M. Alcolea Palafox[1]

Computational chemistry applied to vibrational spectroscopy: A tool for characterization of nucleic acid bases and some of their 5-substituted derivatives

[1] Departamento de Química-Fisica I. Facultad de Ciencias Químicas. Universidad Complutense. Madrid- 28040. Spain., E-mail: alcolea@quim.ucm.es

Abstract:
Computational chemistry can be applied to vibrational spectroscopy in different ways, such as for a better characterization and assignment of all the bands of the experimental spectra, as a tool in the identification of the tautomers present in the gas phase and in the solid state through their spectra and for the simulation of the solid and liquid phase of a compound and the consequent simulation and interpretation of their spectra. In the present study, as an example of the applicability of computational chemistry, the structure and spectra of cytosine and uracil nucleic acid bases and two cytosine derivatives are shown. The FTIR and Raman spectra were analysed with the support of ab initio (Hartree-Fock (HF), MP2) and density functional theory (DFT) (B3LYP, PBE, B-P, etc.) calculations using several basis sets and several scaling equations. The calculations predict an easier tautomerization of cytosine than uracil molecule, but the tautomerization is hindered in the 5-bromocytosine molecule. Thus, in the solid state, this molecule only exists in the amino-oxo tautomeric form.

Keywords: applications of computational chemistry, cytosine molecule, scaling
DOI: 10.1515/psr-2016-0132

1 Introduction

The motivation for the application of computational chemistry to vibrational spectroscopy is to make it a more practical tool [1]. From a practical point of view, the main disadvantage of vibrational spectroscopy is the lack of a direct spectrum–structure relation. This makes it impossible or difficult to determinate the structure of a molecule from its vibrational spectrum. However, vibrational spectroscopy has a number of advantages over other spectroscopic methods, such as Nuclear Magnetic Resonance (NMR). Most of these relate to the inherently greater sensitivity of vibrational spectroscopy, which makes it possible to detect very small amounts. Further advantages are the wider scope of vibrational spectroscopy, e.g. its applicability to solids, liquids and gasses, as well as to adsorbed layers, etc. Instrument costs of IR spectroscopy are also generally lower than for other spectroscopic techniques. It is thus clear that many of the advantages of vibrational spectroscopy could be increased if a method could be found to reliably predict vibrational spectra. Such a method could be used to calculate the expected spectra of proposed structures. Comparison with the observed spectra would confirm the identity of a product, even that of a completely new molecule. Density functional theory (DFT) quantum chemical methods are the most suitable for this purpose, and for many tasks the B3LYP/DFT method is the most commonly used today [2–6].

In general, the computation of the vibrational spectrum of a polyatomic molecule of even modest size is lengthy. The most accurate of the quantum chemical methods is still too expensive and cumbersome to apply in routine research. Thus, one may be forced to work at low level, and consequently, one must expect a high overestimation of the calculated vibrational frequencies. This overestimation can be significantly reduced by the use of transferable empirical parameters for the calculated wavenumbers [1]. The scaling is therefore designed to correct the calculated harmonic wavenumbers to be compared with the anharmonic wavenumbers found by experiment. Thus, the first step after the computation of vibrational wavenumbers is to correct the systematic errors in their values, i.e. the scaling. The present manuscript shows the use of this scaling in the spectra of the cytosine and uracil molecules, which can permit an accurate assignment of wavenumbers.

M. Alcolea Palafox is the corresponding author.

2 Molecules under study

A special characteristic of the molecules under study is their tautomerism. Nucleic acid bases are constituents of DNA and RNA, and they play important roles in the transcription of the genetic code. Although for this process, they naturally occur as one predominant isomer, other minor tautomeric forms also exist. In the famous 1953 publication [7], Watson and Crick stated the importance of tautomeric forms of pyrimidine and purine nucleic acid bases with respect to three-dimensional stacking in DNA. This phenomenon and the hypothesis that rare tautomers may be responsible for DNA mutations have encouraged many chemists to carry out theoretical and experimental studies on the structure of nucleobases in different environments. For this purpose, computational methods have been extensively used as important tools for the interpretation of vibrational spectra [8, 9]. Another interest of these molecules is because the experimental IR wavenumbers in the gas phase are available for the uracil molecule and in an Ar matrix for cytosine molecule. This feature enables us to calculate accurate scaling factors/equations to be used in their derivatives and in related molecules, which facilitates a good match between the scaled and experimental wavenumbers.

Cytosine (in short Cy, Figure 1**b**), also named as 4-amino-2(1H)-pyrimidinone or 4-amino-2-hydroxypyrimidine, is a pyrimidine base and a constituent of nucleotides and, as such, one member of the base pair guanine–cytosine. Uracil (in short U, Figure 1**a**) is also a pyrimidine base and a constituent of nucleotides and, as such, one member of the base pair adenine–uracil. They belong to the group of the most important pyrimidines that play a fundamental role in the structure and function of enzymes and drugs. The importance of Cy and their derivatives is currently indicated by the considerable number of publications devoted to them appearing in the bibliography [10–14].

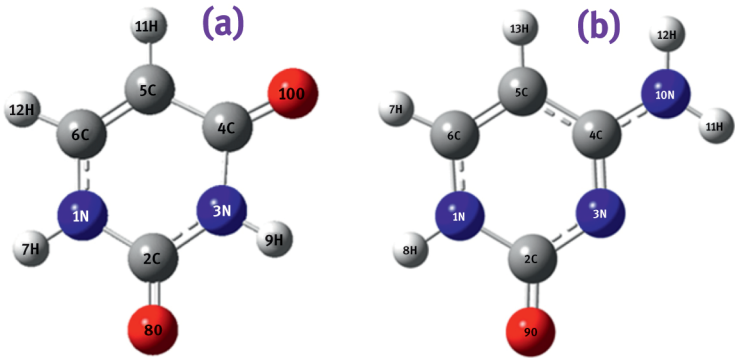

Figure 1: Structures of (a) uracil and (b) cytosine with the labels of their atoms.

Cy contains two labile protons and five conjugated tautomeric sites. Thus, it can exist in various tautomeric forms differing from each other by the position of the proton, which may be bound to either the ring nitrogen atoms or the oxygen atom: amino-oxo (C1), amino-hydroxy (C2a, C2b), imino (C3a, C3b) and (C4) 3H-oxo forms. Of all the possible combinations [15], the six tautomers of Figure 2 are the most important and studied tautomers. This molecule is the only nucleobase where the *enol* tautomer C2b is more stable than the *keto* C1 tautomer in the isolated state. There is a great interest in the tautomerism of Cy because of this property [15, 16].

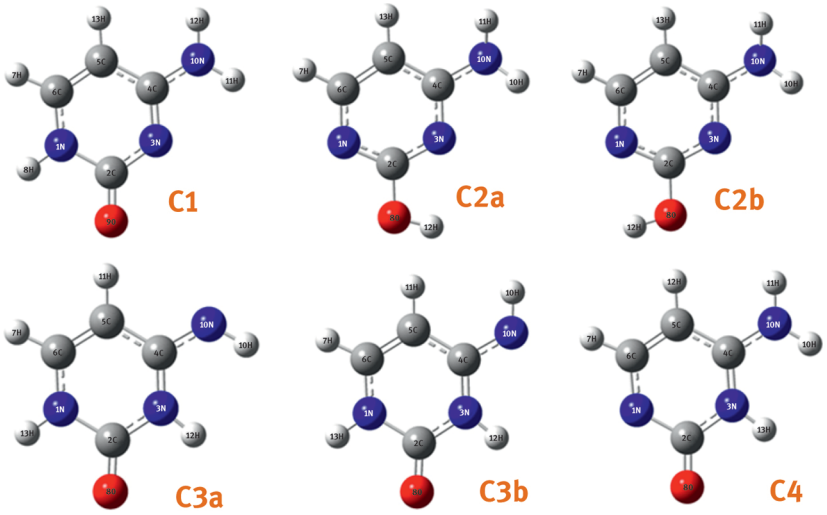

Figure 2: Cytosine tautomers with standard numbering and adopted nomenclature: Nonaromatic 2-oxo form (C1), aromatic 2-hydroxy *trans* form (C2a), aromatic 2-hydroxy *cis* form (C2b), non-aromatic 4-imino *cis* form (C3a), non-aromatic 4-imino *trans* form (C3b) and non-aromatic 3H-amino-oxo form (C4).

The bioactivity of 5-substituted Cy derivatives also generates exceptional interest in their biochemistry and pharmacology, and they are the most interesting and widely studied Cy derivatives. Among these compounds, 5-halogenated derivatives are of special relevance. Transformation of Cy into 5-halogen-cytosine significantly changes its chemical and spectroscopic properties, as well as its in vivo activity. The halogenated pyrimidines were synthesized in the 1950s as potential antitumor agents. Chlorinated pyrimidines are effective mutagens, clastogens and toxicants, as well as extremely effective inducers of sister-chromatid exchanges [17]. These chlorinated adducts can be mutagenic or perturb DNA–protein interactions [18]. Considering the importance of 5-halogencytosine derivatives for medicinal chemistry, their vibrational spectra, taken for low-temperature matrices and for the polycrystalline state, have not been much explored. Although the vibrational spectra of 5-fluoro, 5-chloro and 5-bromo-cytosines have been reported previously on the basis of normal coordinate analysis on semiempirical [19] and DFT [20] methods, there is much controversy in their assignments. Thus, 5-chlorocytosine (5-ClCy) and 5-bromocytosine (5-BrCy) are briefly analysed in the present manuscript.

3 Computational methods

Quantum chemical methods are commonly used to analyse and to interpret the molecular structure and the vibrational spectra of compounds. Its use is shown in the present study of the uracil and cytosine nucleic acid bases, and in two of their derivatives. The molecular structure of Cy is analysed from the data available in the bibliography that have been determined theoretically by quantum chemical methods and experimentally by X-ray diffraction. The values of the geometrical parameters are collected in Table 1. They are the most accurate calculated today for this compound.

