J. Molecular Structure

FT-IR spectra of the anti-HIV nucleoside analogue d4T (Stavudine). Solid state simulation by DFT methods and scaling by different procedures

M. Alcolea Palafox[a,b], D. Kattan[a,b], N.K. Afseth[a]

[a] *Nofima AS - the Norwegian Institute of Food, Fisheries and Aquaculture Research, Osloveien 1, NO-1433 Ås, Norway (alcolea@ucm.es, nils.kristian.afseth@nofima.no)*

[b] *Departamento de Química-Física I, Facultad de Ciencias Químicas,Universidad Complutense, Madrid-28040, Spain*

Abstract

A theoretical and experimental vibrational study of the anti-HIV d4T (stavudine or Zerit) nucleoside analogue was carried out. The predicted spectra in the three most stable conformers in the biological active *anti*-form of the isolated state were compared. Comparison of the conformers with those of the natural nucleoside thymidine was carried out. The calculated spectra were scaled by using different scaling procedures and three DFT methods. The TLSE procedure leads to the lowest error and is thus recommended for scaling. With the population of these conformers the IR gas-phase spectra were predicted. The crystal unit cell of the different polymorphism forms of d4T were simulated through dimer forms by using DFT methods. The scaled spectra of these dimer forms were compared. The FT-IR spectrum was recorded in the solid state in the 400-4000 cm^{-1} range. The respective vibrational bands were analyzed and assigned to different normal modes of vibration by comparison with the scaled vibrational values of the different dimer forms. Through this comparison, the polymorphous form of the solid state sample was identified. The study indicates that d4T exist only in the ketonic form in the solid state. The results obtained were in agreement with those determined in related anti-HIV nucleoside analogues.

Keywords: Scaling procedures, d4T, Stavudine, anti-HIV, FT-IR

1. Introduction

The WHO estimates a total of 36.7 million people living with HIV [1]. This infection urges the scientific community to explore and develop novel and more potent antiviral agents. Different strategies have been employed by researchers in the production of novel antiviral drugs. The compounds synthesized fail within different categories depending on the target of interaction in the HIV replicative cycle [2]. One of the targets that have been envisaged most intensively is reverse transcription, catalysed by reverse transcriptase (RT). The RT associated with HIV is actually the target for three classes of inhibitors: nucleoside RT inhibitors (NRTIs), nucleotide RT inhibitors (NtRTIs), and non-nucleoside RT inhibitors (NNRTIs). The nucleoside-based antiviral agents are synthetic agents whose structure is similar to the naturally occurring nucleosides of DNA and RNA. They are considered as the most common antiviral agents with immune-modulator activity.

There are several NRTIs that have been formally approved for the treatment of HIV infection. One of the most used is d4T. Stavudine (d4T or 3′-deoxy-2,3´-didehydro thymidine) has been one of the first nucleosides synthesized at the Michigan Cancer Foundation [3], and it was approved by the FDA for clinical use again the HIV virus. This anti-viral agent belongs to the NRTIs with also a broad variety of others pharmaceutical behaviors as anti-tumor and antibiotic agents [4,5]. It is widely prescribed in resource-limited settings [6-8], and it is an effective and inexpensive antiretroviral drug, although no longer recommended by WHO for first-line treatment due to toxicity concerns [9-11]. Because of the high cost of alternative drugs, it has not been feasible to replace d4T in most adults [12]. Thus, a new combined antiretroviral therapy (cART), in which d4T was replaced by AZT after 24 weeks of therapy have been proposed [13,14]. Alternatively, several d4T derivatives are being analyzed as possible drugs [15]. Because 2′,3′-dideoxynucleosides (ddNs) are the most important class of nucleosides active against HIV, and d4T is one of the simple structures, it is the interest of the present manuscript.

The bio-activity of d4T compound can be explained by considering that as d4T is attached to the DNA chain, no further growth can take place in the DNA chain due to the absence of the hydroxyl group (O3′-H group) in d4T. This –OH group is necessary for the triphosphorylated nucleotides to grow in the DNA chain. Although d4T is not triphosphorylated, the active form of d4T takes place in the cells by triphosphorylation just like in natural nucleosides and hence, it inhibits the replication of HIV.

Many studies related to the activity of d4T [16] as compared with other prodrugs have been reported [17-19], but little appears on the structure and spectroscopy of this compound. For this purpose, in a previous study the conformational analysis of the d4T structure [20] and other related prodrugs were carried out using MP2 and DFT methods [21-23]. However, the assignments of all the fundamental vibrations of the experimental spectra have not yet been reported, highlighting the importance of the present study.

The main aims of the present study were: (i) to simulate the IR spectrum of d4T in gas phase through the population of the three most stable conformers and to carry out an accurate characterization of all the calculated wavenumbers; (ii) to describe the methodology and accuracy of the different procedures to scale the wavenumbers; (iii) to simulate the crystal unit cell of the three polymorphism forms of d4T by a dimer form using three DFT methods; (iv) to compare the simulated scaled IR spectra of these dimer forms with that recorded by us in the solid state; (v) to identify the dimer form that appears in the solid state sample and to interpret in detail its experimental IR spectrum; (vi) to select which of the B3LYP, X3LYP and M062X DFT methods is the more appropriate to predict the spectra.

2. Experimental

D4T of spectral grade was obtained from M/s Aldrich Chemicals (Milwanke, WI USA). The sample was obtained in powder form, without specification of the crystal form, and it was used as such without any further purification. The d4T sample was prepared in KBr pellet (with 1 mg sample per 300 mg KBr), and FTIR spectra were recorded using a Tensor 27 spectrometer (Bruker Optik GmbH, Germany). The measurements were done in transmission mode in the spectral region from 4000-400 cm^{-1}. Before every sample a background measurement was collected to account for variation in water vapor and CO_2. The resolution was 1 cm^{-1} and the spectrum was an average of 40 scans. The acquisition of spectral data was performed by the OPUS LAB v 6.1 software.

3. Computational Methods

Mainly three levels of theory were used for geometry optimisations of the monomer form: B3LYP/6-31G(d,p), B3LYP/6-311++G(3df,pd) and MP2/6-31G(d,p), while for optimisations and vibrational wavenumbers of the three dimer forms were: B3LYP/6-31G(d,p), X3LYP/6-31G(d,p) and M062X/6-31G(d,p). The vibrational wavenumbers obtained by these DFT methods appear very accurate and they are available in the Gaussian 09 [24] program package. B3LYP is slightly better than the M06L method, in accordance to a previous work carried out by us in a benzene derivative [25]. In general, DFT methods have been applied properly in many studies of drug design [26-29], in canonical nucleosides [30,31] and in nucleosides analogues [32,33]. In particular, the B3LYP/DFT method was chosen mainly because of the following three reasons: (i) this method is one of the most accurate and mostly used today in many biological systems [34,35], including Watson-Crick DNA base mispairs [36-39], i.e. it is the most standard method [40-45]; (ii) many studies have reported that the wavenumbers obtained with this method are better than those determined by other more costly computational ones. Especially the vibrational frequencies of nucleic bases calculated with this method are better than those obtained with MP2 and HF methods [46-49]; and (iii) scaling equations (LSE) are available with this method to reduce the error between theory and experiment [25,46-49].

Without imposing molecular symmetry constraints in the optimization process, the final geometry was obtained by minimizing the energy with regard to all geometrical parameters. For this optimization the Berny algorithm in redundant internal coordinates was used under the TIGHT convergence criterion. The default fine integration grid was employed. All the computed structures are true minima proved by no negative wavenumbers. The optimized structural parameters were used in the vibrational frequency calculations within the harmonic approximation at the same level of theory used for the optimized geometry. The vibrational assignments were interpreted by using the animation option of GaussView 5.1 graphical interface for Gaussian09 program [50]. The zero-point vibrational energies (ZPE) were included in the calculation of the relative energies. The determination of these ZPE was performed with the

wavenumbers maintained unscaled. All quantum chemical calculations have been carried out on the Quipu computer of the "Centro Computacional de la Universidad Complutense de Madrid".

Although several basis sets were used in the present research, only the results with 6-31G(d,p) basis set are discussed, mainly due to small improvements reached with its increment. With these methods and basis set we have previously studied the geometry and thermodynamical parameters of a series of nucleosides analogues [51-54]. Natural atomic charges were computed with the NBO procedure [55]. These atomic charges are of the most accurate today to correlate properties, and they were determined with the keyword POP = NPA.

4. Interaction Energies

The interaction energies were computed on the different polymorphism forms of d4T. These energies were calculated on the B3LYP optimized dimer structures with the 6-31G(d,p) basis set, including the ZPE correction. The interaction energies were also corrected for basis set superposition error (BSSE) using the standard counterpoise (CP) procedure [56-58] as follows:

The total CP corrected interaction energy, ΔE_{AB}^{CP}, between the molecules A and B of the different dimer forms of the nucleoside d4T is calculated according to:

$$\Delta E_{AB}^{CP} = E^{int}(AB) \; + \; E^{def}(AB)$$

where $E^{int}(AB)$ is defined as:

$$E^{int}(AB) = E_{AB}^{AB}(AB) - E_{A}^{AB}(AB) - E_{B}^{AB}(AB)$$

where $E_{AB}^{AB}(AB)$ stands for the electronic energy at the optimized geometry of the whole system (dimer form), and $E_{A}^{AB}(AB)$ (or $E_{B}^{AB}(AB)$) is the electronic energy of the isolated subsystem A (or B) (frozen) in the entire system AB (dimer).

The deformation energy $E^{def}(AB)$ is defined as: $E^{def}(AB) = E_{A}^{def}(AB) + E_{B}^{def}(AB)$

where the deformation energy of monomer, X (≡ molecule A or B) is:

$$E_{X}^{def}(AB) = E_{X}^{X}(AB) - E_{X}^{X}(X)$$

The subscripts represent the molecular system and the superscripts refer to the type of calculation that is done: with the basis set on the whole system (AB) (dimer form), or the basis set of the monomer (X) (monomer form). The parentheses show whether the computation is carried out at the optimised geometry of the whole system (AB), or at the monomer optimised geometry (X).

5. Results and discussion

5.1 Geometry optimization in the isolated state

The structure of d4T, as in other nucleosides, is dominated by three fragments, the thymine ring, the furanose ring and the 5′-hydroxymethyl moiety. The N1 atom of the thymine ring is attached to the C1′ atom of the furanose ring. The labelling of the atoms on d4T and dT molecules as well as the most important endocyclic and exocyclic torsional angles are shown in Scheme 1. They are in accordance to Saenger's notation [59]. In nucleosides (Ns) there are several structural parameters that are considered to characterize their different conformations:

- The exocyclic β(C4′-C5′-O5′-H5′) torsional angle, which defines the orientation of H5′ related to the furanose ring.
- The exocyclic γ(C3′-C4′-C5′-O5′) torsional angle which defines the orientation of O5′ related to the furanose ring.
- The glycosylic χ(O4′-C1′-N1-C2) torsional angle, which defines the perpendicularity of the thymine ring plane related to the furanose ring plane, appearing in three main orientation forms: *anti*, *high-anti*, and *syn*.
- The pseudorotational furanose pucker angle P, defined in the bottom of Table 1.

In a previous study, and through 3D potential energy surface, 25 local minimum by MP2 and several DFT methods were determined [20]. From them, the three most stable conformers in the biological active *anti*-form are plotted in Fig. 1, and the most characteristic parameters are included in Table 1. The highest stability of these three conformers is due to an intramolecular C6H···O5′ hydrogen bond. The CH groups as donor of protons play a significant role in the functioning of biologically important compounds through intermolecular [60] and intramolecular [61] C-H···O H-bonds.

The approximate estimated energy barrier conformer C1 → C3, through β direction in the PES, is calculated to ca. 12.0 kJ/mol. The molecular structure of the best conformer of the thymidine (dT) molecule (conformer C1) is also included in Fig. 1. The torsional angle β=-176.1º of this conformer of dT differs remarkably to that optimized in the best conformer of d4T, β= 63.8º. The orientation of β≈180º of dT is the most appropriate for the first phosphorylation step, which is consistent with that of the dT5′-triphosphate bound to HIV RT [52]. However, the value of β in d4T appears much less feasible, and it can be one of the reasons why the proportion of phosphorylated compounds is in general very small in the majority of prodrugs [52]. In addition, a lower value of P and ν_{max} in d4T than in dT, indicate a worse flexibility of the sugar ring in d4T than in dT, i.e. a worse adaptation to the different enzymatic reactions.

The main structural difference between d4T and dT is the occurrence of the C2′=C3′ double bond. This feature leads to a nearly planar furanose ring with a high extent of rigidity. In addition, the higher presence of low-lying vibrational modes in dT than in d4T is indicative of higher conformational flexibility in dT than in d4T.

5.2 Geometry in the solid state.

Depending on the experimental conditions of the crystal formation, several polymorphism forms appear in the solid state of d4T. Thus, several kinds of crystal forms have been observed, as seen in Table 2: orthorhombic [62], rhombic [63,64], monoclinic [62,65] and triclinic [62,66]. The notation used for the corresponding dimers is as follow: dimer V, dimer I (symmetric dimer), and dimer G or H. The optimized dimer structures of d4T with the B3LYP method are plotted in Fig. 2. For simplicity, the main selected characteristic parameters are only collected in Table 2 for B3LYP, X3LYP and M062X methods. In general, comparable values were obtained by these methods.

Dimer V is formed by an intermolecular hydrogen bond between the hydroxyl oxygen of the first monomer and the hydroxyl hydrogen of the second monomer. Both monomers of the dimer appear in the conformation C3.

Dimer I is obtained by intermolecular H-bonds of the two thymine rings with both molecules in conformation C1: (i) The 1st H-bond between the oxygen atom in the C2=O group of the first monomer (A) and the hydrogen atom in the N3-H group of the second monomer (B), and (ii) the 2nd H-bond between the oxygen atom in C4=O group of the second monomer and the hydrogen atom of the N3-H group of the first monomer.

Dimer G can be formed with conformers C1 (Fig. 1) and with conformers C1 and A2 [20]. In the first type with conformers C1, the dimer is obtained by a rotation of the thymine ring of the second monomer that leads to the formation of two H-bonds: (i) The 1st one between carbonyl oxygen C2=O of the first monomer (A) and the amino hydrogen N3-H of the second monomer (B); and (ii) The 2nd one is formed between the oxygen in the C2=O group of the second monomer and the amino hydrogen N3-H of the first monomer.

The second type of dimer G is formed with conformers C1 and A2. It has the same H-bonds than the first type with conformers C1. Related to the first type, this dimer is obtained by a rotation of the furanose ring in monomer 2.

