. O. Zhurakivsky, M. Ghomi, S. P. Samijlenko and D. M. Hovorum, *J. Phys. Chem. B*, 2007, **111**, 6263; (*b*) Z. Delbederi, C. Fossey, G. Fontaine, S. Benzaria, D. Gavriliu, A. Ciurea, B. Lelong, D. Laduree, A. M. Aubertin and A. Kim, *Nucleosides, Nucleotides Nucleic Acids*, 2000, **19**(9), 1441; (*c*) T. K. Venkatachalam, G. Y. Yu, P. Samuel, S. Qazi, S. Pendergrass and F. M. Uckun, *Eur. J. Med. Chem.*, 2004, **39**, 665; (*d*) A. Yvon-Groussin, P. Mugnier, P. Bertin, M. Grandadam, H. Agut, J. M. Huraux and V. Calvez, *J. Virol.*, 2001, **75**(15), 7184.
24 D. J. Newman, G. M. Cragg and K. M. Snader, *Nat. Prod. Rep.*, 2000, **17**, 215.
25 R. A. Seaton, R. Fox, N. Bodasing, S. E. Peters and Y. Gourlay, *AIDS*, 2003, **17**, 445.
26 L. K. Naeger, N. A. Margot and M. D. Miller, *Antimicrob. Agents Chemother.*, 2002, **46**, 2179.
27 H. J. Yekeler, *THEOCHEM*, 2004, **684**, 223.
28 (*a*) G. M. Ciuffo, M. B. Santillan, M. R. Estrada, L. J. Yamin and E. A. Jáuregui, *THEOCHEM*, 1998, **428**, 155; (*b*) Y. P. Yurenko, R. O. Zhurakivsky, M. Ghomi, S. P. Samijlenko and D. M. Hovorum, *J. Phys. Chem. B*, 2007, **111**, 9655.
29 S. Obika, J. Andoh, T. Sugimoto, K. Miyashita and T. Imanishi, *Tetrahedron Lett.*, 1999, **40**, 6465.
30 V. O. Nava-Salgado, R. Martínez, M. F. R. Arroyo and G. R. Galicia, *THEOCHEM*, 2000, **504**, 69.
31 N. G. Fidanza, F. D. Suvire, G. L. Sosa, R. M. Lobayan, R. D. Enriz and N. M. Peruchena, *THEOCHEM*, 2001, **543**, 185.
32 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin,

D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision B.04)*, Gaussian, Inc., Pittsburg, PA, 2003.
33 (*a*) M. Hoffmann and J. Rychlewski, *Density functional theory (DFT) and drug design. Reviews of Modern Quantum Chemistry*, 2002, **2**, 1767; (*b*) V. I. Danilov, T. Van Mourik, N. Kurita, H. Wakabayashi, T. Tsukamoto and D. M. Hovorum, *J. Phys. Chem. A*, 2009, **113**, 2233; (*c*) Y. V. Il'ichev and J. D. Simon, *J. Phys. Chem. B*, 2003, **107**, 7162; (*d*) O. S. Sukhanov, O. V. Shishkin, L. Gorb, Y. Podolyan and J. Leszczynski, *J. Phys. Chem. B*, 2003, **107**, 2846; (*e*) J. E. Sponer, N. Spackova, P. Kulhanek, J. Leszczynski and J. Sponer, *J. Phys. Chem. A*, 2005, **109**, 2292.
34 M. Alcolea Palafox, N. Iza and M. Gil, *THEOCHEM*, 2002, **585**, 69.
35 M. Alcolea Palafox, O. F. Nielsen, K. Lang, P. Garg and V. K. Rastogi, *Asian Chem. Letts*., 2004, **8**, 81.
36 M. Alcolea Palafox and V. K. Rastogi, *Spectrochim. Acta, Part A*, 2002, **58**, 411.
37 M. Alcolea Palafox, *Recent Research Developments in Physical Chemistry*, Transworld Research Network, India, 1998, p. 2.
38 M. Alcolea Palafox, *Int. J. Quantum Chem*., 2000, **77**, 661.
39 J. E. Carpenter and F. Weinhold, *THEOCHEM*, 1988, **169**, 41.
40 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev*., 1988, **88**, 899.
41 A. Aamouche, G. Berthier, B. Cadioli, E. Gallinella and M. Ghomi, *THEOCHEM*, 1998, **426**, 307.
42 R. W. Williams and A. H. Lowrey, *J. Comput. Chem*., 1991, **12**, 761.
43 J. Tomasi and M. Persico, *Chem. Rev*., 1994, **94**, 2027.
44 R. W. Williams, J. L. Cheh, A. H. Lowrey and A. F. Weif, *J. Phys. Chem*., 1995, **99**, 5299.
45 (*a*) S. Miertus and J. Tomasi, *Chem. Phys*., 1982, **65**, 239; (*b*) M. Cossi, V. Barone, R. Cammi and J. Tomasi, *Chem. Phys. Lett*., 1996, **255**, 327; (*c*) M. T. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys*., 1997, **107**, 3032; (*d*) V. Barone, M. Cossi and J. Tomasi, *J. Chem. Phys*., 1997, **107**, 3210; (*e*) M. Cossi, V. Barone, B. Mennucci and J. Tomasi, *Chem. Phys. Lett*., 1998, **286**, 253; (*f*) V. Barone, M. Cossi and J. Tomasi, *J. Comput. Chem*., 1998, **19**, 404; (*g*) V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995; (*h*) B. Mennucci and J. Tomasi, *J. Chem. Phys*., 1997, **106**, 5151; (*i*) J. Tomasi, B. Mennucci and E. Cancès, *THEOCHEM*, 1999, **464**, 211.