Table 1: Selected equilibrium geometries of cytosine, bond lengths in Å and bond angles in degrees. The calculated values are with different ab initio and DFT methods and basis sets.

Parameters	HF 6-311+G (2d,p)	MP2 6-311++G (2df,2pd)[a]	cc-pVTZ[b]	B3LYP 6-311++G (3df,pd)	B-P 6-311++G (2df,2pd)[a]	BH-LYP 6-311++G (2df,2pd)[a]	CCSD[c] 6-31G(d,p)	Exp.[d]
Bond lengths								
N1-C2	1.3974	1.413	1.4083	1.4226	1.431	1.404	1.413	1.399
C2-N3	1.3604	1.373	1.3684	1.3664	1.373	1.356	1.385	1.356
N3=C4	1.2949	1.313	1.3079	1.3142	1.324	1.300	1.313	1.334
C4-C5	1.4430		1.4259	1.4357			1.457	1.426
C5=C6	1.3387		1.3503	1.3528			1.354	1.337
N1-C6	1.3449		1.3461	1.3495			1.363	1.364
C2=O	1.1938	1.218	1.2147	1.2139	1.226	1.201	1.220	1.237
C4-N9	1.3480	1.362	1.3556	1.3562	1.370	1.342	1.368	1.334
Bond angles								
N-C2-N	116.75		115.98	116.05			116.3	
C2-N=C4	120.35		120.21	120.66			119.6	
N3=C4-C5	123.84		124.25	123.75			124.6	
C2-N1-C6	123.02		123.75	123.28			123.7	
N3-C2=O	124.89		125.18	125.61			124.8	
Torsional angles								
N3=C4-N9-H12	10.08		11.7	6.96			13.7	
C5-C4-N9-H13	178.26		–18.9	–10.91			–25.4	
C2-N3=C4-N9			176.6	178.71			176.9	

[a]Ref [21]. [b]Ref [22]. [c]From ref [23]..[d]From X-ray and neutron diffraction data summarized in a statistical survey of the Cambridge Structural Database [23, 24].

Among the quantum chemical methods, DFT [25, 26] results were selected as the most appropriate. DFT methods provide a very good overall description of medium-size molecules. Moreover, for the wavenumber calculations [1, 27], they appear more accurate than Hartree-Fock (HF) and MP2 methods, and at lower computational cost. Among the DFT methods, the Becke's three-parameter exchange functional (B3) [28] in combination with the correlation functional of Lee, Yang and Parr (LYP) [29], i.e. B3LYP, appears as best and most

frequently used today. The results obtained with several basis sets, differing in size and contraction, are shown in the Table 1–Table 5 of the manuscript. The 6-311++G(3df,pd) basis set is in general too large for wavenumber calculations, due to the computational memory required. The 6-31G(d,p) and 6-311+G(2d,p) basis sets are optimum for this purpose. All the results were determined with the GAUSSIAN 09 [30] program package.

4 Scaling

The vibrational wavenumbers are usually calculated using the simple harmonic oscillator model. Therefore, they are typically larger than the fundamentals observed experimentally. This overestimation in the wavenumbers also depends on the type of vibrational mode and the range considered.

In general, Figure 3 shows the error in the wavenumber of a vibrational mode as computed in a variety of molecular environments, with different methods and sized basis sets. The vertical axis shows the difference between the true value in a given molecule of some particular vibrational mode, and the value computed with different methods and various sizes of self-consistent field (SCF) basis sets. The errors in computing the chosen vibrational mode in many different molecules are found to fall within the shaded area of the diagram. For small basis sets (or semiempirical calculations), the error, or range of uncertainty about the "true" value obtained from experiment, is largely increased. The convergence limit, approached by very large basis sets, still differs from the true value, but this residual error has been found empirically to be remarkably constant for a given parameter and is independent of the molecule studied. The calculations can be done efficiently at the point marked in the figure, and the residual error can be removed by the use of scaling, and therefore give rise to an accurate predicted wavenumber [1].

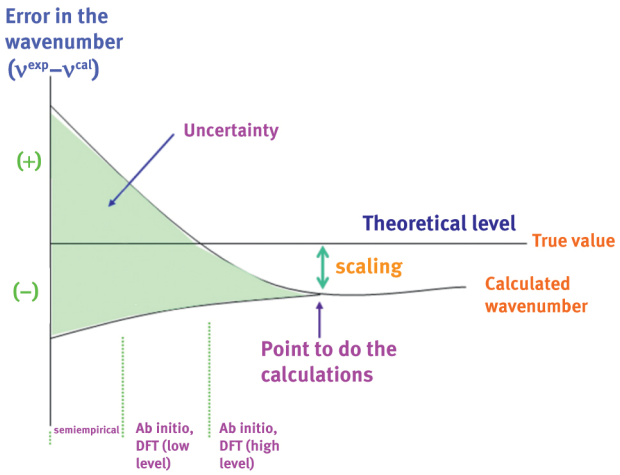

Figure 3: Schematic representation of the error in calculating a vibrational mode in a variety of molecular environments. For wide families of systems, the error is expected to fail within the shaded area.

5 Applications of computational chemistry to vibrational spectroscopy

Among all the possibilities, only four applications are shown here, but they are among the most used today. These applications were performed on the uracil, cytosine, 5-chlorocytosine and 5-bromocytosine molecules. These are the following:

a. Characterization of all the normal modes of a molecule.

b. Accurate assignment of all the bands of a spectrum.

c. Identification of the tautomers present in the gas phase and in the solid state of a compound.

d. Simulation of the crystal unit cell of a compound and the interpretation of its vibrational spectra, which remarkably improves the accuracy in the assignment of its spectra in the solid state.

5.1 Characterization of all the normal modes of a molecule

Computations at the B3LYP/6-31G(d,p) level has been used here for the characterization of all the normal ring modes in Cy (Figure 4) and U (see Ref. [8].). In the Cy molecule, only 30 ring modes are included in the figure. The number of the mode in increasing order appears in the centrum of each ring. Below the plot of each mode appears its calculated wavenumber (in cm^{-1}) and its main characterization. The atomic displacements for each computed wavenumber are determined as XYZ coordinates in the standard orientation, and they are plotted here to identify each vibration.

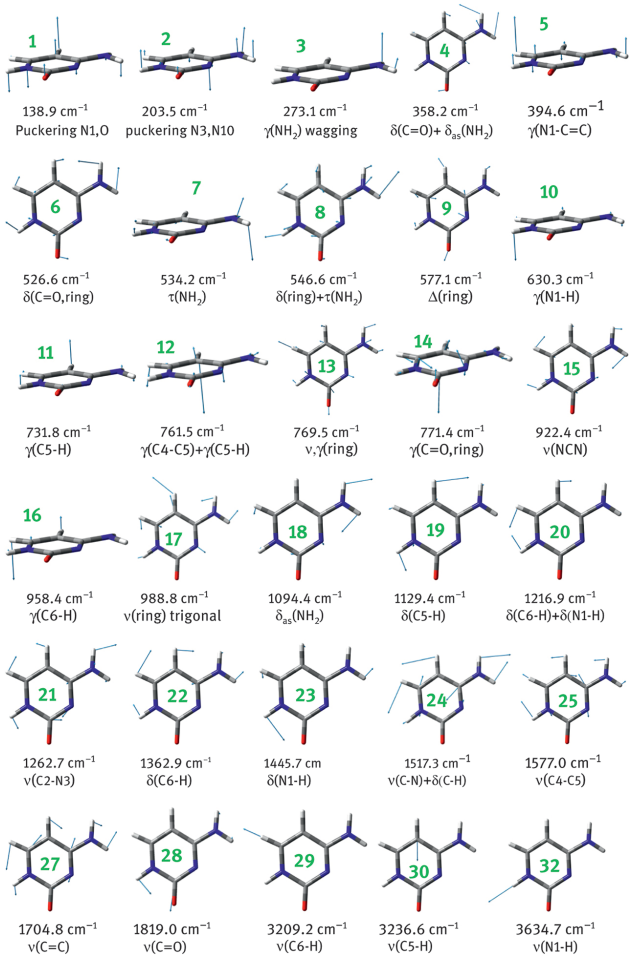

Figure 4: Characterization of the normal vibrational modes in the *keto* tautomer C1 of the cytosine molecule at the B3LYP/6-31G(d,p) level.

It is interesting to observe that a few skeletal ring modes retain a certain resemblance to the pseudo-normal modes of a hypothetical C_6 ring in the skeletal modes of benzene, although the uracil ring has no symmetry within the molecular plane. The previous characterization of the normal modes of a molecule by using computational methods led to an appropriate assignment of its bands, which in the present study correspond to the uracil and cytosine molecules and two of their derivatives.

5.2 Accurate assignment of all the bands of a spectrum

The computational methods provide information about the harmonic vibrational wavenumbers (in cm^{-1}), absolute infrared (IR) intensities (KM/Mole), Raman scattering activities (Å4/amu), Raman depolarization ratios for plane and unpolarized incident light, force constants (mDyne/Å), reduced masses (AMU) and normal coordinates displacement vectors of a molecule after its structure is optimized. All this information helps in the assignment of the bands in the experimental IR and Raman spectra.

5.2.1 Scaling procedures

The calculated wavenumbers generally have a large error for many reasons [1, 8, 9], such as anharmonicity, errors in the computed geometry, Fermi resonance, solvent effects, etc., and it is necessary to correct them. For this purpose, the first step after the calculation of the vibrational wavenumbers is the scaling, which helps to carry out an accurate assignment. To improve these computed wavenumbers, two scaling procedures can be used [8, 27].