The calculated interaction energy is slightly higher in dimer I than in dimer G with monomer C1, which has the lowest value, as seen in Table 3. For simplicity, the values were only shown for the M062X method. The highest value corresponds to dimer V. It is because the deformation energy $E^{def}(AB)$ is slightly larger in dimer I than in dimer G (monomer C1). As expected, dimer V has the lowest deformation energy. This is because only one H-bond is formed and it is through the O5′H group. Therefore, the CP corrected energy is the highest (in negative value) in dimer V.

In these studies, as well as in others X-ray studies [67,68] of related nucleosides, the torsional angle χ is always in *anti* orientation. The optimized thymine ring structure of the molecule is almost planar, in agreement with its experimental X-ray data available [69], and with that determined theoretically [70], as well as that reported by gas-phase electron diffraction in uracil molecule [71].

The optimized structural parameters of these dimer forms were compared with those reported by X-ray, as shown in Fig. 2 and Table 2. The concordance obtained of our calculated values with the experimental X-ray data confirms the validity of our calculations performed on d4T.

5.3 Procedures to improve the frequencies: the scaling

The wavenumbers calculated with the most accurate quantum chemical methods and with the largest basis set don't need to be corrected (scaling). However, these methods are too expensive to be applied for the calculation of the vibrational spectrum of a polyatomic molecule of even a modest size. Thus, one may be forced to work at small level, with the consequence of the deficiency of the theoretical approach, and thus it is expected a large overestimation of the calculated vibrational wavenumbers. However, this error can be

remarkably reduced with the use of transferable empirical parameters (corrections) for the computed wavenumbers, i.e. the scaling. Therefore, with this procedure of scaling it is possible to determine the vibrational wavenumbers of useful accuracy to be obtained from procedures of only modest computational cost.

The theoretical methods predict the vibrational spectra of molecules in the isolated state, but if we have the experimental values in the solid state, the difference between theory and experiment can be very large in the modes involved in intermolecular H-bonds. Thus, to avoid a mistake in the assignment it is recommended to use this methodology and to carry out these two points:

- To simulate with a theoretical model this solid state. In a simple way, in many cases it is enough with the simulation of the crystal unit cell. In the present study with d4T it was carried out through a dimer form.
- To use very accurate scaling procedures for scaling the wavenumbers [25,47-49].

Five procedures of scaling the computed wavenumbers, as described below, can be used satisfactorily depending on the accuracy required:

(i) By using **two overall scale factors procedure (OSF)**: it is the simplest and easiest way and thus it is the procedure generally used in the bibliography to scale the wavenumbers [49].

The equation used is: $\nu^{scal.} = a \cdot \nu^{cal.}$

At the B3LYP/6-31G(d,p) level the following factors are used:

$a = 0.9614$ for high wavenumbers and

$a = 1.0013$ for low values.

However, with this procedure there is a large error in the scaled values, although much lower than without scaling. This error can in some cases make assignment difficult, whereas in other cases contribute to a clear and accurate assignment.

(ii) By using **a linear scaling equation procedure (LSE)**: a linear relationship is established between the calculated and experimental wavenumbers.

$\nu^{scal.} = a + b \cdot \nu^{cal.}$

With the 6-31G(d,p) basis set, the following factors are used:

levels	a	b
B3LYP	29.4±5.3	0.9475±0.0033
X3LYP	27.9±5.2	0.9462±0.0033
M062X	22.7±6.0	0.9417±0.0037

Another equation can also be used with similar result:

$\nu^{scal.} = 34.6 + 0.9447\ \nu^{cal.}$ at B3LYP/6-31G(d,p) level [48].

For the calculation of the “a” and “b” coefficients, this procedure requires the previous computed frequencies of a simple and related structure in which the gas phase vibrational

frequencies have been reported. For d4T, the values calculated in uracil molecule [48] at the same level that in the present study were used. This is connected to the transferability of these equations to related molecules. The LSE procedure developed by one of the authors represents a compromise between accuracy and simplicity, especially for DFT methods, and therefore it was one of the methods used for scaling the wavenumbers, as shown in Table 4.

For the LSE procedure, the linear regression between the computed and experimental frequencies obtained at the B3LYP/6-31G(d,p) level in the uracil molecule leads to a correlation coefficient of 0.9996 with a standard deviation of 16.2. With a larger basis set, such as 6-311++G(3df,pd) the correlation coefficient is slightly improved, i.e. 0.9998, and the standard deviation slightly reduced, i.e. 15.8, which is very close to our results. Thus, the errors obtained in the predicted wavenumbers of the present investigation were very small, the mean deviation was 10 cm^{-1} (1%). These errors are very close to those obtained in other molecules studied earlier by us [48]. For simplicity, Table 4 only shows the scaled frequencies calculated with this LSE procedure.

(iii) By using **two linear scaling equation procedure (TLSE)**: a small improvement compared to the LSE procedure can be reached if two equations, one for high wavenumbers (4000-1500 cm^{-1}) and another for low values (1500-0 cm^{-1}), are used instead of only one equation.

$$\nu^{scal.} = a + b \cdot \nu^{cal.}$$

With the 6-31G(d,p) basis, the following factors are used:

	4000 – 1500 cm^{-1}		1500 – 0 cm^{-1}	
levels	a	b	a	b
B3LYP	34.7±16.0	0.9429±0.0056	3.3±5.9	0.9813±0.0064
X3LYP	27.3±16.6	0.9436±0.0058	3.1±6.0	0.9787±0.0065
M062X	-51.4±22.0	0.9633±0.0076	5.7±6.0	0.9670±0.0064

(iv) By using **a polynomic scaling equation procedure (PSE):** the relationship established between the calculated and experimental wavenumbers was on a second order polynomic form.

$$\nu^{scal.} = a + b_1 \cdot \nu^{cal.} + b_2 \cdot (\nu^{cal.})^2$$

With the 6-31G(d,p) basis, the following factors are used:

levels	a	b_1	$b_2 \cdot 10^6$
B3LYP	3.2±7.4	0.9904±0.0103	-11.2±2.6
X3LYP	3.9±7.6	0.9856±0.0106	-10.3±2.7
M062X	17.3±10.8	0.9506±0.0150	-2.3±3.8

(v) By using **specific scale factors for each mode (SCFEM)**: this procedure requires a previous characterization of the calculated wavenumbers. A specific scaled factor deduced from a related structure (in which the gas phase wavenumbers have been reported) is

subsequently is applied to each specific mode. This procedure leads to the lowest errors, but it requires much more effort, and with a large basis set the improvement is very small. This procedure is recommended only for modes with large anharmonicity.

5.4. Vibrational frequencies

In the isolated state the scaled IR spectrum of conformers C1 and C3 are reproduced as shown in Figs. 3-5. The spectra are divided in two regions: 3700-2800 cm^{-1} (Fig. 4) and 1800-350 cm^{-1} (Fig. 5). A detailed study of the vibrational assignments of these monomers is shown in Tables 4-6. For simplicity, the values obtained by X3LYP and M052X were omitted in the Tables (as well as in the discussion) and only the values by B3LYP and for conformers C1 and C3 are shown and discussed. The analysis of the vibrations is divided in three sections: those corresponding to the thymine ring (Table 4), those of the furanose ring (Table 5) and those corresponding to the CH_3, CH_2 and OH groups (Table 6).

Columns 1-5th of Tables 4-5 correspond to the calculated harmonic wavenumbers ($\nu^{cal.}$, cm^{-1}), scaled wavenumbers ($\nu^{scal.}$) using the LSE procedure, relative IR intensities (A, %), force constants (f, mDyne/Å), and characterization obtained in the isolated state for conformer C1, respectively, while the values for conformer C3 are shown in columns 6-7th. For the normal modes of the uracil ring the notation of ref. [48] was used.

The simulated IR spectrum in the gas phase (Fig. 3) was determined through the population of the three most stable conformers (87.5% of C1, 6.5% of C2 and 5.8% of C3) calculated at the MP2/6-31G(d,p) level and with the Boltzmann's equation. The contribution of the remaining conformers is less than 3%, and therefore it can be omitted. As observed, most of the spectrum corresponds to conformer C1, which appears analysed in detail in Tables 4-6. This spectrum could be compared with the experimental gas/Ar matrix IR spectra when available.

In the solid state the computed vibrational data obtained by B3LYP for dimers V, G and I, and in the thymine ring, furanose ring and in the CH_3, CH_2 and OH groups are shown in Tables 7-9. The analysis of the spectra of these three dimers show that they are very close to those of the monomer, but with noticeable differences in the bands involved in intermolecular H-bonds. The calculation with the X3LYP and M062X methods were also carried out, but the data was omitted in these tables for simplicity. Comparing the scaled spectra of these dimer forms with the experimental IR spectra, the dimer which corresponds to the experimental spectra of the solid state sample was clearly identified. Because it is dimer V, the discussion of the vibrations below mainly refers to this dimer.

In the dimer forms two wavenumbers appear for each vibration, each one corresponds to one molecule of the dimer. Of these two wavenumbers, that wavenumber printed in bold corresponds to that with the highest IR intensity, which is the only one considered in the discussion below.

5.5 Analysis of the vibrational modes

The main characteristics of the spectra are discussed in the following sections. As mentioned above, d4T consists of three fragments: the thymine ring, the furanose ring and the 5′-hydroxymethyl moiety. Some of the molecular vibrations are expected to be specific of each fragment and some other vibrations involve atomic displacements of more than one fragment. Based on this, the fundamental vibrations are therefore analysed. Only the most characteristic modes are discussed in detail. For the remaining modes, Tables 7-9 are self-explanatory.

5.5.1 Thymine ring vibrations

The characterization of the atomic displacements corresponding to the thymine ring normal modes was reported previously by us [70], and it was used as reference in the present study together with the ring normal modes of the uracil molecule [48].

NH modes. The stretching vibrations of N-H and C-H groups are expected in the 2950-3500 cm^{-1} range, with the NH stretching mode in the higher wavenumber region. In thymine, two strong bands have been reported [72] at 3480 and 3434 cm^{-1} in the N-H region. These two bands are insensitive to the substitution on the uracil ring by a methyl group in the 5-position, and thus they appear very close to those of uracil, at 3485 and 3435 cm^{-1}. The spectral positions of these bands were accurately predicted by us in thymine [70] and assigned to the N1-H and N3-H bond, respectively. In the low temperature argon matrix IR spectrum of uridine [73] a band appears at 3482 cm^{-1}, very close to that of uracil, that was assigned to the N3-H stretching. In monomer C1 of d4T, the stretching scaled wavenumber corresponding to this N3-H group (mode **29**, ref. [48]) appears at 3452 cm^{-1}, but in the solid state we have doubts in its identification. In dimer V it is predicted at 3440 cm^{-1} with medium intensity which was not detected because it could be inside of the broad band centered at 3481.6 cm^{-1} and assigned to the ν(OH) mode. Another possibility is to relate the predicted wavenumbers at 3198 cm^{-1} (dimer G) and 3172 cm^{-1} (dimer I), the strongest of the spectrum in these dimers, to the band observed at 3153.8 cm^{-1}. However, this last possibility appears in contradiction with the lacking assignment of the observed band at 3481.6 cm^{-1} by these dimers. Thus, we consider that the N3-H group is involved in intermolecular H-bonds with other molecules of the crystal, which was not simulated in our simplified model with a dimer. This H-bond shifts the wavenumber of ν(N3-H) at about 3170 cm^{-1}, as it is observed in the dimers G and I.

The N3-H in-plane bending (mode **20**) appears strongly coupled with the δ(C6-H) mode. Thus, major contributions of this mode are predicted in the scaled wavenumbers at 1366 and 1352 cm^{-1} in monomer C1, and at 1354 and 1351 cm^{-1} in monomer C3. As expected, this strong coupling remains in dimer V. Thus, the scaled wavenumbers at 1369, 1368, 1353 and 1352 cm^{-1} can be assigned to this mode. The % contribution of δ(C6-H) and δ(N3-H) modes is almost the same in the wavenumbers at 1369 and 1368 cm^{-1}. These predicted values are in accordance to the experimental bands at 1362.3 and 1343.8 cm^{-1}.

The out-of-plane N3-H bending (mode **9**) appears as an almost pure mode and it is predicted at 685 cm^{-1} (monomer C1) and at 683 cm^{-1} (monomer C3). In addition, a small

contribution of this mode appears in several wavenumbers below 600 cm^{-1}. In the solid state this mode is expected to be scaled at similar wavenumber and also as a pure mode. Thus, the scaled values with medium IR intensity at 685 and 684 cm^{-1} in dimer V correspond to this mode, which is in good accordance to the IR band with medium intensity at 695.5 cm^{-1}.

C6-H modes. The characteristic region of the stretching mode is 3000-3100 cm^{-1}. In the IR spectra of pyrimidines it is predicted with such a low intensity that it is difficult to distinguish it from the background. Thus, in thymine it was detected at 3069 cm^{-1} with very weak IR intensity. In the different conformers of d4T the stretching ν(C6-H) (mode **27**) was calculated in the 3104-2849 cm^{-1} range, and especially in conformers C1 it appears at 3216.7 cm^{-1} (scaled at 3073 cm^{-1}) as very weak band. In conformer C3 it was scaled at 3065 cm^{-1}. This mode is to a small degree affected by the dimer formation, appearing at almost the same wavenumber and intensity. These values should thus be very close to those determined in the dimers. The scaled wavenumbers at 3077 and 3062 cm^{-1} (calculated at 3221 and 3205 cm^{-1}, respectively) correspond to this mode, and they were related to the experimental IR band at 3032.5 cm^{-1}.

The bending δ(C6-H) (mode **22**) appears strongly coupled with the δ(N3-H) mode. Thus, a high contribution of this mode is determined in the scaled wavenumbers at 1366 and 1352 cm^{-1} of monomer C1. This strong coupling remains in dimer V, and thus a large contribution of this mode is characterized in the scaled wavenumbers at 1369, 1368, 1353 and 1352 cm^{-1}, as discussed above.

The out of plane bending γ(C6-H) (mode **15**) appears, by contrast, as an almost pure mode. It was predicted with weak IR intensity in the monomers C1 and C3 at 928 and 930 cm^{-1}, respectively. In dimer V it was also characterized as an almost pure mode with weak IR intensity in the scaled wavenumbers at 946 and 922 cm^{-1}, and they were related to the experimental band with very weak intensity at 918.4 cm^{-1}.

C-H modes in CH_3. Because this group is not affected by H-bonds, their modes appear at similar wavenumbers in the monomer as in the dimer forms. The symmetric $\nu_s(CH_3)$ and antisymmetric $\nu_{as}(CH_3)$ stretching are predicted with weak IR intensity, and thus they are difficult to detect in the experimental spectra. However, the predicted wavenumber at 2997 cm^{-1} in the monomer form was nicely related to the very weak experimental band at 2985.5 cm^{-1}.