46 (*a*) B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B*, 1997, **101**, 10506; (*b*) J. Tomasi, B. Mennucci and E. Cancès, *THEOCHEM*, 1999, **464**, 211; (*c*) M. Cossi, V. Barone and M. A. Robb, *J. Chem. Phys*., 1999, **111**, 5295; (*d*) R. Cammi, B. Mennucci and J. Tomasi, *J. Phys. Chem. A*, 2000, **104**, 5631; (*e*) M. Cossi and V. Barone, *J. Chem. Phys*., 2000, **112**, 2427; (*f*) M. Cossi and V. Barone, *J. Chem. Phys*., 2001, **115**, 4708; (*g*) M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Chem. Phys*., 2001, **114**, 5691; (*h*) M. Cossi, G. Scalmani, N. Rega and V. Barone, *J. Chem. Phys*., 2002, **117**, 43; (*i*) M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem*., 2003, **24**, 669.
47 C. Alemán, *Chem. Phys. Lett*., 1999, **302**, 461.
48 O. V. Shishkin, L. Gorb and J. Leszczynski, *Int. J. Mol. Sci*., 2000, **1**, 17.
49 V. I. Danilov, T. van Mourik and V. I. Poltev, *Chem. Phys. Lett*., 2006, **429**, 255.
50 P. U. Civcir, *THEOCHEM*, 2000, **532**, 157.
51 A. F. Jalbout, B. Trzaskowski, Y. Xia, Y. Li, X. Hu, H. Li, A. El-Nahas and L. Adamowicz, *Chem. Phys*., 2007, **332**(2–3), 152.
52 J. M. Gulbis, M. F. Mackay, G. Holan and S. M. Marcuccio, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 1993, **49**, 1095.
53 I. Dyer, J. N. Low, P. Tollin, H. R. Wilson and R. A. Howie, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun*., 1988, **44**, 767.
54 T. Kovacs, L. Parkanyi, I. Pelczer, F. Cervantes-Lee, K. H. Pannell and P. F. Torrence, *J. Med. Chem*., 1991, **34**, 2595.
55 X. Hu, H. Li, W. Liang and S. Han, *J. Phys. Chem. B*, 2004, **108**, 12999.
56 J. P. Henderson, J. Byun, J. Takeshita and J. W. Heinecke, *J. Biol. Chem*., 2003, **278**, 23522.
57 W. Saenger, *Principles in Nucleic Acid Structure*, Springer Verlag, New York, 1984, p. 88.
58 D. M. Close, C. E. Crespo-Hernandez, L. Gorb and J. Leszczynski, *J. Phys. Chem. A*, 2006, **110**(23), 7485.
59 (*a*) G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond*, Oxford University Press, New York, 1999; (*b*) S. K. Panigrahi and G. R. Desiraju, *J. Biosci*., 2007, **32**(4), 677.
60 I. Kulakowska, M. Geller, B. Lesyng, K. L. Wierzchowski and K. Bolewska, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth*., 1975, **407**, 420.
61 R. D. Brown, P. D. Godfrey, D. McNaughton and A. P. Pierlot, *J. Am. Chem. Soc*., 1988, **110**, 2329.
62 Y. He, C. Wu and W. Kong, *J. Phys. Chem. A*, 2004, **108**, 943.
63 S. K. Kim, W. Lee and D. R. Herschbach, *J. Phys. Chem*., 1996, **100**, 7933.
64 N. J. Kim, Y. S. Kim, G. Jeong, T. K. Ahn and S. K. Kim, *Int. J. Mass Spectrom*., 2002, **219**, 11.
65 F. Piuzzi, M. Mons, I. Dimicoli, B. Tardivel and Q. Zhao, *Chem. Phys*., 2001, **270**, 205.
66 H. Kang, K. T. Lee and S. K. Kim, *Chem. Phys. Lett*., 2002, **359**, 213.
67 S. Arai, T. Chatake, T. Ohhara, K. Kuruhara, I. Tanaka, N. Suzuki, Z. Fujimoto, H. Mizuno and N. Niimura, *Nucleic Acids Res*., 2005, **33**, 3017.
68 C. M. Marian, F. Schneider, M. Kleinschmidt and J. Tatchen, *Eur. Phys. J. D*, 2002, **20**, 357.
69 A. L. Sobolewski, W. Domcke, C. Dedonder-Lardeux and C. Jouvet, *Phys. Chem. Chem. Phys*., 2002, **4**, 1093.
70 (*a*) E. S. Kryachko, M. T. Nguyen and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 2001, **105**(10), 1934; (*b*) O. V. Shishkin, L. Gorb and J. Leszczynski, *Int. J. Mol. Sci*., 2000, **1**, 17; (*c*) V. I. Danilov, T. Mourik and V. I. Poltev, *Chem. Phys. Lett*., 2006, **429**, 255.
71 H. Yekeler and D. Özbakir, *J. Mol. Model*., 2001, **7**(4), 103.
72 (*a*) T. Frigato, D. Svozil and P. Jungwirth, *J. Phys. Chem. A*, 2006, **110**, 2916; (*b*) S. Kim, S. E. Wheeler and H. F. Schaefer, *J. Chem. Phys*., 2006, **124**(20), 204310.
73 S. X. Tian, C. F. Zhang, Z. J. Zhang, X. J. Chen and K. Z. Xu, *Chem. Phys*., 1999, **242**, 217.
74 M. Orozco, B. Hernández and F. J. Luque, *J. Phys. Chem. B*, 1998, **102**, 5228.
75 B. Blicharska and T. Kupka, *J. Mol. Struct*., 2002, **613**, 153.