(a) The first procedure uses a single overall scale factor for the calculated wavenumbers, ν^{exp}/ν^{calc}. This is the easiest way and thus is the procedure generally used in the bibliography to scale wavenumbers. To correct the overestimation of the calculated wavenumbers, several authors have reported scale factors for different levels. The most complete set of values has been determined by Scott and Radom [31], with the particularity of using two scale factors, one for the high- and medium-wavenumber vibrations, and another for low wavenumbers.

(b) A remarkable improvement in the accuracy of the scaled wavenumbers is obtained if a linear relationship is established between the calculated and experimental wavenumbers. This procedure called linear scaling equation (LSE) uses a scaling equation to correct the computed wavenumbers of a molecule at a specific level of theory. This scaling equation, obtained previously for a specific molecule (or group of molecules) with experimental gas-phase wavenumbers available (or Ar matrix values), can be used in related molecules and in their derivatives due to the good transferability of their parameters, which enables accurate assignment of the experimental values for many compounds. This procedure, developed by us, represents a compromise between accuracy and simplicity, and the results obtained are accurate enough for the standard today in large molecules [1, 8, 9].

Following this last procedure, the scaling equations for Cy and U molecules collected in Table 2 were obtained at different levels of computation.

Table 2: Linear scaling equations $\nu^{scaled} = a + b \cdot \nu^{calculated}$ obtained in cytosine and uracil molecules.

Levels of computation	*a*	*b*	Correlation coeffient, r
Cytosine molecule			
HF/6-31+G(d,p)	–4.1	0.8965	0.9997
HF/6-31+G(2d,p)	–14.3	0.9053	0.9997
B3LYP/6-31+G(d,p)	16.3	0.9560	0.9999
B3LYP/6-31+G(2d,p)	6.2	0.9631	0.9999
B3LYP/6-311+G(2d,p)	4.8	0.9671	0.9999
Uracil molecule			
HF/6-31G(d,p)	5.7	0.8928	0.9997
HF/6-31++G(d,p)	10.5	0.8938	0.9998
MP2/6-31G(d)	34.5	0.9372	0.9996
BP86/6-31G(d,p)	46.0	0.9678	0.9998
BLYP/6-31G(d,p)	46.4	0.9718	0.9998
B3P86/6-31G(d,p)	34.1	0.9389	0.9999
B3LYP/6-31G(d)	30.8	0.9468	0.9999
B3LYP/6-31G(d,p)	34.6	0.9447	0.9999
B3LYP/6-311+G(2d,p)	30.8	0.9538	0.9999
B3LYP/6-311++G(3df,pd)	31.9	0.9512	0.9999
B3LYP/dgdzvp	39.2	0.9472	0.9999
B3PW91/6-31G(d)	30.1	0.9421	0.9999
B3PW91/6-31G(d,p)	34.9	0.9393	0.9999

Table 3: Comparison of the calculated harmonic wavenumbers ($\nu^{cal.}$, cm^{-1}) at the B3LYP/6-31+G(d,p) level, the relative (*A*, %) infrared intensities, and scaled wavenumbers with the experimental values ($\nu^{exp.}$, cm^{-1}) with the % relative IR intensity in parenthesis. The characterization obtained (in parenthesis are the %PED contribution) in the normal ring modes of the cytosine molecule is also included together with the values of the absolute error obtained by different procedures.

No.	Calculated				$\nu^{exp.,\ c}$		Scaled[d]			Absolute error ($\omega^{cal.} - \nu^{exp.}$) in cm–1				
										Calculated[e]		With a factor[f]	With LSE[g]	
	$\nu^{cal.}$	$\nu^{cal.,\ a}$	*A*	Characterization[b]	Ne	Ar	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)

1	134	134	0	(100 %) puckering N1	220 (47)	236 (30)	134	144	138	–34	12	–34	–27	9
2	202	202.3	1	(100 %) puckering N3	342 (2)	343 (5)	156	165	202	15	15	15	16	8
3	156	257.4	31	(95 %) γ(NH_2) wagging	397 (6)	400 (4)	202	209	245	–2	2	–1	–3	–7
4	358	359.2	0	(43 %) δ(C=O) + (42 %) $\delta_{as}(NH_2)$ + (15 %) δ,Γ(ring)	511 (56)	507 (22)	358	359	351	20	16	21	13	3
5	398	398.9	3	(45 %) γ(N1-C=C) + (30 %) γ(ring)+ (25 %) τ(NH_2)	525 (32)	520 (57)	399	397	393	13	13	14	6	–1
6	527	520.3	2	(80 %) τ(NH_2)	531 (17)	535 (16)	528	520	510	10	11	11	2	–3
7	533	533.5	0	(60 %) δ(C=O)+ (25 %) δ,Γ(ring) + (15 %) $\delta_{as}(NH_2)$	571 (3)	568 (8)	534	526	519	9	11	10	0	–4
8	545	547.2	0	(80 %) δ(ring) mainly in N1, N3 + (20 %) τ(NH_2)	614 (27)	613 (17)	546	537	532	11	12	12	0	–5
9	577	580.7	0	(95 %) δ(ring) mainly in O8	711 (3)	710 (3)	578	568	564	14	19	–14	–2	–2
10	624	618.8	8	(70 %) γ(N1-H) + (25 %) γ(C4-N9)	717 (8)	717 (8)	625	613	608	46	51	17	29	29
11	724	722.8	5	(62 %) γ(C5-H) + (27 %) γ(ring) + (11 %) τ(NH_2)	749 (5)	747 (4)	696	708	708	22	28	–8	4	6
12	763	765.5	1	(75 %) δ,γ(ring) + (25 %) $\delta_s(NH_2)$	767 (3)	767 82)	734	746	746	6	22	–24	–12	–1
13	769	770.4	1	(43 %) γ(C4-C5)+ (30 %) γ(C5-H)+ (27 %) γ, δ(ring)	948 (0)	955 (1)	739	751	753	33	33	–5	6	3
14	773	789.2	5	(65 %) γ(C-N3=C) mainly in C2 + (35 %) γ(ring)	1085 (10)	1083 (12)	743	755	766	5	14	–37	–27	–20
15	924	923.5	1	(65 %) ν(ring) + (35 %) $\delta_{as}(NH_2)$	1103 (3)	1091 (6)	888	900	900	37	37	–7	4	2
16	962	963.8	0	(75 %) γ(C6-H) + (25 %) γ(C5-H)	1198 (34)	1196 (29)	925	936	933	24	25	–23	–13	–14
17	988	989.6	0	(80 %) ν(ring) + (20 %) $\delta_{as}(NH_2)$	1237 (11)	1244 (5)	950	961	958	20	17	–29	–19	–23
18	1088	1094.9	7	(50 %) $\delta_{as}(NH_2)$ + (35 %) δ(C=O) + (15 %) δ(ring)	1324 (17)	1320 (16)	1046	1056	1063	39	37	–13	–4	–7
19	1128	1126.6	0	(50 %) δ(C5-H)+(30 %) δ(N1-H) +(20 %) ν(C=C-N)	1382 (9)	1382 (17)	1084	1095	1093	61	63	5	14	16
20	1220	1213.3	6	(50 %) δ(HC6-NH)+(35 %) δ(C5-H)+(15 %) ν(N-C6)	1441 (47)	1439 (51)	1173	1183	1182	69	67	11	19	18
21	1264	1251.5	4	(48 %) ν(C2-N3)+22 % ν(C-N9)+17 % $\delta(NH_2)$+13 % δ(ring)	1540 (18)	1539 (21)	1215	1225	1221	31	29	–30	–22	–23
22	1359	1355.5	8	(37 %) δ(C6-H) + (34 %) ν(C4-N9) + (22 %) ν(ring)	1569 (18)	1570 (14)	1307	1316	1313	67	68	4	11	14
23	1443	1441	12	(42 %) δ(N1-H)+ (30 %) ν(C-N3-C) +(22 %) ν(N1-C6)	1625 (100)	1622 (91)	1387	1396	1398	70	65	5	12	9
24	1508	1499.9	23	(40 %) ν(C-N9)+(25 %) δ(C6-H)+(20 %) δ(ring)+(15 %) $\delta(NH_2)$	1725 (58)	1718 (100)	1450	1458	1457	56	49	–12	–6	–10
25	1570	1562.2	21	(55 %) ν(C4-C5) + (32 %) ν(ring) + (13 %) $\delta(NH_2)$	3474 (33)	3471 (12)	1509	1517	1516	139	129	0	–4	2
26	1637	1632.6	17	(90 %) $\beta_s(NH_2)$	3575 (36)	3564 (18)	1574	1581	1584	152	135	7	3	3
27	1692	1681.4	64	(50 %) ν(C=C) + (30 %) ν(N3=C4) + (12 %) ν(ring)	3618 (50)	3600 (18)	1627	1634	1631					
28	1774	1758	100	(80 %) ν(C=O) + (20 %) ν(ring)			1706	1712	1708					
29	3211	3191.7	0	(75 %) ν(C6-H) + (25 %) ν(C5-H)			3087	3086	3090					
30	3236	3216.8	0	(75 %) ν(C5-H) + (25 %) ν(C6-H)			3111	3110	3116					
31	3610	3590.6	11	(98 %) $\nu_s(NH_2)$			3471	3467	3473					
32	3629	3616.1	10	(98 %) ν(N1-H)			3489	3486	3499					
33	3752	3719.3	6	(100 %) $\nu_{as}(NH_2)$			3607	3603	3603					

Rms[h]	53.5	49	17.2	13.5	12	

[a]With the 6-311+G(2d,p) basis set. [b]Abbreviations: ν, stretching; δ, in-plane bending; γ, out-of-plane bending. [c]Ref. [33]. [d]Wavenumbers scaled: (a) with the scale factor of 1.0013 for calculated wavenumbers lower than 800 cm^{-1}, and the scale factor of 0.9614 for higher wavenumbers [31]. (b) With the corresponding scale equations of Table 2. (c) With the 6-311+G(2d,p) basis set and with the corresponding scale equation of Table 2. [e] Absolute error in the calculated wavenumbers: (d) from the second column. (e) From the third column. (f) Error calculated from column (a). (g) Errors in the scaled wavenumbers: (g) From column (b). (h) From column (c). h Rms, defined as $(\sum (\omega^{cal.} - \nu^{exp.})^2/n)^{1/2}$, where the sum is over all the modes, n, and $\nu^{exp.}$ corresponds to the Ar matrix values.