The IR bands at 1455/1451 and 1437/1433 cm^{-1} in thymine were assigned as $\delta_{as}(CH)$ in CH_3, while the band at 1395 cm^{-1} was related to the symmetrical mode. In d4T, the antisymmetric mode of the CH_3 group was predicted (scaled) at 1437 and 1048 cm^{-1} in monomer C1, and at 1438, 1050 cm^{-1} in dimer V, and they were nicely related to the experimental band at 1053.4 cm^{-1}.

The γ_{as}(C-H) out-of-plane mode in CH_3 was scaled in the monomer form at 1014 cm^{-1} with weak IR intensity and at similar wavenumber in the dimers, and it was nicely associated to the very weak experimental band at 1010.6 cm^{-1}.

C=O modes. The stretching modes are known to be very sensitive to the polynucleotide conformations. The typical pattern of the absorption bands due to the C=O stretching vibrations of uracil and its derivatives is always very complex. There is a strong coupling with the C=C stretching mode, although weaker in the case of thymine than in other uracil derivatives, which permits a clear identification of the mode. The wavenumbers of the two strong IR bands calculated in conformer C1 at 1746 (1740 cm^{-1} in C3) and 1729 cm^{-1} (1725 cm^{-1} in C3) are close to those observed in uracil and thymine (1759 and 1731 cm^{-1}) molecules, and they were assigned to the ν(C2=O) and ν(C4=O) modes, respectively. The IR intensity for the ν(C4=O) mode is higher than for ν(C2=O), which is in accordance to the experimental spectrum.

In the experimental spectrum of the solid state, the stretching ν(C4=O) (mode **25**) was related to the highest peak of the IR spectrum at 1691.7 cm^{-1}. By contrast, ν(C2=O) (mode **26**) was not assigned to an experimental band because it can be involved in the broad band centred at 1691.7 cm^{-1}. Thus, it has been deconvoluted in several peaks that appear at 1695.1, 1690.1, 1687.9, 1685.5 and 1683.3 cm^{-1}. One of these peaks could correspond to ν(C2=O). All these peaks can be explained by other intermolecular H-bonds/interactions of the C=O groups in the solid state. These peaks were not reproduced in our simplified model.

The out-of-plane vibration γ(C2=O) (mode **11**) was scaled at 747 cm^{-1} in the conformer C1 and at 744 cm^{-1} in conformer C3, while γ(C4=O) (mode 10) appears scaled at 757 cm^{-1} for both conformers. The IR intensity is predicted as weak for both modes. In dimer V they were predicted with weak/medium IR intensity at 749 cm^{-1} for γ(C2=O) and at 754 cm^{-1} for γ(C4=O), in accordance to the weak experimental bands at 760.2 and 765.7 cm^{-1}, respectively. A noticeable contribution of mode 11 was also observed in the scaled value at 743 cm^{-1} in dimer V, in excellent agreement with the IR band detected at 743.6 cm^{-1}.

C-N modes. The C-N stretching vibrations of the uracil ring are mainly described in the modes **17**, **18** and **21**. These modes appear coupled with δ(C1′-H) and with the in-plane or out-of-plane modes of the CH_3 group. They are to a little extent affected by the H-bonds, and thus their wavenumbers are almost the same in the monomer than in the dimer forms. Therefore, the scaled wavenumber with medium IR intensity at 1430 cm^{-1} in dimer V (at 1431 cm^{-1} in monomer) was well related to the experimental band with medium intensity at 1423.4 cm^{-1}. Also the scaled values at 1213 and 1143 cm^{-1} were accurately assigned to the experimental bands at 1226.9 and 1142.9 cm^{-1}, respectively. The calculated frequencies in the monomer form are in accordance to that reported by Gunasekaran *et al.* [74].

The stretching vibration in the glycosylic bond N1-C1′ is identified in the scaled wavenumber at 1103 cm^{-1} in dimer V (at 1098 cm^{-1} in monomer) and it appears at a medium/strong IR intensity. The experimental bands observed at 1115.1 and 1107.2 cm^{-1} were assigned to this mode.

The C-N in-plane bendings are calculated to spread out in many vibrations below 900 cm^{-1}. The main contribution appears ascribed to mode **16**, which was scaled at 771 cm^{-1} in dimer V and related to the experimental bands detected at 783.6 and 778.7 cm^{-1}.

5.5.2 Furanose ring vibrations

C-O modes. The stretching mode ν(C4′-O) appears with a high contribution in the scaled wavenumber at 1086 cm^{-1} in dimer V, but it is predicted at 1055 cm^{-1} in dimer G and I. This shift is due to the different values for the χ, β and γ torsional angles of dimer V as compared to dimers I and G, as shown in Table 2. The value of these torsional angles highly affects the wavenumbers of the furanose ring vibrations.

Others noticeable contributions of this stretching mode have been detected in the scaled wavenumbers at 983 and 973 cm^{-1} in dimer V. This mode appears associated with others ring vibrations that is difficult to detect. These values are similar to that calculated in AZT at 975 cm^{-1}, and to that reported by S. Gunasekaran *et al.* [75].

C-H modes. The wavenumber of the stretching ν(C-H) mode are scaled in the dimers in the 3104-2849 cm^{-1} range. Accordingly, in the experimental IR spectra of the solid state several bands observed in the 3112.9-2823.5 cm^{-1} range were assigned to this mode. The scaled value in the monomer is almost the same as in the three dimers. The C-H modes of the furanose ring are thus not affected by the intermolecular H-bonds.

The symmetric stretching mode in the C2′H=C3′H moiety was predicted at 3102 cm^{-1} and related accurately with the experimental band at 3112.9 cm^{-1}, while the asymmetric mode predicted at 3081 cm^{-1} was associated to the bands at 3097.7 and 3071.4 cm^{-1}.

The stretching modes corresponding to C1′-H and C4′-H groups and scaled at 2956 and 2851 cm^{-1} were also well related to the experimental bands at 2949.4 and 2823.5 cm^{-1}, respectively.

The in-plane bending δ(C-H) mode appears spread out in several vibrations in the 1650-950 cm^{-1} range, in good accordance with those observed in the experimental spectra. The symmetric mode of the ring is predicted at 1253 cm^{-1} in dimer V, while the asymmetric modes are found at 1131 and 1087 cm^{-1}, in accordance to the experimental bands observed at 1268.7, 1136.1 and 1091.9 cm^{-1}, respectively.

The main contribution of the asymmetric out of plane bending γ_{as}(C-H) in the C2′H=C3′H moiety is scaled in conformer C1 at 944 cm^{-1}, the same wavenumber as found in the dimers, in excellent concordance to that detected in the IR spectrum of 956.9 cm^{-1}. The symmetric mode is predicted at 691 cm^{-1}, but this was not observed experimentally.

5.5.3 CH_2OH group vibrations

OH modes. The presence of only one band in the experimental IR spectrum at 3481.6 cm^{-1} with strong intensity and assigned to ν(O5′-H), and the appearance of ν(C=O) bands with strong intensity indicates that this compound in the solid state appears only in the *keto* form. The stretching mode ν(O-H) was characterized and predicted in conformer C1 at 3628 cm^{-1}

and at 3659 cm^{-1} in conformer C3. However, bands were not detected at ca. 3600 cm^{-1} in the experimental IR spectrum of the solid state. The closest band appears at 3481.6 cm^{-1} as broad and with strong intensity. This shift in the wavenumber can only be due to this OH group appears intermolecular H-bonded. In dimer V the scaled wavenumber at 3614 cm^{-1} corresponds to free OH, and the wavenumber at 3453 cm^{-1} to H-bonded OH. This last wavenumber is close to the experimental band at 3481.6 cm^{-1}, as shown in Table 9. The broadness of the experimental band could indicate that other vibrational modes appear close to this wavenumber, as in the present case with the ν(N3-H) stretching mode.

The δ(OH) in-plane bending appears strongly coupled with δ_{as}(C-H) in CH_2 mode, with almost the same contribution (%) of both modes. Thus, in monomer (free OH) it is predicted at 1373 cm^{-1} (Table 6), but due to the intermolecular H-bond it should appear at higher wavenumbers. Thus, in dimer V the predicted wavenumber at 1424 cm^{-1} (calculated at 1471 cm^{-1}) should be assigned to this mode, in accordance to the experimental band at 1407.7 cm^{-1}.

The γ(O-H) out-of-plane bending appears simulated at 459 cm^{-1} in the monomer and at 736 and 408 cm^{-1} in dimer V. The wavenumber at 408 cm^{-1} can be related to free OH and that at 736 cm^{-1} to H-bonded OH. Because the weak IR intensity predicted for this mode, it was not assigned to any experimental band.

C-O modes. The main contribution of the C5′-O stretching is predicted in dimer V at 1043 cm^{-1} and strongly coupled with ν_{as}(ring, CH) of the furanose ring. In AZT the stretching mode was calculated at 1035 cm^{-1}, in good agreement with our results.

The in-plane bending is also predicted as a noticeable contribution in the scaled wavenumber at 852 cm^{-1} in dimer V, and strongly coupled with CH_2 and the furanose ring vibrations. The experimental band observed at 851.9 cm^{-1} was accurately assigned to this mode.

The out-of-plane mode was also accurately predicted in dimer V at 660 cm^{-1} and good related to the experimental band at 661.1 cm^{-1}. This mode also appears strongly coupled with CH_2 and ring vibrations of the furanose and uracil rings.

C-H modes. The antisymmetric stretching ν_{as}(C-H) for the CH_2 group is scaled in dimer V at 2936 cm^{-1} with medium intensity, in good accordance to the experimental IR band observed at 2933.1 cm^{-1}. The symmetric mode was also accurately predicted at 2890 cm^{-1} and related to the experimental band at 2883.5 cm^{-1}.

The scissoring (β) mode is predicted at 1467 cm^{-1} with weak IR intensity and it could be related to the experimental IR band at 1467.1 cm^{-1} with strong intensity. The wagging (ω) mode is scaled with medium intensity at 1380 cm^{-1} in excellent accordance to the experimental band at 1383.8 cm^{-1}. The twisting (τ) mode was predicted at 1207 cm^{-1} with weak intensity and related to the very weak band observed at 1181.1 cm^{-1}.

The rocking (Γ) mode for the CH_2 group was predicted in monomer C1 at 878 cm^{-1} and at 886 cm^{-1} in dimer V with weak IR intensity, in good accordance to the weak IR band detected at 881.2 cm^{-1}.

6. Identification of the dimer

The first step after the analysis and assignments of the different bands of the experimental spectra is to know what polymorphic form appears in the solid state of the d4T sample. Because Aldrich Chemical has not mentioned the crystal form of the sample, it was identified here by comparing the IR scaled frequencies of the three dimer forms to the IR experimental ones. For this purpose, several bands of the experimental IR spectrum were selected as characteristics for this identification, as shown in Table 11. The experimental and theoretical Raman wavenumbers were also used to confirm the dimer selected. The two most characteristic bands from the experimental IR spectrum are the following:

The band at **3481.6 cm^{-1}**: It was selected because in ν(O-H) mode, the difference between the scaled wavenumbers in dimers G and I and the experimental one observed in the solid state is almost 145 cm^{-1}, but in dimer V it is only 28 cm^{-1}. In dimer V two scaled wavenumbers are predicted at 3614 and 3453 cm^{-1} for this mode. The value with the highest IR intensity at 3614 cm^{-1} is very close to that scaled in the monomer at 3628 cm^{-1}. Thus, the another frequency at 3453 cm^{-1} was selected as more appropriate because it is associated with H-bonds, and it is in accordance with the experimental IR band in the solid state at 3481.6 cm^{-1}.

The band at **1691.7 cm^{-1}**: It was mainly selected by the highest IR intensity observed experimentally, which is in accordance to that obtained theoretically in dimer V with IR (100%), and scaled at 1727 cm^{-1}. The remaining dimers have much lower IR intensity and they are far from the highest value.

These two bands appear to indicate that the solid state sample corresponds to dimer V. The remaining experimental bands of Table 11 also confirm this feature, although the difference is smaller.

7. Analysis of the rms errors

Table 11 shows the errors obtained using four of the scaling procedures recommended to scale the calculated wavenumbers and with the three DFT methods used with the 6-31G(d,p) basis set. For OSF scaling factors with the X3LYP and M062X methods have not yet been reported, and thus the rms error was not calculated. Due to the large error in the predicted wavenumber of the ν(OH) mode by dimers G and I, as compared to the experimental value at 3481.6 cm^{-1}, this wavenumber was not included in the calculation of the rms errors. In addition the experimental band at 3153.8 cm^{-1} was not included, due to the doubts in its assignment. In this way, we try to confirm in the remaining wavenumbers that dimer V corresponds to the

experimental solid state sample. The rms errors in dimers I and G were only calculated with the B3LYP method, because dimer V was the one selected as the solid state sample.

By comparing the results of Table 11, the following was observed:

(i) The LSE procedure leads to a remarkable improvement in the accuracy of the scaled wavenumbers as compared to OSF, although scaling equations are less available than overall scale factors.

(ii) The use of two scaling equations (TLSE) leads to a noticeable improvement of the results. It is the best procedure, at least with the results obtained in d4T. With the PSE procedure we also see an improvement as compared to LSE. Thus, these procedures are those recommended for an accurate scaling of the wavenumbers, and especially TLSE.

(iii) As expected, the simulation of the crystal unit cell by a dimer form leads to lower rms errors than those obtained with the monomer form C1. This feature is observed with the three DFT methods used.

(iv) The rms errors obtained in dimer V and B3LYP are always smaller than those obtained in dimers I and G. This feature is observed with the four scaling procedures used. The results with dimer G appear slightly better than those with dimer I.

(v) With the values obtained in d4T, the scaled wavenumbers determined with the X3LYP method are slightly better than those with B3LYP, although the difference is very small. Thus, both methods are recommended to be used to calculate the wavenumbers. With the M062X method, larger errors than with B3LYP and X3LYP are obtained. This feature is observed in the monomer as well as in the dimer forms, and with the LSE, TLSE and PSE procedures. Thus, this M062X/DFT method is not recommended to be used to predict the spectra. This feature is in accordance with that observed in a benzonitrile derivative with the M06L method [25].

8. CONCLUSIONS

The knowledge of geometry and vibrational spectra of d4T and related prodrugs could lead to a comprehensive understanding necessary for the development of effective drugs for the future. The most important findings of this study are the following:

(1) The simulated scaled IR vibrational spectrum of the anti-HIV nucleoside analogue d4T was carried out in detail for the first time in the isolated state. An accurate characterization was performed and the % contribution of the different modes to each calculated vibrational wavenumber was determined.