5.2.2 Assignment of all the bands in the cytosine molecule

The vibrational bands computed in tautomer C1 of the Cy molecule with B3LYP and two different basis sets are examined in Table 3. The first column refers to the numbers assigned to the calculated vibrations, and they are given in increasing order of wavenumbers. These normal modes of Cy structure appear plotted in Figure 4. The second and the third column list the calculated wavenumbers with the 6-31G(d,p) and 6-311+G(2d,p) basis sets, respectively. The relative IR intensities of the fourth column were obtained by normalizing the computed values to the intensity of the strongest line, no. 28. Small differences appear between the calculated values with both basis sets. The value of the IR intensity helps to carry out the match of theoretical to experimental wavenumbers. Although wavenumber calculations on Cy have been reported with other quantum chemical methods [22, 32], the values shown in Table 3 represent the most accurate today.

The characterization established by B3LYP for each calculated wavenumber is shown in the fifth column. The percent contribution of the different modes to a computed wavenumber appears in parentheses. Contributions lower than 10 % were not considered. The sixth and seventh columns collect the experimental wavenumbers reported in neon and argon [33, 34] matrices. The most detailed study corresponds to an Ar matrix [34], and thus their values were selected as reference.

The scaled wavenumbers of the eight to tenth columns were obtained following the scaling procedures mentioned above. These scaled wavenumbers refer to an isolated molecule, and thus they can be directly compared to the experimental wavenumbers in an Ar matrix. From this match, the experimental bands can be assigned. This scaling was carried out in the calculated wavenumbers of tautomers C1, C2a, C2b, C3a and C3b of Cy. When the scaled wavenumbers of these tautomers are compared with the experimental values reported in an Ar matrix, it was not possible to identify the characteristic bands corresponding to the *enol* (C2a, C3b) or *imino* (C3a, C3b) tautomers. Thus, only the *keto* form C1 appears clearly in gas phase. However, in other studies, both C1 and C2 tautomers have been detected with a small amount of C3 [35].

In addition, with the help of computational methods, it is possible to modify the assignment reported by other authors. Thus, in Cy, the calculated C2=O stretch (scaled at 1,712 cm^{-1}, ninth column) is predicted as the strongest IR band in accordance with that observed in an Ar matrix spectrum at 1,718 cm^{-1} [34]. However, Radchenko et al. [36] pointed out that the most intense IR band in an Ar matrix at 1,620 cm^{-1} corresponds to the ν(C=C) stretch. We have corrected all errors in Table 3. For the discussion of the remaining bands, Table 3 is self-explanatory.

5.2.3 Assignment of all the bands in the uracil molecule

Table 4 shows the results obtained in the U molecule with the corresponding scaling equations of Table 2. The first column refers to the notation used for the ring normal modes [8, 27]. The main characterization of these modes appears in the second column. The third and fourth columns list the calculated wavenumbers with the B3LYP method and the 6-31G(d,p) and 6-311+G(2d,p) basis sets, respectively. The column with relative intensities was obtained by dividing the computed values by the intensity of the strongest line. Although wavenumber calculations on uracil have been determined by other authors using MP2 [37] and B3LYP DFT methods [38, 39], the values shown here represent the most accurate today.

Table 4: Comparison of the experimental wavenumbers (ν^{exp}, cm^{-1}) and calculated harmonic values ($\nu^{cal.}$, cm^{-1}) with the 6-31G(d,p) basis set, together with their relative (A, %) infrared intensities, and characterization obtained in the normal ring modes of uracil molecule with several methods. Absolute error (Δ) obtained in the calculated and scaled (ν^{scal}) wavenumbers.

No.	Characterization[a]	B3LYP	Infrared[c]	Absolute error $\Delta(\nu^{cal.} - \nu^{exp})$ in cm^{-1}	B3LYP[b,f]

		ν^{cal}	$\nu^{cal.,\ b}$	A	Ar matrix[d]	Gas[d]	Gas[e]	B3LYP	B3LYP[b]	B3PW91	ν^{scal}	$\Delta(\nu^{scal}-\nu^{exp})$
1	Puckering N3	150	147	0	185 w	185 w	374 vw	–15	–20	–14	171.1	-
2	Puckering N1	170	165	0	391 m	377 m	395 w	–6	–4	–6	188.2	3
3	δ(OCNCO)	385	387	4	411 m	411 m	512 w	1	0	1	400	9
4	γ(C=C-H12)	396	395	4	516.5 m	527 m	545 w	7	11	6	407.6	13
5	δ(ring)	519	523	4	536.4 m	588 w	659.5 w	5	7	6	529.7	18
6	δ(ring) + δ(C=O)	541	543	1	559 w	556 m	717.4 vw	–1	–12	0	548.8	13
7	δ_{as}(ring) + δ(C=O)	558	547	1	551.2 m	633 m	756.5 w	18	15	25	552.6	–6
8	γ(N1-H)	563	560	8	662.1 s	769 s	802 w	27	17	32	565	20
9	γ(N3-H)	687	677	14	718 w	810 s	952 w	12	11	14	676.6	18
10	γ(C4=O) + γ(C5-H)	729	728	2	759.2 sh?	946 vw	972 sh	–5	8	2	725.2	8
11	γ(C2=O)	752	765	10	804 s	974 w	990 sh	–10	–13	–3	760.5	4
12	ν(ring)	772	769	0	958.3 w	999 w?	1082 m	11	17	12	764.3	5
13	γ(C5-H) + γ(C4=O)	813	819	10	963 w	1089 s?	1172 s	13	10	18	812	10
14	ν(C-C) + δ(N-H)	965	962	1	982 w	1228 m	1356 sh	–2	0	1	948.4	–4
15	γ(C6-H)	970	972	0	1073 w	1360 m	1387 s	0	4	3	957.9	–14
16	δ(NCC)	990	994	1	1184 vs	1380 s	1400 s	18	13	24	978.9	–11
17	ν(ring) + δ(C5-H)	1091	1086	1	1217.4 w	1396 ms	1461 s	26	19	37	1066.7	–6
18	ν(C-N)+δ(C6H,N1H)	1198	1191	14	1359.3 vw	1480 s	1641 s	44	40	50	1166.8	–5
19	δ(C5-H) + δ(N-H)	1231	1227	1	1388.7 vs	1632 vs	1756 vs	26	26	31	1201.2	–16
20	δ(N3-H) + δ(C-H)	1382	1382	3	1399.6 vs	1688 vs	3124 m	20	19	28	1349	–7
21	ν(C-N) + δ(N3-H)	1407	1406	20	1472 ms	1734 vs	3436 s	22	23	26	1371.9	–15
22	δ(C6-H) + δ(N-H)	1422	1423	3	1644 m	3076 w	3484 s	45	37	57	1388.1	–12
23	δ(N1-H) + ν(N1-C)	1506	1498	18	1741 vs	3101 w		49	29	60	1459.6	–1
24	ν(C=C)	1690	1670	12	1757.5 vs	3427 w		67	16	85	1623.7	–17
25	ν(C4=O)	1808	1757	100	3434.5 s	3450 w		89	35	108	1706.7	–34
26	ν(C2=O)	1845	1791	92	3484.3 s			145	124	155	1739.1	–17
27	ν(C6-H)	3221	3200	1				140	118	154	3083	7
28	ν(C5-H)	3264	3242	0				184	156	**206**	3123.1	–1
29	ν(N3-H)	3620	3592	11				174	152	197	3456.9	21
30	ν(N1-H)	3658	3636	17							3498.9	15
							Rms[g]	66	54	75		13.7

[a] Abbreviations: ν, stretching; δ, in-plane bending; γ, out-of-plane bending. [b]With the 6-311+G(2d,p) basis set. [c]Notation: sh, shoulder; vs, very strong; s, strong; m, medium; w, weak; vw, very weak. [d]Refs. [40, 41]. [e]Ref. [42]. [f]With the scaling equations. [g]Rms, defined as $(\sum(\omega^{cal.}-\nu^{exp.})^2/n)^{1/2}$, where the sum is over all the modes, n, and $\nu^{exp.}$ corresponds to the values underlined.

Unfortunately, few studies have reported the gas-phase vibrational spectrum of U. The experimental values selected in the seventh to ninth columns were those reported in an argon matrix [40, 41] and in the gas phase [40, 42]. For the determination of the scaling equations, the values reported in the gas phase by Colarusso et al. [42] were selected as reference, since the assignments given there correspond most closely to our own. These wavenumbers appear underlined in Table 4. In the lack of these data, the values obtained in an Ar matrix were used. The scaling equations obtained in this way and listed in Table 2 have a good transferability to uracil derivatives [27]. With this correction (scaling) of the calculated wavenumbers, in general, they are remarkably close to the experimental wavenumbers, and thus they can be used for the assignments.

In the assignment of the uracil modes, Table 4 is self-explanatory, and only a few comments need be made: The N-H stretches appear as pure modes and with strong intensity, and thus they were clearly characterized and assigned. The N-H in-plane bendings are more complex because they appear mixed with other modes. These N-H modes have noticeable contributions to seven calculated vibrations. The description of mode 22 is complex. Mainly, it was characterized as δ(C6-H)+ δ(N-H) with some contribution of ν(C-N), in agreement with the assignment reported by Harsányi et al. [40] and Colarusso et al. [42]. Mode 21 is also complex and it has

been characterized as ν(C-N)+ δ(N3-H) in agreement with Harsányi et al. [40] but in contrast to Colarusso et al. [42] and Lés et al. [43].