(2) In the isolated state the trend of the five first optimum conformers is the same by all the methods and levels used. The spectra corresponding to the monomers C1 and C3 are very similar, with the exception of the modes corresponding to the O5′-H groups, due to their different orientations. The two monomers C1 and C3 differ in intensity in the modes of

the groups involves in the torsional angles. In the IR spectrum the bands of conformer C3 are in general more intense than those of conformer C1.

(3) The predicted spectrum in gas phase was determined taking into account the contribution of the three most stable conformers. The final spectrum mostly corresponds to conformer C1, which was accurately analyzed.

(4) Four scaling procedures were used to improve the accuracy in the assignment of the wavenumbers. The scaled values with TLSE are the best. The rms errors with PSE are slightly improved compared to using LSE. In general, very low errors were obtained with the scaled values. Therefore, the assignments seem to be the most accurate today.

(5) The equilibrium geometry and the harmonic wavenumbers for the possible crystal unit cells of d4T were obtained by three DFT methods, and the results obtained were compared to the experimental values. The accuracy reached in the calculated wavenumbers with B3LYP and X3LYP appears similar, and thus both methods are recommended to be used to simulate the spectra

(6) Dimer V was assigned as the dimer that appears in the solid state sample, through the comparison of the scaled wavenumbers in the different dimers to the values obtained experimentally in the IR and Raman spectra. Several bands were selected for this purpose. In general, the experimental spectrum in the solid state suggests that the simulated spectrum corresponds to dimer V. Therefore, for the first time we describe here a procedure to identify the crystal form of a polymorphic compound through DFT methods and IR spectroscopy.

(7) The simulation of a dimer form remarkably improves the assignment carried out in the IR spectrum of d4T, with the sample in the solid state as compared to the monomer form.

(8) The stretching bands corresponding to N-H modes in dimer G and dimer I appear at lower wavenumbers than those corresponding to monomer C1 and dimer V. This fact is because the G and I dimers are stabilized by two N-H···O=C intermolecular hydrogen bonds.

(9) The stretching mode ν(O5′-H5′) appears at a higher wavenumber in C3 than in C1 monomer, in accordance to the calculated O5′-H5′ distance, lower for C3 than for C1. However the intramolecular O5′···H6 hydrogen bond distance is shorter for C3 than for C1 conformer. This fact is due to the repulsion between H5′ (sugar) and H6 (pyrimidine base) hydrogen atoms.

Acknowledgements

MAP wish to thank to BSCH-UCM PR26/16 for financial support. MAP is also thankful to B. Manchobas Pantoja for some of the calculations carried out in the present publication.

References

[1] W.H. Organization. http://www.who.int/gho/hiv/en/

[2] E.D. Clercq, Int. J. Antimicrobial agents 33 (2009) 307-320.

[3] J.P. Horwitz, J. Chua, M.A. Da Rooge, M. Noel, I.L. Klundt, J. Org. Chem. 31 (1) (1966) 205-211.

[4] R.A. Seaton, R. Fox, N. Bodasing, S.E. Peters, Y. Gourlay, Aids 17 (2003) 445.

[5] L.K. Naeger, N.A. Margot, M.D. Miller, Antimicrob. Agents CH 46 (2002) 2179.

[6] L. Taisheng, G. Fuping, L. Yijia, Z. Chengda, H. Yang, L. Wei, H. Yun, L. Hongzhou, X. Jing, H. Aiqiong, L. Yanling, T. Xiaoping, W. Hui, Z. Tong, G. Guiju, L. Junkang, Z. Xiaoying, W. Xinhua, S. Yongtao, B. Jinsong, L. Ling, W. Huanling, Chinese Med. J. 127 (1) (2014) 59-65.

[7] T. Li, Y. Dai, J. Kuang, J. Jiang, Y. Han, Z. Qiu, et al., PLoS One 3 (2008) e3918.

[8] D. Podlekareva, D. Grint, I. Karpov, A. Rakmanova, K. Mansinho, N. Chentsova, I. Zeltina, M. Losso, M. Parczewski, J.D. Lundgren, A. Mocroft, O. Kirk, HIV Med. 16 (9) (2015) 533-543.

[9] A.T. Brennan, M.-A. Davies, J. Bor, et al. AIDS 31 (1) (2017) 147-157.

[10] D. Nayak, A. Boxi, S. Ashe, et al. Materials Sci. Engineer. C-Materials Biol. Appl. 73 (2017) 406-416.

[11] E. Kampira, K. Dzobo, J. Kumwenda, et al. Omics-A J. Integrative Biol. 18 (7) (2014) 438-445.

[12] J.J. van Oosterhout, J. Mallewa, S. Kaunda, N. Chagoma, Y. Njalale, E. kampira, et al. PLoS One 7 (2012) e42029.

[13] N. Phanuphak, J. Ananworanich, N. Teeratakulpisarn, T. Jadwattanakul, S.J. Kerr, N. Chomchey, et al. Antiviral Ther. 17 (2012) 1521-1531.

[14] A. Singh, A. Hemal, S. Agarwal, et al. Int. J. Std Aids 27 (13) (2016) 1145-1152.

[15] K.A.S. Alameed, F. Shareef, Asian J. Chem. A 25 (17) (2013) 9789-9794.

[16] B. Sandhya, A.H. Hegde, J. Seetharamappa, Molec. Biol. Reports 40 (5) (2013) 3817-3827.

[17] M.M. Conradie, M. van de Vyver, E. Andrag, et al. Calcified Tissue Intern. 101 (4) (2017) 422-432.

[18] H. Cassim, K. Otwombe, E. Lazarus, et al. PLOS ONE 12 (7) (2017) e0180645.

[19] L.K. Naeger, N.A. Margot, M.D. Miller, Antimicrob. Agents Chemotherapy 46 (2002) 2179-2184.

[20] M. Alcolea Palafox, N. Iza, J. Molec. Struct.1028 (2012) 181-195.

[21] M. Alcolea Palafox, J. Talaya. J. Phys. Chem. B 114 (2010) 15199-15211.

[22] M. Alcolea Palafox, N.Iza, Phys. Chem. Chem. Phys. 12 (2010) 881-893.

[23] M. Alcolea Palafox, N. Iza, M. de la Fuente, R. Navarro, J. Phys. Chem. B 113 (2009) 2458-2476.

[24] Gaussian 09, Revision D.01, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian, Inc., Wallingford CT, 2009.

[25] M. Alcolea Palafox, Rachna Rastogi, Anupama, M. Jane Alam, Daisy Bhat, V.K. Rastogi. Asian J. Phys. 25 (2) (2016) 189-219.

[26] A.A. El-Sayed, A. Tamara Molina, M.C. Álvarez-Ros, M. Alcolea Palafox, J. Biomol. Struct. Dyn. 33 (4) (2015) 723-748.

[27] M. Alcolea Palafox. Chem. Informatics 1 (2:11) (2015) 1-13.

[28] M. Alcolea Palafox, P. Posada-Moreno, A.L. Villarino-Marín, C. Martinez-Rincon, I. Ortuño-Soriano, I. Zaragoza-García. J. Comp.-Aided Molec. Design 25 (2) (2011) 145-161.

[29] M. Alcolea Palafox, N. Iza. Struct. Chem. 24 (3) (2013) 967-980.

[30] Y.P. Yurenko, R.O. Zhurakivsky, M. Ghomi, S.P. Samijlenko, D.M. Hovorun. J. Phys. Chem. B, 111 (2007) 9655-9663; & 112 (2008) 1240-1250.

[31] Y.P. Yurenko, R.O. Zhurakivsky, S.P. Samijlenko, M. Ghomi, D.M. Hovorun. Chem. Phys. Lett. 447 (2007) 140-146.

[32] Y. P Yurenko,. R. O. Zhurakivsky, M. Ghomi, S.P. Samijlenko, D.M. Hovorun, J. Phys. Chem. B 111 (22) (2007) 6263-6271.

[33] A.G. Ponomareva, Y.P. Yurenko, R.O. Zhurakivsky, Tanja van Mourik, D.M. Hovorun. Phys. Chem. Chem. Phys. 14 (2012) 6787-6795.

[34] S.P. Samijlenko, Y.P. Yurenko, A.V. Stepanyugin, D.M. Hovorun. J. Phys. Chem. B, 114 (2010) 1454-1461.

[35] Ol'ha O. Brovarets', R.O. Zhurakivsky, D.M. Hovorun, Chem. Phys. Lett. 578 (2013) 126-132.

[36] Ol'ha O. Brovarets', R.O. Zhurakivskya, D.M. Hovorun, J. Biomol. Struct. Dyn. 33 (2015) 674-689.

[37] Ol'ha O. Brovarets', D.M. Hovorun. Phys. Chem. Chem. Phys. 16 (2014) 15886-15899; & 9074-9085.

[38] Ol'ha O. Brovarets', D.M. Hovorun, J. Biomol. Struct. Dyn. 33 (2015) 925-945; & 2297-2315.

[39] Ol'ha O. Brovarets', R.O. Zhurakivsky, D.M. Hovorun, J. Comput. Chem, 35 (2014) 451-466.

[40] V. Balachandran, M. Murugan, A. Nataraj, M. Karnan, G. Ilango, Spectrochim. Acta A, 132 (2014) 538-549.

[41] V. Arjunan, S. Thirunarayanan, G. Durga Devi, S. Mohan, Spectrochim. Acta A 150 (2015) 641-651.

[42] V. Arjunan, M.K. Marchewka, A. Raj, H. Yang, S. Mohan, Spectrochim. Acta A 135 (2015) 540-550.

[43] M. Szafran, A. Komasa, M. Anioła, A. Katrusiak, Z. Dega-Szafran, Vib. Spectrosc., 2016, 84, 92-100.

[44] J.S. Singh, Spectrochim. Acta A 117 (2013) 502-518.

[45] M. Szafran, A. Katrusiak, Z. Dega-Szafran, I. Kowalczyk, J. Molec. Struct. 1005 (1-3) (2011) 144-151.

[46] M. Alcolea Palafox, V.K. Rastogi, Spectrochim. Acta A 58 (3) (2002) 411-440.

[47] M. Alcolea Palafox, Int . J. Quantum Chem. 77 (2000) 661-684.

[48] M. Alcolea Palafox, N. Iza, M. Gil, J. Mol. Struct. (Theochem) 585 (2002) 69-92.

[49] M. Alcolea Palafox, Recent Res. Devel. in Physical Chem.2., Transworld Research Network, India, (1998) 213.

[50] GaussView 5.1, Gaussian Inc., Wallingford CT, 2009.

[51] M.C. Alvarez-Ros, M. Alcolea Palafox. Pharmaceuticals 7 (2014) 695-722.

[52] M. Alcolea Palafox. Struct. Chem. 25 (1) (2014) 53-69.

[53] M. Alcolea Palafox. J. Biomol. Struct. Dyn. 32 (5) (2014) 831-851.

[54] M.C. Alvarez-Ros, M. Alcolea Palafox. J. Molec. Struct. 1047 (2013) 358-371.

[55] J.E. Carpenter, F. Weinhold, J. Molec. Struct. (Theochem) 169 (1988) 41-62.

[56] D. Kattan, M. Alcolea Palafox, S. Kumar, D. Manimaran, H. Joe, V.K. Rastogi, Spectrochim. Acta A 123 (2014) 89-97.

[57] S.F. Boys, F. Bernardi, Molec. Phys. 19 (1970) 553-566.

[58] M. Alcolea Palafox, V.B. Jothy, S. Singhal, I. Hubert Joe, S. Kumar, V.K. Rastogi, Spectrochim. Acta A 116 (2013) 509-517.

[59] Saenger W. Principles in Nucleic Acid Structure; Springer Verlag Publishers: New York, 1984.

[60] Ol'ha O. Brovarets', Y.P. Yurenko, D.M. Hovorun. J. Biomol. Struct. Dyn. 33 (2015) 1624-1652; & 32 (2014) 993-1022.

[61] Y.P. Yurenko, R.O. Zhurakivsky, S.P. Samijlenko, D.M. Hovorun. J. Biomol. Struct. Dyn, 29(1) (2011) 51-65.

[62] M. Mirmehrabi, S. Rohani, M.C. Jennings, Acta Cryst. section C, 61 (2005) o695.

[63] G.V. Gurskaya, E.N. Tsapkina, Abstracts of the twelfth European Crystallographic Meeting, Moscow (1989), vol. 2, p 380.

[64] E.N. Tsapkina, The Molecular and Crystalline Structures of Nucleic analogs of Nucleic Acid Biosynthesis Substrates (in Russian), Dissertation for Doctorate I Chemical Sciences, Institute of Molecular Biology, Moscow (1989).

[65] W.E. Harte, Jr. J.E. Starret, Jr. J.C Martin, M.M. Mansuri, Biochem. Biophys. Res. Comm. 175 (1) (1991) 298.

[66] G.V. Gurskaya, A.V. Bochkarev, A.S. Zhdanov, N.B. Dyatkina, A.A. Kraevskii, Mol. Biol. 25 (2) (1991) 401.

[67] J.M. Gulbis, M.F. Mackay, G. Holan, S.M. Marcuccio, Acta Crystallog. C49 (1993) 1095.

[68] I. Dyer, J.N. Low, P. Tollin, H.R. Wilson, R.A. Howie, Acta Crystallog. C44 (1988) 767.

[69] K. Ozeki, N. Sakabe, J. Tanaka, Acta Crystallogr. B 25 (1969) 1038.

[70] V.K. Rastogi, Chattar Singh, Vaibhav Jain, M. Alcolea Palafox. J. Raman Spectrosc. 31 (2000) 1005-1012.

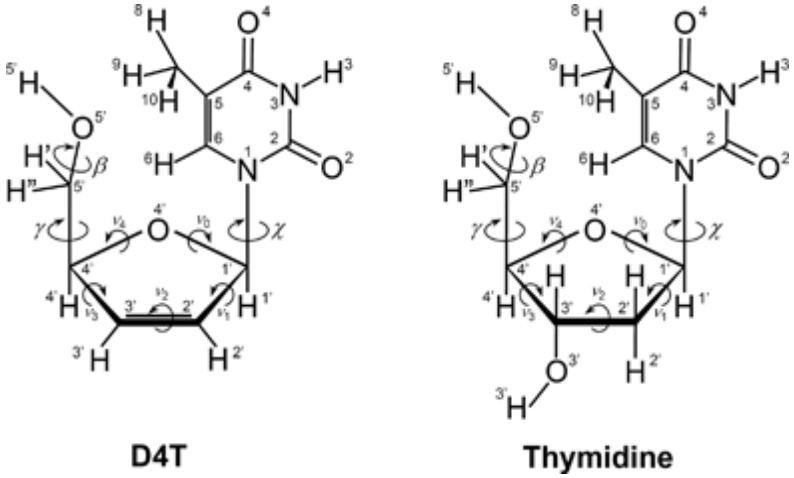
[71] G. Ferenczy, L. Harsányi, B. Rozsondai, I. Hargittai, J. Molec. Struct. 140 (1986) 71-77.