In the gas phase, the γ(N1-H) out-of-plane bending mode is clearly related to the band at 556 cm^{-1} [40] or at 545 cm^{-1} [42], in disagreement with Aamouche et al. [37], who related this band to the two calculated wavenumbers corresponding to δ(C2=O) and γ(N1-H) modes. The assignment of the very weak gas-phase band at 672 cm^{-1} [40] or at 692 cm^{-1} [42] should be changed and related to the γ(N3-H) mode.

The calculated C2=O and C4=O stretches are the strongest IR bands, and they are predicted with similar intensity. The scaled wavenumbers of both modes are at 1,739.1 and 1,706.7 cm^{-1} (fourteenth column in Table 4), modes C2=O and C4=O, respectively, in good agreement with the gas-phase results. The C=O out-of-plane bending appears in general mixed with γ(C-H) modes. The description of mode 13 as γ(C5-H)+ γ(C4=O) is in agreement with Colarusso et al. [42] and other authors [38, 43, 44]. The band reported in the gas phase [42] at 717.4 cm^{-1} and in an argon matrix at 707.4 and 719 cm^{-1} by Ivanov et al. [41] and Szczepaniak et al. [38], respectively, were related here to mode 10, defined as γ(C4=O) + γ(C5-H).

As we show here, with accurate scaling, it is possible to carry out a good assignment and in some cases to correct that reported by other authors.

5.2.4 Scaling in Cy and U molecules: accuracy of the different methods

The accuracy of the two scaling procedures mentioned above on the Cy molecule and the two theoretical methods used is shown in Table 3. The absolute error obtained in the wavenumbers is shown in the last five columns. They were calculated with the experimental values in an Ar matrix. In the last row, the root mean square (rms) error of each level and procedure is determined. With the LSE procedure, the errors obtained in the predicted wavenumbers were very small; the mean deviation was 13 cm^{-1} (1 %). The values of these rms errors were very close to those obtained in other related compounds studied by us [9, 45, 46]. This good match of scaled to experimental wavenumbers helps in the assignment and analysis of the experimental fundamental modes.

Similar to the Cy molecule, the absolute errors obtained in the calculated and scaled wavenumbers of the U molecule are shown in Table 4. The largest values correspond to the calculated wavenumbers and they are marked in bold type. However, with the scaling equation they are remarkably reduced as shown in the last column of the Table. The bottom of the table shows the rms error obtained for the calculated and scaled wavenumbers at several computational levels.

Calculations at other levels were also carried out on the U and Cy molecules. The errors obtained in the U molecule and by the two scaling procedures are shown in Table 5. As can be seen, remarkable differences appeared between the HF and B3LYP methods. In both cases, the large rms error in the calculated wavenumbers is remarkably reduced with the scaling, especially by HF. However, the error obtained after scaling by HF continues to be higher than by B3LYP. That is the HF method should not be used for assignment of the experimental wavenumbers.

Table 5: Errors obtained in the calculated and scaled wavenumbers of the uracil modes by the different procedures and methods.

Method	Calculated wavenumbers			Scaled wavenumbers with an overall factor			Scaled wavenumbers with LSE		
	rms	Largest error (cm^{-1})		rms	Largest error (cm^{-1})		rms	Largest error (cm^{-1})	
		Positive	Negative		Positive	Negative		Positive	Negative
HF/6-31G(d,p)	184	427	6	23	53	37	22.6	46	57
HF/6-31++G(d,p)	177	418	11	37	53	95	16.7	27	50
MP2/6-31G(d)	82	187	44	33	56	50	25.4	50	66
BP86/6-31G(d,p)	35	86	44	34	54	54	18.1	34	32
BLYP/6-31G(d,p)	34	73	49	24	46	44	19.6	34	36
B3P86/6-31G(d,p)	77	207	14	21	44	40	15	32	26
B3LYP/6-31G(d,p)	66	184	15	25	50	41	13.8	24	23
B3LYP/6-311+G(2d,p)	54	156	20				13.7	21	34
B3PW91/6-31G(d,p)	75	206	14				14.9	30	26
M06-2X/6-31G(d,p)	79	192	13				19.1	53	31
M06-L/6-31G(d,p)	70	190	13				19	49	25

With the use of a scaling equation, a remarkable reduction of the error is obtained by B3LYP and thus of the risk to a mistake. The best for this purpose is the 6-311+G(2d,p) basis set shown in the last three columns of Table 5. Larger basis sets represent an excessive increase in the computational cost for a very slight improvement. The largest absolute error corresponds to vibration no. 12, a ring bending. Large errors also appear in the vibrations nº 21 and 25, corresponding to C-N and C-C stretches, respectively.

Finally, it can be concluded that the LSE procedure leads to very low errors, and lower than using an overall scaling factor. With B3LYP, the wavenumbers are significantly close to the experimental and much better than by HF. This B3LYP method is also the best among the DFT methods used, and it is the recommended method for carrying out an assignment of the vibrational bands.

5.3 Identification of the tautomers present in the isolated state

5.3.1 Tautomerism in nucleic acid bases

The nucleic acid bases can undergo *keto-enol* tautomerism. Much of the interest of the tautomerism is due to the fact that tautomers induce alterations in the normal base pairing, leading to the possibility of spontaneous mutations in the DNA or RNA helices. On the basis of geometry, ionization potential and dipole moment, it is virtually impossible to decide which arrangement is actually the most stable one. However, computational methods applied to the vibrational spectra may help in deciding the most stable arrangement (tautomer). The absorption bands due to C=O, NH and OH groups give the most straightforward information about the tautomeric forms present, because they correspond to the characteristic, well-localized vibrations of the functional groups directly involved in tautomeric changes. Thus, special attention was paid on these modes.

All the tautomers in the Cy molecule appear with relative energies much lower than their counterparts in the U molecule. Two features can explain it: (i) the negative charge on N1 atom is lower in Cy than in U, with longer N1-H bond length. (ii) The negative charge on the O2 atom is higher in Cy than in U, with longer C=O bond length. Both features favor tautomerism in Cy molecule. Thus, there is a special interest to study this molecule and its derivatives [15, 16, **47**].

The Cy molecule has been studied intensively by IR spectroscopy in the free monomeric form (in low-temperature inert matrices [34, 36]) and in the crystalline phase at room [48] or at low temperatures [49]. The different aggregate states can contain two tautomeric forms [33, 50]: the gas phase with both the *enol* and the *keto* forms, whereas in the crystalline state and polar solvents only the *keto* form has been observed [34]. For its interpretation, the theoretical vibrational spectrum of Cy has been predicted at different levels of approximation [34, 51].

5.3.2 Tautomerism in 5-bromocytosine

Similar to the Cy molecule, 5-BrCy can exist in various tautomeric forms. However, this molecule has been much less studied than Cy, and thus it is discussed here. The computational methods have been applied to vibrational spectroscopy to simulate (scaled) the IR spectrum of the different tautomers of 5-BrCy. In the isolated state, the *enol* form C2b of 5-BrCy is the most stable one. The next most stable tautomer is the *enol* form C2a, 3.18 kJ/mol above C2b in 5-BrCy, and 12.96 kJ/mol above C2b in the Cy molecule with the MP2 method. The bromine atom in position 5 favors the tautomerism with much lower relative energies (about 66 %) in 5-BrCy than in Cy.

In the isolated state, the harmonic vibrational bands computed in the 5-BrCy ring are shown in Table 6. The second column lists the calculated wavenumbers with the 6-31G(d,p) basis set in tautomer C1. The third column collects their relative IR intensities (A) in percent. They were obtained by normalizing the computed value to the intensity of the strongest band. The assignment with the calculated percent potential energy distribution (PED) of the different modes for each vibration appears in the fourth column. Contributions lower than 10 % were not considered. The scaled wavenumbers in the fifth column correspond to tautomer C1 and those in the ninth column to tautomer C2b. With these data obtained theoretically, the experimental spectrum in an Ar matrix reported by Jaworski et al. [52] can be now analyzed in detail. For this task, the scaled values were directly compared to the experimental values, and from this comparison it was possible to separate those corresponding to tautomer C1 (sixth column), from those to tautomer C2b (tenth column). In Table 6, only the experimental bands with high intensity are shown, and those selected in the comparison with the scaled values are underlined.

Table 6: Comparison of the calculated harmonic wavenumbers (ν^{cal}, cm^{-1}), relative infrared intensities (A, %), scaled harmonic wavenumbers (ν^{scal}, cm^{-1}), experimental IR values ($\nu^{exp.}$), and characterization obtained in gas phase of both tautomers of 5-BrCy molecule at the B3LYP/6-31G(d,p) level.