[72] T. III Rush, W.L. Peticolas, J. Phys. Chem. 99 (40) (1995) 14647-14658.

[73] A.Y. Ivanov, S.A. Krasnokutski, G.G. Sheina, Y.P. Blagoi, Spectrochim. Acta A 59 (2003) 1959.

[74] S. Gunasekaran, S. Kumarsen, R.A. Balaji, G. Anand, S. Seshadri, Pramana J. Phys. 1291 (2008) 71.

[75] S. Gunasekaran, R.T. Kumar, S. Ponnusamy, Ind. J. Pure Appl. Phys. 884 (2007) 45.



d4T

Scheme 1. Definition of the endocyclic and exocyclic torsional angles for d4T and deoxythymidine molecules according to the Saenger's notation [59].

LEGENDS OF FIGURES

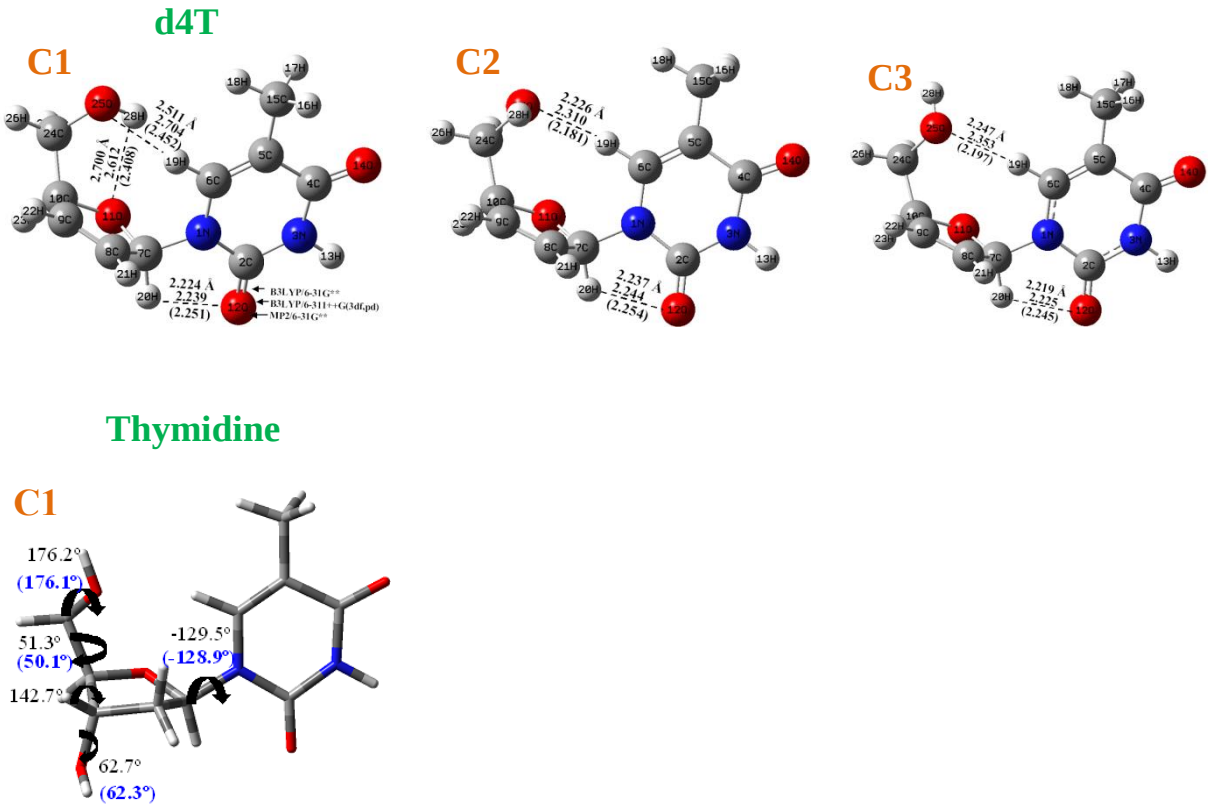
Fig. 1. Geometry of the three most optimum conformers of d4T and the values of the strongest intramolecular H-bonds/distances determined at B3LY/6-31G(d,p), B3LYP/6-311++G(3df,pd) and MP2/6-31G(d,p) (values in parentheses) levels. The molecular structure of the best conformer of the thymidine molecule is also included.

Fig. 2. Two views of the optimized dimer structures of d4T at the B3LYP/6-31G(d,p) and B3LYP/6-311++G(3df,pd) (in parentheses) levels. The dimer G appears with two conformers: C1 and C1-A2.

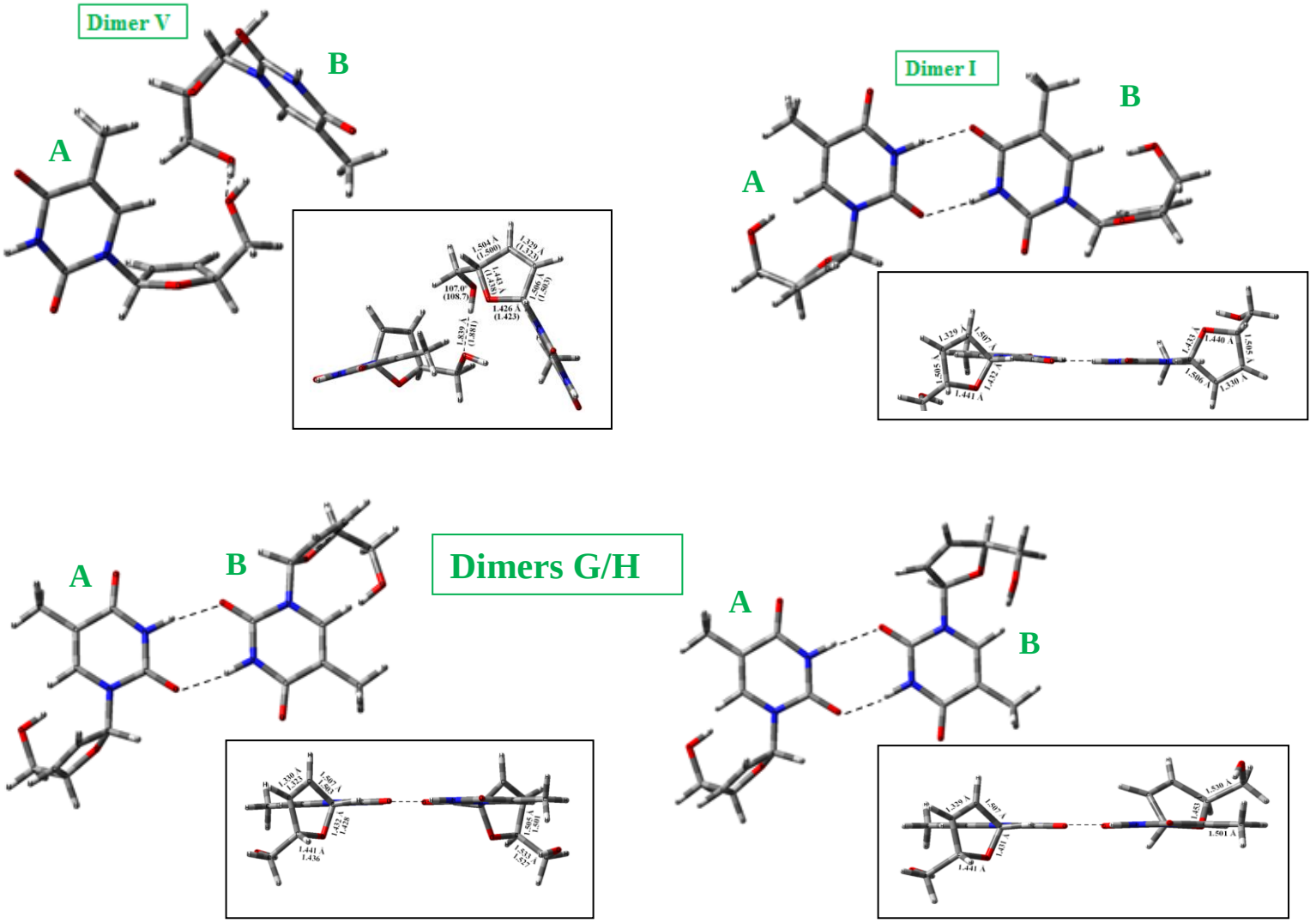
Fig. 3. (a) Calculated IR spectra of the three most stable conformers C1, C2 and C3. (b) Simulated IR spectrum in gas phase of d4T through the population of the three best conformers.

Fig. 4. (a) Theoretical scaled IR spectrum in the isolated state of C1 and C3 conformers of d4T using the LSE procedure from uracil values at the B3LYP/6-31G(d,p) level, (b)-(d) theoretical scaled IR spectra of the three dimers of, and (e) experimental IR spectrum in the solid state in the 3700-2800 cm^{-1} range.

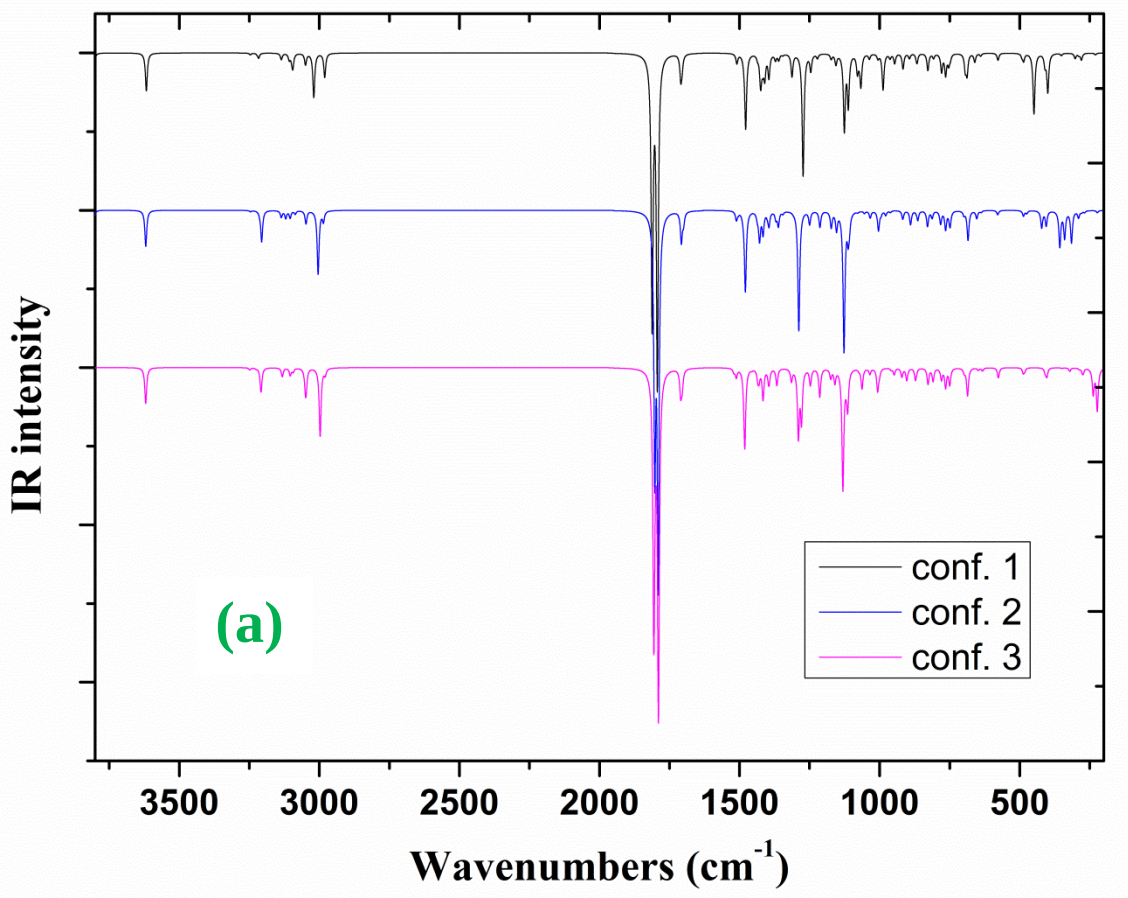
Fig. 5. (a) Theoretical scaled IR spectrum in the isolated state of C1 and C3 conformers of d4T using the LSE procedure from uracil values at the B3LYP/6-31G(d,p) level, (b)-(d) theoretical scaled IR spectra of the three dimers of, and (e) experimental IR spectrum in the solid state in the 1800-350 cm^{-1} range.

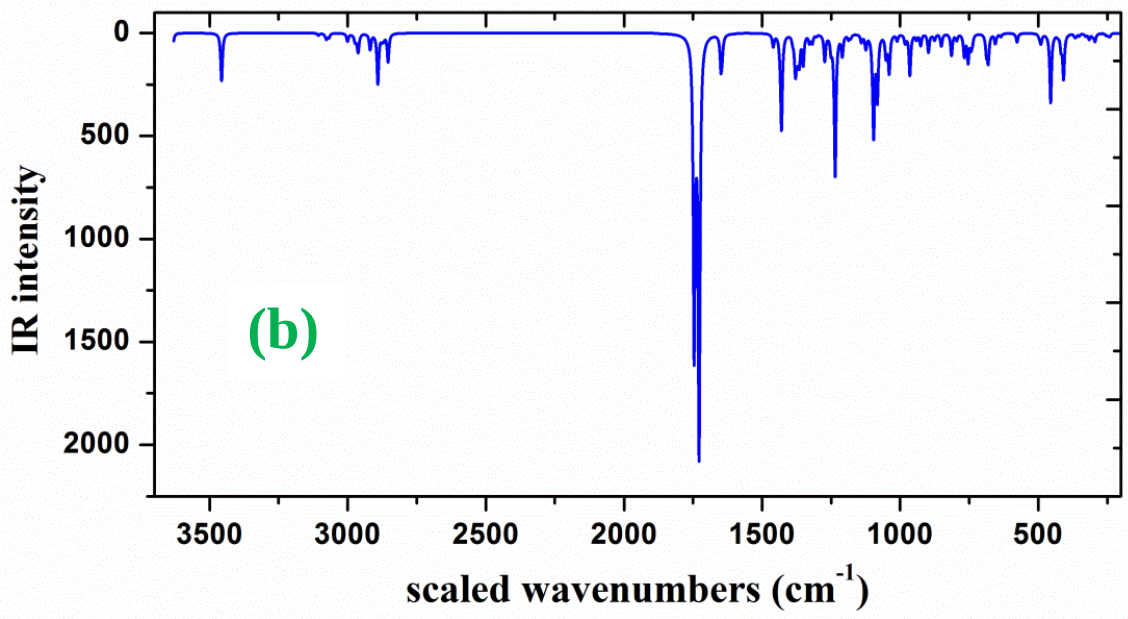

d4T
C1
C2
C3
2.511 Å
2.704
(2.452)
2.700 Å
2.612
(2.408)
2.224 Å
2.239
(2.251)
B3LYP/6-31G**
B3LYP/6-311++G(3df,pd)
MP2/6-31G**
2.226 Å
2.310
(2.181)
2.237 Å
2.244
(2.254)
2.247 Å
2.353
(2.197)
2.219 Å
2.225
(2.245)
Thymidine
C1
176.2°
(176.1°)
51.3°
(50.1°)
142.7°
62.7°
(62.3°)
-129.5°
(-128.9°)

(Fig. 1)

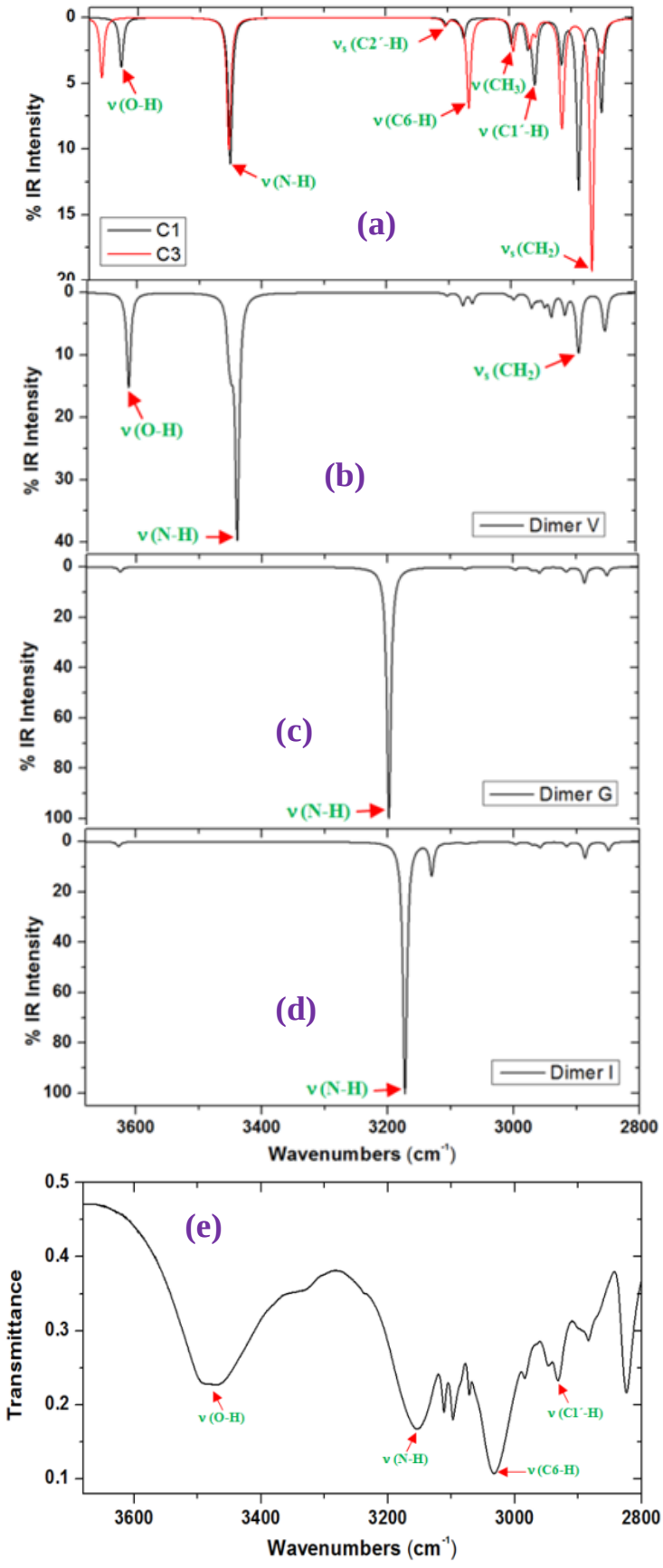

Dimer V
A
B
Dimer I
A
B
Dimers G/H
A
B
A
B
with conformer C1
with conformers C1-A2

(Fig. 2)

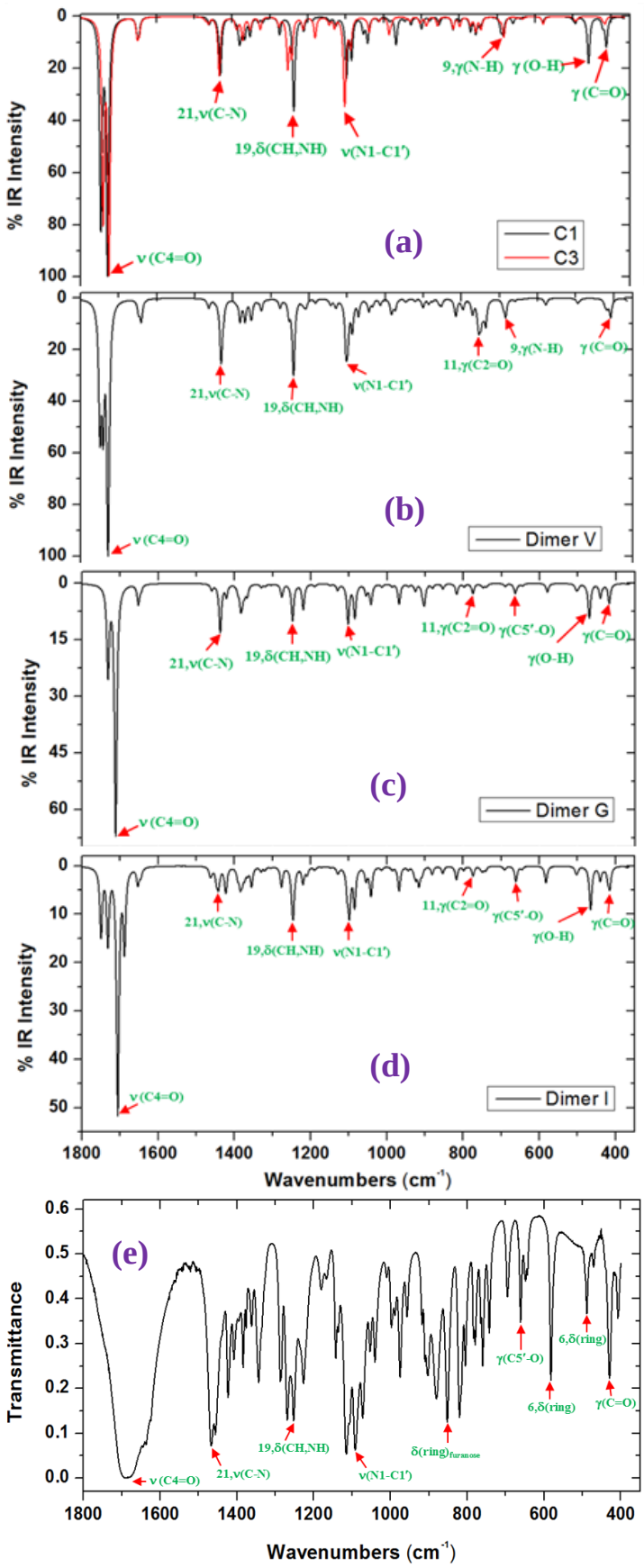

IR intensity
(a)
conf. 1
conf. 2
conf. 3
3500
3000
2500
2000
1500
1000
500
Wavenumbers (cm-1)


IR intensity
0
500
1000
1500
2000
(b)
3500
3000
2500
2000
1500
1000
500
scaled wavenumbers (cm-1)

(Fig. 3)


% IR Intensity
ν (O-H)
ν (N-H)
ν s (C2´-H)
ν (CH3)
ν (C6-H)
ν (C1´-H)
ν s (CH2)
C1
C3
(a)
(b)
Dimer V
(c)
Dimer G
(d)
Dimer I
Wavenumbers (cm-1)
Transmittance
(e)

(Fig. 4)


% IR Intensity
(a)
C1
C3
21,ν(C-N)
19,δ(CH,NH)
ν(N1-C1')
9,γ(N-H)
γ(O-H)
γ(C=O)
ν(C4=O)
(b)
Dimer V
11,γ(C2=O)
(c)
Dimer G
γ(C5'-O)
(d)
Dimer I
Wavenumbers (cm-1)
Transmittance
(e)
6,δ(ring)
δ(ring)furanose

(Fig. 5)

Table 1. Comparison of the most characteristic parameters in the three most stable conformers in the *anti* form of dT and d4T, calculated at the MP2/6-31G(d,p) level. The exocyclic torsional angles and pseudorotational angle P are in degrees, and the energy increments are in kJ·mol^{-1}.

	Conf.	χ	β	γ	P[a]	S[b]	ν_{max}[c]	Δ*E*
d4T	C1	-103.6	63.8	60.6	77.2	$^{0}T_{4}$	8.8	0
	C2	-134.4	-63.8	49.3	95.0	$^{0}T_{1}$	7.7	6.431
	C3	-122.3	165.0	43.7	83.3	$^{0}T_{4}$	8.1	6.719
dT	C1	-128.9	-176.1	50.1	163.3	^{2}E	37.0	0
	C2	-125.1	174.9	49.3	165.8	^{2}E	36.9	2.686
	C3	-134.1	67.7	64.5	24.7	$^{4}_{3}T$	35.4	4.063

[a] Definition: $tgP = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\ \nu_2(\sin(36) + \sin(72))}$ [b]Notation used from ref. [59]. [c] $\nu_{max} = \nu_2 / \cos P$

Table 2. Calculated and experimental values of several selected characteristic parameters of the three dimer forms of the d4T molecule. The exocyclic torsional angles and the pseudorotational angle P are in degrees. The calculated values correspond to the B3LYP, X3LYP and M062X methods, respectively.

dimer form	**χ**	**β**	**γ**	**ν[e]**	**P**	**S**	**ν_{max}**
dimer V:							
Molecule A: by B3LYP	-122.9	167.2	45.5	-130.1	85.1	^{o}E	7.7
X3LYP	-122.8	167.9	45.8	-130.0	85.1		7.8
M062X	-148.2	162.2	46.5	-127.2	88.5		5.7
Molecule B: by B3LYP	-114.1	88.1	58.0	-129.1	80.2	$^{0}_{4}T$	7.6
X3LYP	-113.8	88.3	57.7	-128.9	80.0		7.6
M062X	-106.5	89.8	56.2	-123.4	63.5		2.5
dimer I:							
Molecule A: by B3LYP	-108.7	64.9	62.0	-129.7	80.9	$^{0}_{4}T$	8.0
X3LYP	-108.6	64.8	62.2	-129.6	80.7		7.9
M062X	-105.2	63.2	62.5	-128.9	79.7		8.4
Molecule B: by B3LYP	-107.7	64.4	62.8	-129.5	80.7		7.5
X3LYP	-107.8	64.3	63.0	-129.3	80.4		7.4
M062X	-104.6	63.3	63.2	-127.9	77.0		7.2
X-ray: orthorhombic[a]: Molecule. A	-102.1		50.5	-129.0			
Molecule B	-117.2		62.0	-127.6			
Rhombic[b]: Molecule A	-102.0		52.0				
Molecule B	-159.9						
dimer G/H:							
monomer C1: Molecule A	-107.5	64.3	62.3	-130.0	81.8	^{0}E	8.1
Molecule B	-107.5	64.3	62.3	-130.0	81.9		8.1
monomer A2: Molecule A: by B3LYP	-174.8	58.3	44.9	-127.0	90.4		5.7
X3LYP	-175.3	58.4	45.0	-126.4	90.4		4.3
M062X	-176.5	44.5	41.2	-123.7	96.5		3.4
C1 Molecule B: by B3LYP	-108.6	64.6	62.5	-130.0	81.0		8.2
X3LYP	-108.6	64.6	62.7	-129.8	80.8	$^{0}_{4}T$	8.1
M062X	-110.2	64.2	64.0	-128.8kj`ñ	78.1		8.6
X-ray: triclinic[c]: Molecule A	-172.6		54.1	-125.6	90.4	^{0}E	4.8
Molecule B	-85.1		55.5	-128.8	103.6	$^{0}_{1}T$	6.2
monoclinic[d]: Molecule A	-174.1		53.8	-123.1			
Molecule B	-118.0		60.6	-130.5			

[a] Ref. [62] [b]Ref. [64]. [c]**Ref.** [62,66]. [d]Refs. [62,65]. [e]ν(**N1-C1**′-O4′-C4′).

Table 3. Interaction energies $E^{int}(AB)$ (kJ mol^{-1}), where A is one molecule of the dimer and B is the another molecule, deformation energies E^{def}, and the total CP corrected interaction energy, ΔE^{CP}_{AB}. The calculations are at the M062X/6-31G(d,p) level.

Base pair	$E^{def}{}_A(AB)$	$E^{def}{}_B(AB)$	$E^{def}(AB)$	$E^{int}(AB)$	ΔE^{CP}_{AB}
Dimer V	4.744	1.038	5.782	-87.569	-81.787
Dimer I	6.014	5.924	11.938	-70.901	-58.963
Dimer G/H: monomer C1	**5.540**	5.531	11.071	-68.168	-57.097
Monomer A2	5.678	5.676	11.354	-72.441	-61.087

Table 4. Comparison of the calculated harmonic wavenumbers (ν^{cal}, cm^{-1}), scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %), force constants (f, mDyne/Å), and characterization obtained in the isolated state, with the conformers C1 and C3 of the uracil ring of d4T.