No	Tautomer C1					Tautomer C2b			
	ν^{cal}	A	Characterization	$\nu^{scal,\,a}$	Exp.[b]	ν^{cal}	A	$\nu^{scal,\,a}$	Exp.[b]
1	3630	9	(95 %) ν(N1-H)	3464	3454, 3441	-	1	-	1588
2	3228	0	(99 %) ν(C6-H)	3084	1752, 1724, 1718	3189	5	3047	1556
3	1822	100	(79 %) ν(C=O) + (14 %) δ(N1-H)	1756	1629, 1618, 1609	1635	39	1579	1442, 1437
4	1696	49	(41 %) ν(C=C) + (38 %) ν(C-N) + (12 %) $\beta_s(NH_2)$	1637	1507	1598	70	1544	1300, 1296
5	1556	19	(49 %) ν(N3-C4) + (34 %) ν(N1-C-C) + (11 %) $\beta_s(NH_2)$	1505	1434	1489	6	1441	1258
6	1499	17	(48 %) ν(N3CC5) + (24 %) $\delta(NH_2)$ + (20 %) ν(N1-C6)	1451	1420, 1418	1335	3	1295	1055
7	1435	7	(40 %) δ(N1-H) + (33 %) ν(N-C4)+(16 %)Δ(ring)	1390	1286	1312	6	1274	950
8	1331	0	(49 %) δ(C-H) + (29 %) ν(CCC) + (12 %) δ(N1-H)	1292	1242, 1234, 1220	1057	1	1033	976
9	1269	3	(42 %) ν(C2-N3)+ (25 %)ν(CN10)+(21 %) Δ(ring)	1233	1178, 1171	977	3	957	789
10	1197	7	(72 %) δ(H-N1-C6-H) + (17 %) ν(ring)	1165	773	988	7	967	779
11	1039	6	(94 %) ν(ring)	1016	611, 605	799	1	789	652
12	937	1	(91 %) γ(C6-H)	920	545	795	0	785	601
13	916	1	(65 %) ν(ring) + (35 %) $\delta_{as}(NH_2)$	900		731	2	725	530, 526, 518
14	772	4	(84 %) γ(N-CO-N) mainly in C2 + (16 %) Γ(ring)	764		650	1	648	
15	764	1	(74 %) ν(ring) + (26 %) $\delta_s(NH_2)$	756		581	1	583	
16	745	0	(66 %) γ(C4-C)+(21 %) Γ(ring) + (13 %) $\tau(NH_2)$	738		531	0	536	
17	627	2	(67 %) ν(ring) + (21 %) $\delta_{as}(NH_2)$ + (12 %) ν(C-Br)	627		455	2	464	
18	625	8	(71 %) γ(N1-H) + (16 %) Γ(ring) +(13 %) $\tau(NH_2)$	625		369	1	383	
19	576	1	(45 %) Δ(ring) + (33 %) δ(C=O) + (22 %) $\delta(NH_2)$	579		282	1	301	
20	536	0	(70 %) Δ(ring) mainly in N1, N3 + (30 %) $\delta_{as}(NH_2)$	541		300	3	318	
21	396	1	(55 %) γ(N1-CH=C) + (30 %) Γ(ring) + (15 %) $\tau(NH_2)$	409		220	0	242	

22	389	0	(36 %) δ(C=O) + (28 %) Δ(ring) + (26 %) $\delta_{as}(NH_2)$	402	201	18	224
23	282	0	δ(C-Br) + (29 %) Δ(ring) + (12 %) $\delta(NH_2)$	301	84		113
24	255	0	(85 %) puckering N1 +(15 %) $\tau(NH_2)$	275			
25	205	10	(51 %) puckering N3 + (49 %) $\gamma(NH_2)$	228			
26	204	1	(41 %) δ(C-Br) + (31 %) Δ(ring) + (28 %) $\delta(NH_2)$	227			
27	77	1	(30 %) Γ(ring) + (46 %) γ(C=O) + (24 %) $\gamma(NH_2)$	107			

[a]With the scaling equation: $\nu^{scaled} = 34.6 + 0.9447 \cdot \nu^{calculated}$, ref [8].. [b] IR in Ar matrix, ref [52].. The highest IR intensity is shown underlined.

Through the comparison of these spectra, we have identified, separated and assigned the bands corresponding to the different tautomers, which in this case are only the amino-oxo form (tautomer C1) and the amino-hydroxy form C2b which is the most stable form in accordance to the computations, Figure 5. Due to many bands between both conformers appear very close in frequency, the IR intensity criterion was followed in these cases. Thus, the assignment reported previously by Jaworski et al. [52] has been improved by using scaling, and all bands accurately assigned to both tautomers. Bands corresponding to tautomer C3b, simulated theoretically, were not found in the experimental spectrum.

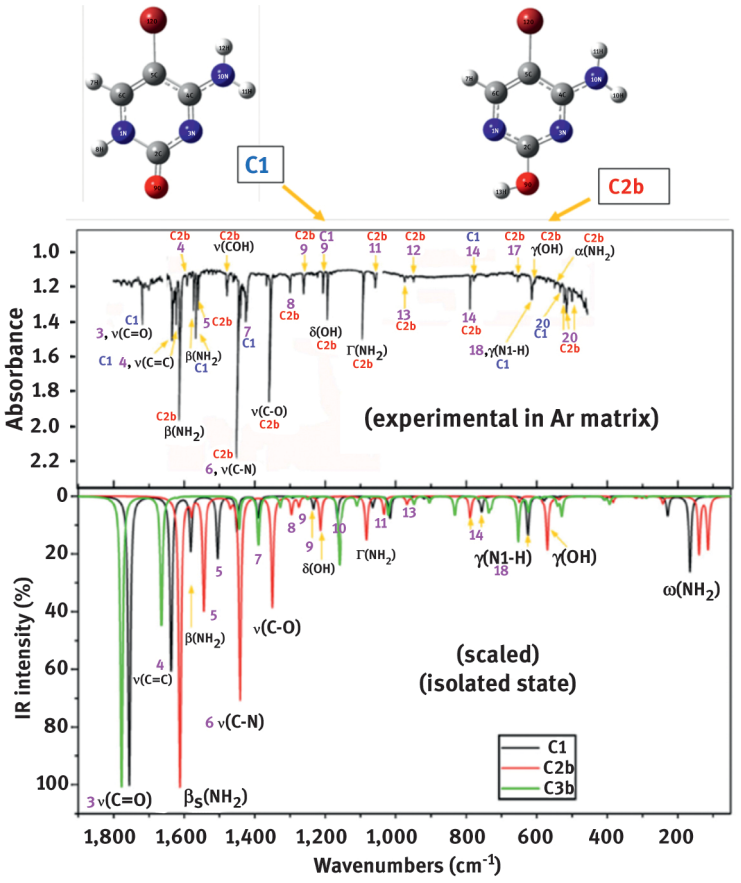

Figure 5: Assignment to tautomers C1 and C2b of the experimental IR spectrum in argon matrix of 5-bromocytosine.

5.4 Simulation of the crystal unit cell of a compound and the interpretation of its vibrational spectra

Figure 6 shows the optimized simulated tetramer form of the Cy molecule obtained from the crystal unit cell reported by X-ray [53]. The molecules adopt their amino-oxo tautomeric form, in which they are associated

to form ribbons stabilized by N-H···O hydrogen bonds that involve N1-H and NH_2 groups and the carbonyl oxygen atom. Four molecules appear in the crystal unit cell. Unfortunately, the detailed analysis of this structure has not yet been reported. Thus, the molecule of 5-chlorocytosine (5-ClCy) has been selected as example.

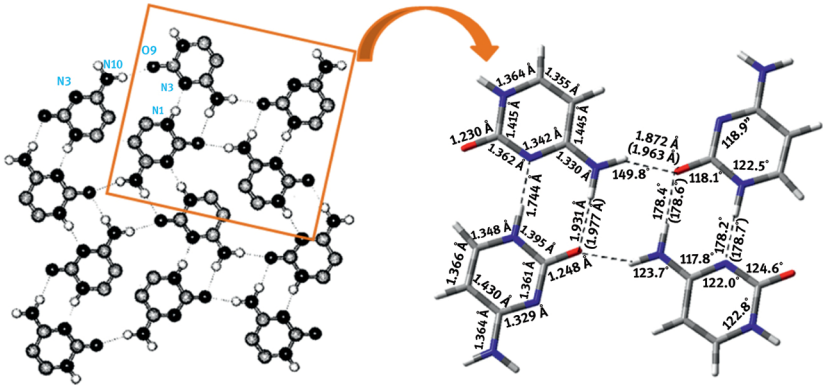

Figure 6: Schematic representation of the crystal structure of cytosine in the solid state [16, 53]. Hydrogen bonds are formed by all three available protons and they are indicated by dashed lines. At the right side appears the simulation at the B3LYP/6-31G(d,p) level of the crystal unit cell of Cy through a tetramer form. In parentheses appears the values obtained at the O3LYP/6-31G(d,p) level.

5.4.1 5-Chlorocytosine

The crystal unit cell of 5-ClCy was simulated at the B3LYP/6-31G(d,p) level by a tetramer form, Figure 7. Its simulated spectrum appears closer to the experimental spectrum in the solid state than the simulated spectrum of the monomer form. Figure 8 only shows the comparison between the experimental (IR) and that scaled with the tetramer form. Similar agreement is observed with the Raman spectrum.

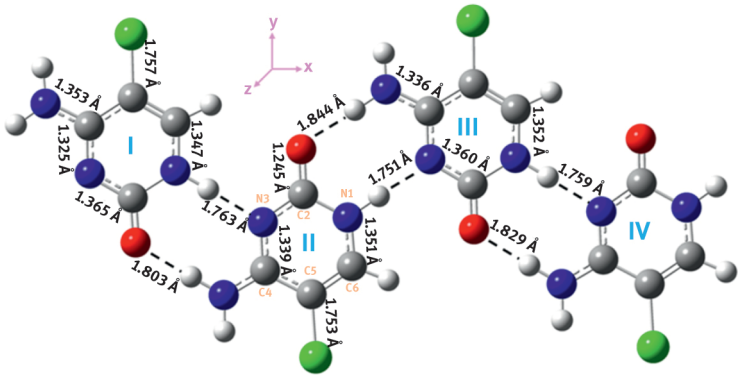

Figure 7: Simulation of the crystal unit cell of 5-chlorocytosine by a tetramer form at the B3LYP/6-31G(d,p) level.

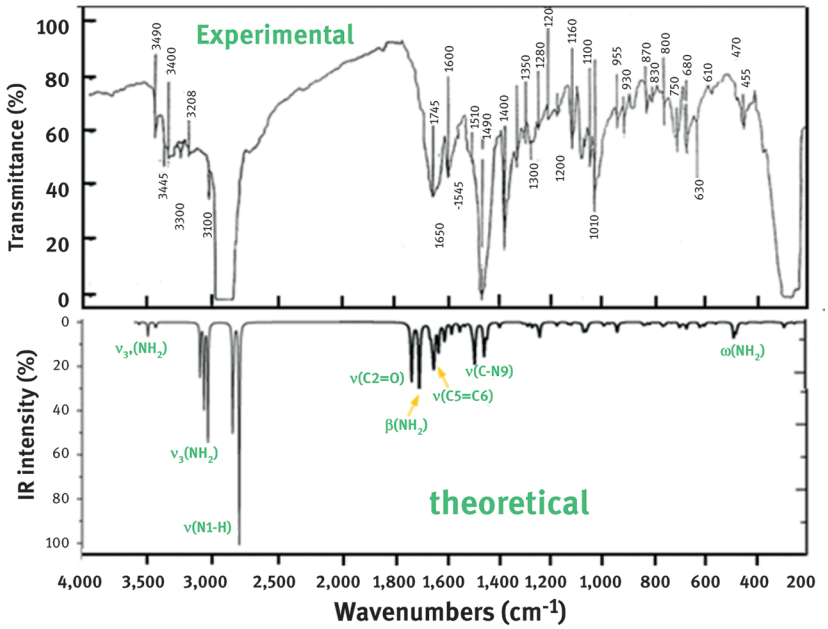

Figure 8: Experimental and theoretical IR spectra of 5-chlorocytosine in the 4,000–200 cm^{-1} range. The predicted scaled spectrum was obtained using the LSE procedure in the tetramer form at the B3LYP/6-31G(d,p) level.

The scaled and experimental wavenumbers in the monomer and tetramer forms are collected in Table 7, together with the corresponding assignment. They are the most accurate today for this molecule. The scaling equation $\nu^{scaled} = 4.8 + 0.9671 \cdot \nu^{calculated}$ deduced from the Cy molecule (Table 3) at the B3LYP/6-311+G(2d,p) level was used for the monomer, while $\nu^{scal} = 34.6 + 0.9447 \cdot \nu^{cal}$ at the B3LYP/6-31G(d,p) level was used for the tetramer. Four wavenumbers appear in the tetramer form for each characterized vibration corresponding to its four molecules. Of these four wavenumbers, those determined in the molecules II and III (see Figure 7) are underlined. These wavenumbers appear affected by the internal H-bonds of the crystal, and they represent better the solid state. Among these wavenumbers, that with the highest IR intensity appears in bold type, while that with the highest Raman intensity is shown in italic type. The main characterization carried out, and corresponding to the tetramer form, is shown in the last column.

Table 7: Comparison of the experimental wavenumbers (cm^{-1}) and harmonic calculated wavenumbers (ν^{cal}, cm^{-1}), relative infrared intensities (A, %), relative Raman scattering activities (S, %), scaled wavenumbers (ν^{scal}, cm^{-1}), and characterization obtained in 5-ClCy molecule calculated at B3LYP/6-311+G(2d,p) level.

Monomer				Tetramer	Exp.[b]		Characterization
ν^{cal}	A	S	ν^{scal}	$\nu^{scal, a}$	IR	Raman	
3731.4	9	23	3602.2	3590, 3523, ***3521***, 3520	3451 s	3448 ms	ν_{as}(N-H) in NH_2
3611.7	9	100	3487.3	3468, 2872, 2864, ***2822***	3320 w	3320 ms	ν(N1-H)
3599.9	13	74	3476	3457, 3121, 3093, ***3063***	3182 w	3174 m	ν_s(N-H) in NH_2
3207.9	0	54	3099.8	***3091***, 3085, 3084, 3081	3090 w	3081 m	ν(C6-H)
1761.1	100	25	1711.3	1738, **1707**, 1686, *1666*	1708 w	1708 ms	ν(C2=O)
1678.5	46	6	1632.1	**1633**, *1610*, 1608, 1608	1630 b	1626 ms	ν(C5=C6) + ν(ring) + $\beta_s(NH_2)$
1627	20	7	1582.6	1658, *1653*, **1650**, 1582	1655 ms	1651 ms	$\beta_s(NH_2)$ + ν(C-N)
1542.6	19	12	1501.6	**1550**, 1540, 1530, *1500*	1610 s	1604 s	ν(CCCN3) + δ(N1-H)
1485.1	24	4	1446.4	1503, *1496*, **1492**, 1455	1525 ms	1538 ms	ν(C-N10) + ν(NC6) + $\beta_s(NH_2)$ +δ(CH)
1430.7	8	5	1394.2	1454, 1450, ***1445***, 1395	1510 ms	1512 m	ν(N1C6) + δ(N1-H) +ν(C4C5)+ $\beta_s(NH_2)$
1325.9	0	7	1293.7	1303, ***1302***, 1300, 1297	1472 s	1470 s	δ(C-H) + δ(N1-H) + ν(C-N10)
1256	3	9	1226.6	1285, 1284, ***1267***, 1256	1405 ms	1394 w	ν(C2-N3) + ν(C-N10) + ν(ring)
1195.6	8	2	1168.6	*1245*, 1240, **1237**, 1170	1342 s	1342 m	δ(N1-H) + δ(C-H)
1092.7	3	1	1069.9	1127, 1126, **1116**, *1086*	1295 s	1294 m	$r(NH_2)$ + ν(ring)
1060.9	8	3	1039.3	*1067*, 1063, **1056**, 1050	1227 s	1220 s	ν(ring) + ν(C5=C6)
932.6	1	0	916.2	**937**, *935*, 934, 908	1128 vs	1130 m	γ(C6-H)

916.9	1	1	901.1	***965***, 958, 939, 922	1055 vs	1060 w	δ(ring) + r(NH_2)
780.4	4	0	770.1	763, ***755***, 754, 748	1034 vs	1028 vs	γ(C=O) + γ(ring)
771.9	0	15	762	***788***, 785, 780, 774	890 ms	892 ms	ν(ring) + ν(N1-C) + δ(NH_2)
750.7	0	0	741.6	**725**, *724*, 723, 722	972 w	962 ms	γ(C4) + γ(C=O) + γ(ring)
658.5	3	2	653.2	*692*, 685, 676, **666**	740 s	748 m	ν(C-C*l*) + ν(ring)
612.3	9	0	608.8	989, 986, 981, ***615***	785 w	785 m	γ(N1-H)
583.4	1	5	581.1	608, *604*, 598, **596**	652 ms	654 vs	δ(C=O) + δ(ring) + r(NH_2)
562	0	0	560.6	**834**, 825, 810, *565*	982 w	618 m	τ(NH_2)
542	0	2	541.4	592, 576, 559, ***551***	620 w	588 ms	δ(ring) + δ(C=O) + r(NH_2)
410.2	0	4	414.9	**467**, 453, 437, *431*	595 ms	540 ms	δ(C-C*l*, ring) + r(NH_2)
397.5	1	0	402.7	442, **441**, 438, *411*	820 w	438 m	γ(CCN) + γ(N1-H) + τ(NH_2)
352.4	0	2	359.4	*391*, 387, 383, **375**	525 ms	390 vw	δ_{as}(NC4CC*l*)+δ(C=O)+δ(ring)+r(NH_2)
267.1	0	0	277.5	310, **308**, 299, *295*	435 s	295 ms	γ(C-C*l*) + γ(ring) + τ(NH_2)
240.9	1	1	252.4	**285**, 283, *277*, 267	395 m	250 w	δ_s(N-C4-C-C*l*) + r(ring)
205.1	4	0	218	**243**, 238, *233*, 219	290 b	482 ms	γ(N3) + ω(NH_2) + γ(C=O)
123.5	20	0	139.7	***480***, 479, 473, 156	251 ms		ω(NH_2)
80.8	0	0	98.7	***153***, 144, 130, 128	472 m		γ(ring) puckering on O, C*l*

[a]With the equation at the B3LYP/6-31G(d,p) level: $\nu^{scal} = 34.6 + 0.9447\ \nu^{cal.}$ [b]Ref. [20].

The absence of a ν(OH) band in the 3,500–3,700 cm^{-1} range and the appearance of ν(C=O) modes as strong bands indicate that in the solid state this compound is only in the *keto* form. Another feature is that the majority of the fundamental modes absorb in relatively narrow wavenumber ranges. We have changed the assignments of several bands reported by other authors [20] according to our scaled values. A good match between the scaled wavenumbers in the tetramer to the experimental ones was found, and much better than that with the monomer form. That is the simulation of the crystal unit cell is appropriate for an accurate assignment of the experimental bands in the solid state.

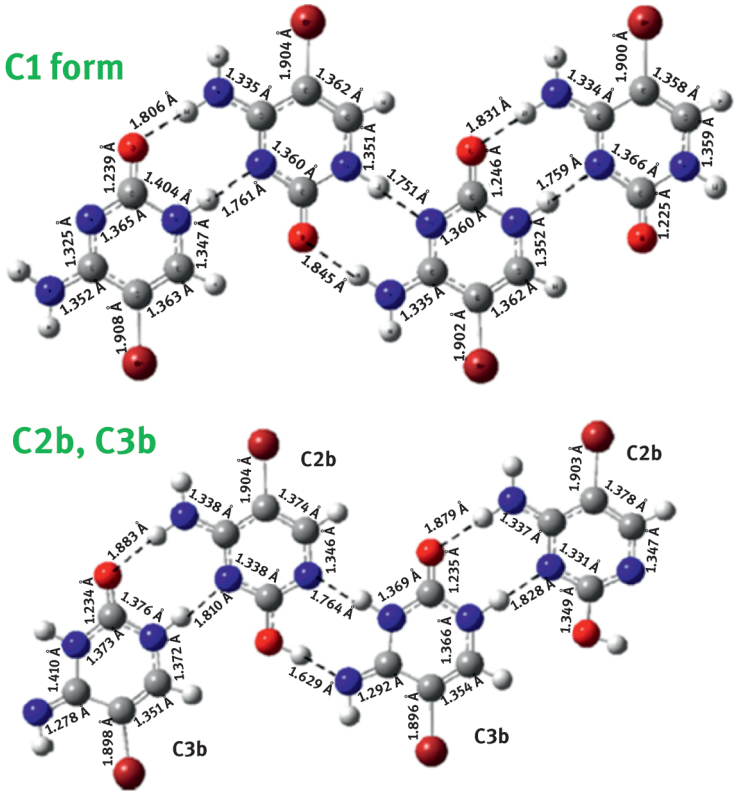

Figure 9: Simulation by a tetramer form of two possible crystal unit cells of 5-bromocytosine with three tautomeric forms.