Conformer C1					Conformer C3	
ν^{cal}	ν^{scal},[a]	A[b]	f	Characterization[c]	ν[a]	A[b]
3617.0	3452	12	8.30	(100%) 29,ν(N-H)	3454	11
3216.7	3073	2	6.66	(99%) 27,ν(C6-H)	3065	7
1811.8	1746	51	20.42	(74%) 26,ν(C2=O) + (13%) δ(N-H) + (11%) ν(C4=O)	1740	79
1793.2	1729	100	17.41	(68%) 25,ν(C4=O) + (18%) δ(N-H)	1725	100
1709.5	1650	8	10.84	(45%) 24,ν(C5=C6) + (20%) δ(C6-H) + (15%) ν(C2′=C3′) + (12%) ν(C=O)	1650	7
1477.9	1431	23	3.43	(40%) 21,ν(C-N) + (35%) δ(C1′-H) + (20%) $\delta s(CH_3)$	1434	30
1409.7	1366	7	2.07	(33%) 22,δ(N-H) + (26%) δ(C6-H) + (20%) δ(C1′-H)	1354	2
1395.0	1352	5	2.30	(40%) 20,δ(N-H) + (37%) δ(C6-H) + (12%) δ(C1′-H)	1351	3
1272.9	1237	37	1.81	(55%) 19,δ(C-H, N-H) + (35%) δ(C1′-H)	1243	6
1245.4	1211	5	4.90	(75%) 18,ν(C-N) + (15%) $\omega(CH_3)$	1213	5
1172.6	1142	2	3.47	(65%) 17,ν(N3-C4) + (16%) δ(C1′-H) + (10%) $\Gamma(CH_3)$	1144	3
1125.6	1098	19	2.12	(25%) ν(N1-C1′) + (17%) $\delta(CH)_{furanose}$+ (17%) ν(C5′-O) + (13%) ν(C4′-O) +(13%) δ(ring,CH) +(11%) $\delta_s(CH_2)$	1103	35
945.9	928	3	0.69	(88%) 15,γ(C6-H) + (10%) $\gamma_{as}(CH_3)$	930	2
916.1	900	2	1.44	(50%) 14,ν(ring) + (20%) $\gamma(ring)_{furanose}$ + (18%) $\Gamma(CH_3)$ + (10%) $\Gamma(CH_2)$	904	3
779.1	771	2	2.00	(70%) 16,δ(CN,ring) + (16%) $\gamma(ring)_{furanose}$ + (12%) $\gamma(CH_3)$	771	3
764.7	757	6	2.19	(78%) 10,γ(C4=O, N-H) + (17%) $\gamma_{as}(CH_3)$	757	6
753.8	747	3	2.33	(70%) 11,γ(C2=O, N-H) + (15%) $\gamma_s(C\text{-}H)_{furanose}$ + (10%) $\gamma_{as}(CH_3)$	744	0
748.1	741	1	1.88	(30%) 11,γ(C2=O) + (26%) 12,δ(ring) + (19%) γ_s(C-H) in C2′H=C3′H + (11%) $\gamma_s(CH_3)$	743	4
688.1	685	6	0.35	(90%) 9,γ(N-H)	683	8
639.6	639	1	1.22	(40%) 7,δ(ring) + (23%) $\delta(CH_2)$ + (17%) $\delta(ring)_{furanose}$+ (17%) $\Gamma(CH_3)$	633	1
577.7	580	2	1.14	(38%) 6,δ(ring) + (27%) $\gamma_s(ring)_{furanose}$ + (22%) $\delta(CH_2OH)$ + (13%) $\Gamma(CH_3)$	579	3
492.2	500	1	0.54	(28%) γ(N1-C1′) + (22%) δ,$\gamma_s(ring)_{furanose}$ + (20%) γ(ring) + (15%) δ(O-H) + (10%) $\Gamma(CH_2)$	495	2
486.2	494	2	0.88	(60%) 6,δ(ring) + (19%) $\delta_s(ring)_{furanose}$ + (15%) $\gamma_s(CH_3)$	490	1
409.9	422	3	0.52	(50%) 5,δ,γ(ring,C=O) + (18%) $\gamma_s(CH_3$+ (15%) $\gamma(C=C)_{furanose}$ + (15%) δ(O-H))	416	3
399.9	412	12	0.33	(36%) δ,γ(ring, C=O) + (34%) δ(O-H) + (20%) γ(C=C,ring)$_{furanose}$	364	0
351.6	367	1	0.28	(40%) 4,γ(ring) mainly in C5=C6 + (33%) $\tau(ring)_{furanose}$ + (16%) $\delta(CH_2OH)$ + (11%) $\Gamma(CH_3)$	338	1
289.0	308	1	0.20	(43%) 3,δ(ring,C=O) + (36%) $\delta(CH_3)$ + (16%) $\delta_s(ring)_{furanose}$	296	1
272.7	292	2	0.18	(39%) τ(ring) mainly in C5=C6 + (20%) $\Gamma(CH_3)$ + (34%) $\tau(ring)_{furanose}$	259	7
166.3	192	0	0.14	(62%) 1,γ(puckering ring on N3) + (22%) $\tau(ring)_{furanose}$ + (12%) $\delta(CH_2OH)$	193	1
105.5	134	0	0.03	(47%) γ(C=O) + (32%) $\gamma_s(CH_3)$ + (11%) $\delta_s(ring)_{furanose}$ + (10%) γ(C5′H_2OH)	135	0
41.9	74	0	0.01	(37%) 1,γ(puckering ring on N3) + (33%) $\delta(CH_2OH)$ + (16%) $\gamma(CH_3)$ + (14%) $\delta(ring)_{furanose}$	68	0

[a]With the LSE: ν^{scal} =34.6+0.9447·ν^{cal} [48] [b]Normalized to the highest value. [c]The normal modes of the uracil ring were labeled according to ref. [48].

Table 5. Comparison of the calculated harmonic wavenumbers (ν^{cal}, cm^{-1}), scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %), force constants (f, mDyne/Å), and characterization obtained in the isolated state, with the conformers C1 and C3 of the furanose ring and for the C5′H_2OH group of d4T.

Conformer C1					Conformer C3	
ν^{cal}	$\nu^{scal,\ a}$	A^b	f	Characterization	$\nu^{scal,\ a}$	A^b
3246.7	3102	1	6.87	(99%) ν_s(C-H) in C2′H=C3′H	3104	1
3223.1	3079	0	6.66	(98%) ν_{as}(C-H) in C2′H=C3′H	3082	0
3098.1	2961	5	6.14	(98%) ν(C1′-H)	2957	1
2980.8	2851	4	5.66	(84%) ν(C4′-H) + (16%) ν(C5′-H)	2849	2
1705.0	1645	6	11.30	(47%) ν(C2′=C3′) + (14%) ν(C5=C6) + (12%) δ(C2′-H) + (12%) δ(C3′-H)	1645	10
1425.6	1381	12	1.87	(27%) δ(C1′-H) + (26%) ω(CH_2) + (18%) 22,δ(C6-H) + (18%) ω(CH_3)	1390	7
1371.5	1330	2	1.95	(46%) δ_s(C-H) in C2′H=C3′H + (25%) δ(C1′-H) + (17%) δ(C5′H_2)	1326	5
1359.8	1319	2	1.77	(38%) δ(C1′-H) + (24%) δ(C4′-H) +(17%) δ(OH) + (15%) δ_s(CH_2)	1325	5
1344.9	1305	0	1.47	(55%) δ(C4′-H) + (27%) δ(CH_2) + (10%) δ(C2′-H)	1276	4
1312.6	1275	7	1.36	(79%) δ_s(C-H)$_{furanose}$ + (10%) τ(CH_2)	1253	24
1155.6	1126	3	1.58	(28%) δ_{as}(C-H)$_{furanose}$ + (20%) δ(ring) + (19%) ν(C5-CH_3) + (16%) ν(C5′-O) + (11%) δ(CH_2)	1130	6
1111.5	1085	18	1.25	(45%) δ_{as}(C-H)$_{furanose}$ in C2′H=C3′H + (23%) ν(C5′-O) + (16%) δ(CH_2) + (12%) ν(C4′-O)	1092	12
1078.9	1054	5	1.22	(25%) δ(C-H)$_{furanose}$ + (22%) ν(C4′-O) + (17%) δ(O-H) + (15%) Γ(CH_3) + (12%) Γ(CH_2)	1087	11
1066.5	1042	6	2.29	(36%) ν_{as}(ring,CH)$_{furanose}$ + (34%) ν(C5′-O) + (18%) Γ(CH_3) + (16%) δ(ring) + (14%) Γ(CH_2)	1038	6
1005.6	985	2	1.45	(70%) ν_s(ring,CH)$_{furanose}$ + (15%) Γ(CH_2)	986	7
987.3	967	11	1.37	(58%) ν_{as}(ring,CH)$_{furanose}$ + (28%) Γ(CH_2) + (10%) δ(OH)	980	2
962.4	944	1	0.71	(85%) γ_{as}(C-H) in C2′H=C3′H	940	1
866.9	854	1	1.24	(55%) δ,γ(ring)$_{furanose}$ + (25%) Γ(CH_2) + (18%) δ(C5′-O)	858	4
828.0	817	2	1.77	(37%) δ(ring)$_{furanose}$ + (27%) γ(C2′-H, C3′-H) + (16%) γ(OH) + (11%) Γ(CH_2)	816	5
807.8	798	1	1.56	(48%) δ(ring)$_{furanose}$ + (30%) γ_s(C-H) in C2′H=C3′H + (12%) δ(CH_2)	799	4
694.5	691	5	0.60	(30%) γ_s(C-H) in C2′H=C3′H + (23%) γ(N3-H) + (18%) γ(ring)$_{furanose}$ + (15%) γ_s(CH_2)	694	2
660.7	659	3	1.01	(30%) γ(C5′-O) + (23%) γ,δ(ring)$_{furanose}$ + (23%) 7,δ(ring) + (14%) γ_s(CH_2)	647	1
302.2	320	2	0.15	(44%) δ,γ(ring)$_{furanose}$ + (30%) δ_s(CH_2OH) + (24%) 5,δ(ring)	306	0
279.9	299	2	0.19	(35%) γ(ring)$_{furanose}$ + (29%) δ_s(CH_2OH) + (24%) 3,δ(ring,C=O) + (12%) Γ(CH_3)	293	2
229.8	252	1	0.12	(42%) γ(ring)$_{furanose}$ + (33%) γ(C5′H_2OH) + (14%) γ(CH_3) + (11%) 3.δ(ring)	245	12
138.3	165	1	0.05	(33%) γ(C5′H_2OH) + (25%) 1,γ(ring, C=O) + (27%) γ(ring)$_{furanose}$ + (15%) γ(CH_3)	168	0
126.1	154	1	0.03	(35%) τ(ring)$_{furanose}$ + (30%) γ(COH) in CH_2OH+ (20%) δ_s(CH_3) + (15%) Γ(CH_2)	159	0
69.2	100	0	0.01	(35%) γ(C5′H_2OH) + (20%) 1,γ(puckering N3 ring) + (16%) γ(ring)$_{furanose}$ + (15%) γ(CH_3)	109	0
51.4	83	0	0.01	(35%) γ(C5′H_2OH) + (29%) 1,γ(puckering N3 ring) + (15%) γ(CH_3) + (12%) γ(ring)$_{furanose}$	89	0

[a]With the LSE: ν^{scal} =34.6+0.9447·ν^{cal} [48] [b]Normalized to the highest value.

Table 6. Comparison of the calculated harmonic wavenumbers (ν^{cal}, cm^{-1}), scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %), force constants (f, mDyne/Å), reduced mass (μ) and characterization obtained in the isolated state, with the conformers C1 and C3 for the CH_3, CH_2 and OH groups of d4T.

Group	Mode	Conformer C1					Conformer C3	
		ν^{cal}	$\nu^{scal,\,a}$	A^b	f	μ	$\nu^{scal,\,a}$	A^b
OH	ν	3803.7	3628	2	9.09	1.1	3659	4
	δ	1416.5[c]	1373	9	1.55	1.3	1372	12
	γ	449.5	459	19	0.17	1.5	421	1
CH_2	ν_{as}	3094.8	2958	2	6.20	1.1	2916	5
	ν_s	3019.9	2887	7	5.74	1.1	2866	20
	β	1509.0	1460	0	1.47	1.1	1462	3
	ω	1423.5	1379	10	1.82	1.5	1386	6
	τ	1221.8[d]	1189	1	1.02	1.2	1181	8
	Γ	893.3	878	2	1.15	2.5	887	4
CH_3	ν_{as}	3135.8	2997	2	6.39	1.1	2994	3
	ν_{as}	3107.6	2970	2	6.26	1.1	2967	2
	ν_s	3049.5	2915	2	5.67	1.0	2913	4
	δ_s	1509.4	1461	2	1.50	1.1	1462	3
	δ_{as}	1484.6	1437	1	1.36	1.0	1437	1
	ω	1435.3	1391	1	1.67	1.4	1410	0
	δ_{as}	1072.5	1048	1	1.00	1.5	1048	1
	γ_{as}	1036.4	1014	2	1.10	1.7	1012	2
	Γ	135.9	163	0	0.01	1.3	168	0

Notation: ν (stretching), δ (in-plane bending), β (scissoring), γ (out-of-plane bending), Γ (rocking), τ (twisting), ω (wagging). [a]With the LSE: $\nu^{scal} = 34.6 + 0.9447 \cdot \nu^{cal}$ [48]. [b]Normalized to the highest value. [c]Strongly coupled with $\tau(CH_2)$. [d]Strongly coupled with δ(O-H).

Table 7. Comparison of the scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %) and characterization obtained in the isolated state with monomer C1and the dimers V, G and I of the thymine ring of the d4T molecule. Two wavenumbers appear for each vibration of the dimers, and that with the highest IR intensity is shown in bold letters.

Monomer C1	Dimer V		Dimer G		Dimer I		Characterization	Exp IR
$\nu^{scal,\ a}$	$\nu^{scal,a}$	A^b	$\nu^{scal,\ a}$	A^b	$\nu^{scal,\ a}$	A^b		
3452	3449, **3440**	13	**3198**, 3168	100	**3172**, 3130	100	29,ν(N-H)	3153.8 vs
3073	**3077**, 3062	4	**3076**, 3076	0	3075, **3074**	0	27,ν(C6-H)	3032.5 s
1746	**1749**, 1741	58	1731, **1731**	22	1749, **1732**	15	26,ν(C2=O) + δ(N-H) + ν(C4=O)	
1729	1728, **1727**	100	**1711**, 1701	66	**1706**, 1688	51	25,ν(C4=O) + δ(N-H)	1691.7 vs
1650	**1648**, 1646	9	1653, **1652**	1	**1653**, 1648	3	24,ν(C5=C6) + δ(C6-H) + ν(C2′=C3′)	1652.9 vw
1431	1432, **1430**	23	1437, **1437**	13	**1438**, 1437	3	21,ν(C-N) + δ(C1′-H) + $\delta s(CH_3)$	1423.4 m
1366	**1369**, 1368	8	1374, **1374**	2	**1373**, 1373	1	22,δ(N-H) + δ(C6-H) + δ(C1′-H)	1362.3 vw
1352	1353, **1352**	7	1367, **1367**	3	1366, **1355**	4	20,δ(N-H) + δ(C6-H) + δ(C1′-H)	1343.8 m
1237	**1242**, 1240	31	1250, **1248**	10	1248, **1245**	8	19,δ(C-H, N-H) + δ(C1′-H)	1252.3 m
1211	1219, **1213**	4	**1220**, 1220	4	**1220**, 1209	3	18,ν(C-N) + $\omega(CH_3)$	1226.9 w
1142	1144, **1143**	2	**1149**, 1149	0	1172, **1148**	0	17,ν(N3-C4) + δ(C1′-H) + $\Gamma(CH_3)$	1142.9 w
1098	**1103**, 1098	23	1104, **1102**	10	1104, **1098**	9	ν(N1-C1′)+$\delta(CH)_{furanose}$+ ν(C5′-O)	1115.1 m, 1107.2 m
928	**946**, 922	2	**926**, 926	2	928, **924**	2	15,γ(C6-H) + $\gamma_{as}(CH_3)$	918.4 vw
900	902, **900**	5	908, **907**	2	**916**, 908	4	14,ν(ring) + $\gamma(ring)_{furanose}$ + $\Gamma(CH_3)$	910.1 vw, 901.9 vw
771	**771**, 769	8	**798**, 798	1	**799**, 798	1	16,δ(CN,ring) + $\gamma(ring)_{furanose}$ + $\gamma(CH_3)$	783.6 w, 778.7 w
757	757, **754**	10	**775**, 774	3	775, **773**	1	10,γ(C4=O, N-H) + $\gamma_{as}(CH_3)$	765.7 w
747	**749**, 744	14	**753**, 752	0	**754**, 753	0	11,γ(C2=O, N-H) + $\gamma_s(C\text{-}H)_{furanose}$ + $\gamma_{as}(CH_3)$	760.2 w
741	**743**, 741	4	749, **748**	1	751, **748**	1	11,γ(C2=O) + 12,δ(ring)	743.6 m
685	685, **684**	8	691, **691**	1	692, **690**	0	9,γ(N-H)	695.5 m
639	**644**, 637	1	647, **644**	0	**646**, 644	0	7,δ(ring) + $\delta(CH_2)$ + $\delta(ring)_{furanose}$+ $\Gamma(CH_3)$	648.7 m, 643.2 vw
580	629, **580**	2	583, **580**	2	**583**, 580	3	6,δ(ring) + $\gamma_s(ring)_{furanose}$ + $\delta(CH_2OH)$	582.3 s
500	**578**, 496	3	505, **504**	1	505, **504**	1	γ(N1-C1′) + δ,$\gamma_s(ring)_{furanose}$ + γ(ring)	
494	495, **494**	2	500, **500**	1	**501**, 500	1	6,δ(ring) + $\delta_s(ring)_{furanose}$ + $\gamma_s(CH_3)$	488.7 vw, 471.3 m
422	422, **418**	4	**441**, 432	4	**440**, 432	3	5,δ,γ(ring,C=O) + γ_s(**CH_3**+ $\gamma(C=C)_{furanose}$	429.6 s
412	416, 415	4	418, **418**	3	420, **415**	4	δ,γ(ring, C=O) + δ(O-H) + $\gamma(C=C,ring)_{furanose}$	407.9 w
367	**365**, 363	1	**369**, 369	0	**370**, 367	0	4,γ(ring) mainly in C5=C6 + $\tau(ring)_{furanose}$	
308	**311**, 308	1	**316**, 314	0	318, **314**	0	3,δ(ring,C=O) + $\delta(CH_3)$ + $\delta_s(ring)_{furanose}$	
292	**294**, 292	0	**296**, 296	0	296, **295**	0	τ(ring) mainly in C5=C6 + $\Gamma(CH_3)$	
192	252, **195**	0	203, **200**	0	**207**, 200	0	1,γ(puckering ring on N3) + $\tau(ring)_{furanose}$	
134	**88**, 86	3	92, **86**	0	**91**, 86	0	γ(C=O) + $\gamma_s(CH_3)$ + $\delta_s(ring)_{furanose}$	

[a]With the LSE: ν^{scal} =34.6+0.9447·ν^{cal} [48] [b]Normalized to the highest value.

Table 8. Comparison of the scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %) and characterization obtained in the isolated state with monomer C1and the dimers V, G and I of the furanose ring of the d4T molecule. Two wavenumbers appear for each vibration of the dimers, and that with the highest IR intensity is shown in bold letters.

Monomer C1	Dimer V		Dimer G		Dimer I		Characterization	Exp IR
$\nu^{scal,\ a}$	$\nu^{scal,\ a}$	A^b	$\nu^{scal,\ a}$	A^b	$\nu^{scal,\ a}$	A^b		
3102	3104, **3102**	1	**3102**,3101	0	**3102**,3101	0	ν_s(C-H) in C2′H=C3′H	3112.9 w
3079	3084, **3081**	0	3079**,3079**	0	**3080**,3079	0	ν_{as}(C-H) in C2′H=C3′H	3097.7 m, 3071.4 m
2961	2962, **2956**	1	2966,**2966**	1	2967,**2963**	0	ν(C1′-H)	2949.4 m
2851	**2851**, 2848	7	2851, **2850**	2	2849, **2849**	2	ν(C4′-H) + ν(C5′-H)	2823.5 m
1645	1642, **1640**	12	1645, **1645**	1	1645, **1645**	1	ν(C2′=C3′) + ν(C5=C6) + δ(C2′-H) + δ(C3′-H)	1645.7 vw
1381	1394, **1391**	1	1392, **1391**	2	**1392**, 1391	2	δ(C1′-H) + ω(CH_2) + 22,δ(C6-H) + ω(CH_3)	1376.4 m
1330	1336, **1326**	5	1330, **1330**	1	**1330**, 1330	0	δ_s(C-H) in C2′H=C3′H + δ(C1′-H) + δ(C5′H_2)	
1319	**1325**, 1324	5	1320, **1320**	1	1320, **1320**	0	δ(C1′-H) + δ(C4′-H) + δ(OH) + δ_s(CH_2)	
1305	1308, **1276**	5	1305, **1304**	0	1305, **1304**	0	δ(C4′-H) + δ(CH_2) + δ(C2′-H)	1287.4 w
1275	1275, **1253**	4	1276, **1276**	4	**1276**, 1276	3	δ_s(C-H)$_{furanose}$ + τ(CH_2)	1268.7 m
1126	**1131**, 1128	4	1130, **1130**	1	**1130**, 1128	1	δ_{as}(C-H)$_{furanose}$ + δ(ring) + ν(C5-CH_3) + ν(C5′-O)	1136.1 vw
1085	1088, **1087**	13	1085, **1085**	8	1085, **1084**	4	δ_{as}(C-H)$_{furanose}$ in C2′H=C3′H + ν(C5′-O) + δ(CH_2)	1091.9 m
1054	**1086**, 1071	16	1055, **1055**	2	**1055**, 1055	2	δ(C-H)$_{furanose}$ + ν(C4′-O) + δ(O-H) + Γ(CH_3)	1072.5 w
1042	**1043**, 1033	9	1043, **1042**	4	1042, **1041**	5	ν_{as}(ring,CH)$_{furanose}$ + ν(C5′-O) + Γ(CH_3) + δ(ring)	1040.8 vw
985	985, **983**	7	**987**, 987	1	987, **985**	0	ν_s(ring,CH)$_{furanose}$ + Γ(CH_2)	988.6 vw
967	978, **973**	7	969, **969**	4	**968**, 967	3	ν_{as}(ring,CH)$_{furanose}$ + Γ(CH_2) + δ(OH)	974.8 w
944	944, 943	2	944, **944**	1	**944**, 944	0	γ_{as}(C-H) in C2′H=C3′H	956.9 m
854	858, **852**	5	**881**, 874	0	**881**, 881	1	δ,γ(ring)$_{furanose}$ + Γ(CH_2) + δ(C5′-O)	851.9 m
817	818, **814**	4	854, **854**	1	854, 853	1	δ(ring)$_{furanose}$ + γ(C2′-H, C3′-H) + γ(OH) + Γ(CH_2)	820.3 m
798	800, **796**	4	**818**, 818	2	**818**, 818	2	δ(ring)$_{furanose}$ + γ_s(C-H) in C2′H=C3′H + δ(CH_2)	805.6 m
691	692, 691	4	**740**, 739	1	**741**, 740	0	γ_s(C-H) in C2′H=C3′H + γ(N3-H) + γ(ring)$_{furanose}$	
659	660, 659	8	**665**, 658	3	**662**, 659	3	γ(C5′-O) + γ,δ(ring)$_{furanose}$ + 7,δ(ring) + γ_s(CH_2)	661.1 m
320	337, **329**	2	328, **323**	0	**328**, 323	0	δ,γ(ring)$_{furanose}$ + δ_s(CH_2OH) + 5,δ(ring)	
299	296, **295**	2	304, **302**	1	303, **301**	1	γ(ring)$_{furanose}$ + δ_s(CH_2OH) + 3,δ(ring,C=O) + Γ(CH_3)	

[a]With the LSE: $\nu^{scal} = 34.6 + 0.9447 \cdot \nu^{cal}$ [48] [b]Normalized to the highest value.

Table 9. Comparison of the scaled wavenumbers (ν, cm^{-1}), relative IR intensities (A, %), and characterization obtained in the isolated state and the dimers V, G and I for the CH_3, CH_2 and OH of d4T, and experimental IR wavenumbers. Two wavenumbers appear for each vibration of the dimers, and that with the highest IR intensity is shown in bold letters.

Group	Mode	Monomer C1	Dimer V		Dimer G		Dimer I		Exp IR
		$\nu^{scal,\ a}$	$\nu^{scal,\ a}$	A^b	$\nu^{scal,\ a}$	A^b	$\nu^{scal,\ a}$	A^b	
OH	ν	3628	**3614**, 3453	12	3626, **3626**	1	**3628**, 3627	1	3481.6 s
	δ	1373	**1424**, 1183	8	1380, **1380**	3	**1380**, 1379	1	1407.7 vw
	γ	459	**736**, 408	4	**470**, 469	8	467, **465**	6	
CH_2	ν_{as}	2958	2946, **2936**	7	2958, **2958**	1	2958, **2958**	1	2933.1 m
	ν_s	2887	2893, **2890**	11	**2886**, 2886	3	2887, **2886**	3	2883.5 m
	β	1460	**1467**, 1456	1	1461, **1461**	0	1461, **1460**	0	1467.1 s?
	ω	1379	1382, **1380**	8	1386, **1384**	5	1385, **1384**	3	1383.8 w
	τ	1189	1380, **1207**	5	1190, **1190**	1	**1190**, 1189	0	1181.1 vw
	Γ	878	**886**, 876	4	**903**, 881	5	**907**, 881	1	881.2 w
CH_3	ν_{as}	2997	3004, **2996**	2	2996, **2996**	1	2997, 2996	1	2985.5 vw
	ν_{as}	2970	**2968**, 2967	2	**2970**, 2970	1	2972, **2971**	0	
	ν_s	2915	**2914,** 2913	4	2915, **2915**	2	2916, **2915**	1	
	δ_s	1461	1464, **1463**	4	1459, **1459**	1	**1464**, 1459	2	
	δ_{as}	1437	1446, **1438**	1	1484, **1484**	2	**1438**, 1437	0	
	ω	1391	1405, 1404	8	**1421**, 1418	3	1425, **1422**	5	
	δ_{as}	1048	**1050**, 1047	2	**1048**, 1048	2	**1048**, 1047	3	1053.4 vw
	γ_{as}	1014	1013, **1013**	3	**1014**, 1014	1	**1015**, 1013	0	1010.6 vw
	Γ	163	184, **170**	0	168, **166**	0	**170**, 167	0	

[a]With the LSE: $\nu^{scal} = 34.6 + 0.9447 \cdot \nu^{cal}$ [48] [b]Normalized to the highest value.

Table 10. Comparison of the selected scaled wavenumbers (ν^{scal}, cm^{-1}), relative IR intensities (A, %), relative Raman intensities (S, %), characterization obtained in the isolated state with monomer C1 and the dimers V, G and I, and the experimental IR and Raman frequencies obtained in the solid state. Two wavenumbers appear for each vibration of the dimers, and that with the highest IR intensity is shown in bold letters, and that with the highest Raman intensity is shown in italic letters.

Monomer C1	Dimer V			Dimer G			Dimer I			Characterization	Exp IR	Exp Raman
$\nu^{scal,\ a}$	$\nu^{scal,a}$	A^b	S^b	$\nu^{scal,\ a}$	A^b	S^b	$\nu^{scal,\ a}$	A^b	S^b			
3628	*3453*, **3614**	12	45	*3626*, **3626**	1	0	**3628**, *3627*	1	2	ν(O-H)	3481.6 w	-
1729	*1728*, **1727**	100	34	**1711**, *1701*	66	4	**1706**, *1688*	51	4	ν(C4=O)+δ(N-H)	1691.7 vs	1690.2 vs
1237	**1242**, *1240*	31	3	*1250*, **1248**	10	0	*1248*, **1245**	8	12	δ(C-H, N-H)+δ(C1′-H)	1252.3 w	1235.6 m
320	*337*, **329**	2	0	328, ***323***	0	0	**328**, *323*	0	1	$\delta,\gamma(ring)_{furanose}+\delta_s(CH_2OH)$	-	346.2 vw
192	*252*, **195**	0	0	203, ***200***	0	0	**207**, *200*	0	0	1,γ(puckering ring on N3)	-	249.5 vw

[a]With the LSE: $\nu^{scal} = 34.6+0.9447 \cdot \nu^{cal}$ [48] [b]Normalized to the highest value.

Table 11. Rms errors obtained after scaling the calculated wavenumbers of d4T by different DFT methods.

Structure	Method	Scaling procedure			
		OSF	LSE	TLSE	PSE
Monomer C1	B3LYP	24.0	15.9	10.5	13.6
	X3LYP	-	14.2	10.5	13.2
	M062X	-	19.6	11.7	18.9
Dimer V	B3LYP	21.9	12.6	8.2	11.2
	X3LYP	-	12.5	8.1	11.2
	M062X	-	18.1	11.5	18.0
Dimer I	B3LYP	22.6	16.3	14.5	15.2
Dimer G	B3LYP	21.8	16.0	14.1	15.2