5.4.2 5-Bromocytosine

X-ray data of 5-BrCy has not yet been reported, and thus the tautomer present in the solid state is not known, but it is possible to resolve this problem by simulation of the crystal unit cell of all the possible tautomers using

as reference the structure reported in related molecules, which in the present case is 5-ClCy. The crystal data of this unit cell (the tetramer form) was the starting point used for the simulation of the solid state (**Figure 9**).

In the isolated,state the molecule of 5-BrCy is calculated to be full planar, and the small deviations from planarity observed in crystals of Cy [53] are due to the intermolecular H-bonds. Our simulations in the tetramer form also show these deviations, in general with values slightly higher than in the crystal. Due to the orientation of the bromine atom in the tetramer form, and its almost zero interaction with other atoms, its bond length is almost the same as that in the isolated state. The simulations were carried out at the B3LYP/6-31G(d,p) level [54]. The tetramer form with tautomer C1 is 30.5 kJ/mol more stable than with tautomers C2b-C3b. The simulated spectra of these tetramer forms appear closer to the experimental spectrum in the solid state than the simulated spectrum of the monomer form. As an example of this concordance, the experimental (IR) and the scaled spectrum of this molecule are compared in Figure 10, for simplicity only in the dimer form and in the 3,700–2,700 cm^{-1} range. Similar agreement is observed in the other regions, as well as in the Raman spectrum.

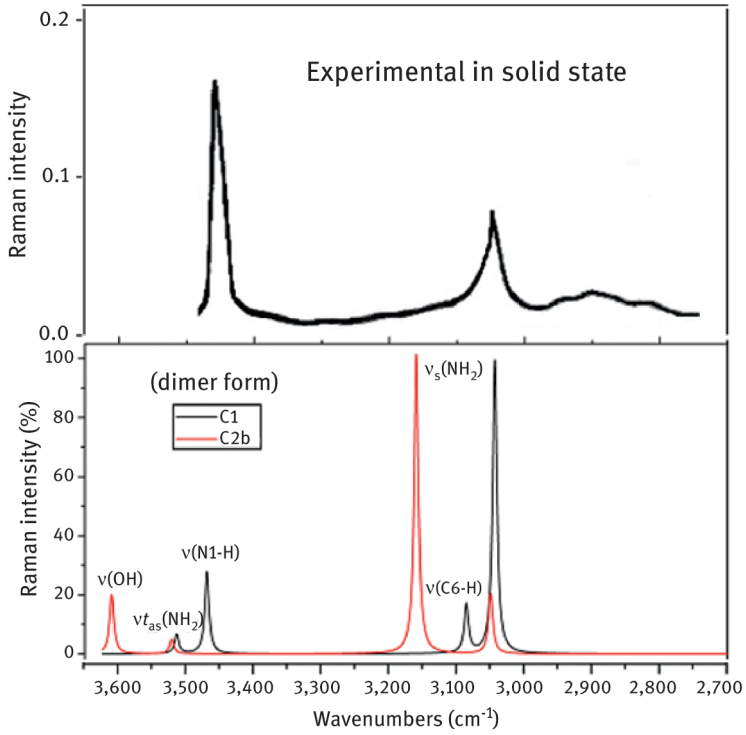

Figure 10: Experimental and theoretical scaled Raman spectra in the dimer form of 5-bromocytosine in the 3,650–2,700 cm^{-1} range and in the C1 and C2b tautomers using the scale equation $\nu^{scaled} = 34.6 + 0.9447\ \nu^{calculated}$

The scaled values of the ring vibrations in two of the tetramer forms are collected in Table 8. Four wavenumbers appear for each vibration corresponding to the four 5-BrCy molecules of the tetramer. Of these four wavenumbers, that with the highest IR intensity is shown in bold type, while that with the highest Raman intensity is shown in italic type. These scaled tetramer values can be directly compared to the experimental IR and Raman data reported in the solid state [20], [54]. We have underlined the experimental bands selected for this comparison. The absolute error (Δ) between the observed and the scaled wavenumbers is lower in tetramer C1 than in C2b-C3b. I.e. The spectrum of crystalline solid 5-BrCy shows that this molecule exists in the crystal in only the amino-oxo form (C1).

Table 8: Comparison of the scaled harmonic wavenumbers (ν^{scal}, cm^{-1}) obtained in the tetramer forms of the ring of 5-BrCy with the experimental IR and Raman values, and the absolute error $\Delta(\nu^{scal}-\nu^{exp})$ (in cm^{-1}) obtained. The highest calculated IR intensity is shown in **bold** type while the highest Raman intensity is printed in *italic* type.

Tetramer (C1)		**Tetramer (C2b and C3b)**		**Experimental in the solid state**	
ν^{scal}	**Δ**	ν^{scal}	**Δ**	**IR[a]**	**IR[b]**
3,466, 2,870, 2,859, ***2,818***	–82	3,459, *2,969*, 2,948, **2,816**	–84	3,210 w	2,900 b,vs
3,089, **3,085**, 3,084, 3,082	15	3,095, 3,093, ***3,054***, 3,052	–16	3,095 w	3,070 vs
1,737, **1,705**, 1,684, *1,665*	–5	***1,730***, 1,726	20	1,726 w	1,710 vs

1,633, *1,606*, 1,605, 1,604	23	1,662, ***1,656***, 1,617, 1,612	46	1,625 w, 1,610 m	
1,488, 1,448, *1,443*, 1,394	16	***1,593***, 1,546, 1,493, 1,484	121	1,510 w	1,510 m
1,497, *1,496*, **1,456**, 1,452	–54	1,539, *1,465*, 1,445, **1,437**	–73	1,472 vs, 1,455 w	1,433 s
1,548, 1,539, *1,529*, 1,493	–22	*1,532*, 1,528, **1,522**, 1,361	–48	1,570 ms	1,556 m
1,320, **1,303**, 1,302, *1,301*	–39	*1,334*, **1,330**, 1,303, 1,278	–12	1,342 vs	1,337 m
1,284, 1,284, **1,266**, *1,248*	–19	1,421, **1,352**, *1,311*, 1,280	67	1,285 ms	1,285 s
1,270, *1,243*, **1,239**, 1,169	9	1,301, *1,288*, **1,231**, 1,216	1	1,230 m	1,235 s
1,044, 1,041, **1,034**, 1,030	19	1,050, ***1,046***, 1,042, 1,021	31	1,065 ms, 1,015 s	1,066 s, 1,004 m
992, 946, ***943***, 920	–12	976, ***959***, 946, 943	4	975 ms, 955 ms	943 w
966, 957, 939, 928	53	***1,001***, 998, 979, 965	88	900 w	913 m
766, ***760***, 758, 750	15	**783**, 780, 738, *736*	38		745 m
787, 784, 780, 773	9	792, *786*, 783, **775**	–3	776 ms	778 m
740, 737, *737*, **735**	10	***730***, 729, 724, 723	5	725 s	725 w
665, 659, 649, 641	19	673, 660, *647*, **633**	–13	632 m	646 vs
995, 988, *942*, **615**	15	964, **897**, 893, *655*	297		600 s
604, 598, *596*, 586	24	616, 608, 601, ***598***	18	580 m	
590, 573, 557, *551*	45	589, 567, 555, ***549***	4	545 ms	
451, 448, 444, ***416***	–3	*471*, **466**, 432, 419	47	430 ms	440 s, 419 w
458, 445, *430*, 415	3	**457**, 441, 422, *410*	2	495 w	470 m, 455 w
313, 309, **308**, 304	28	***316***, 309, 304, 303	36	380 w, 280 b	-
303, **301**, 292, 289		**328**, *324*, 308, 305		-	-
244, *238*, 232, 219	–2	*247*, **240**, 201, 182	0	-	-
258, 255, *246*, 238	41	**278**, 257, *245*, 235	73	-	-
150, 142, **130**, *123*		***147***, 138, 130, 127		-	-

[a] Ref [20]. [b] Ref [54].

The preference for only one tautomeric form in the crystal than in the isolated state is a clear indication of the importance of intermolecular interactions, in particular H-bonding, to determine the structure of the condensed phase. Because molecules in the crystal are involved in hydrogen bonding and in other intermolecular interactions, it is expected that these H-bonds significantly change the spectrum obtained in the solid state compared to that in the gas phase (or in Ar/Ne matrices). The bands corresponding to the stretching and bending vibrations of the N1-H, C=O and NH_2 groups change more drastically than those related to the ring or to the CH group. Thus, the calculations of the tetramer form help to interpret the spectra in the crystalline solid.

6 Summary and conclusions

i. Several examples of the applicability of the computational methods to analysing the vibrational spectra of cytosine, uracil, 5-chlorocytosine and 5-bromocytosine nucleic acid bases are shown. DFT quantum chemical methods were selected as the most appropriated, in particular B3LYP, which was the main method used here.

ii. The B3LYP/DFT method was used to characterize the ring normal vibration modes of uracil and cytosine molecules through the atomic displacement vectors involved in each calculated vibrational frequency. This application of the computational methods permits a clear identification and assignment of all the bands observed experimentally.

iii. An accurate simulation of the vibrational spectra of the molecules under study was performed using DFT methods. These theoretical/scaled spectra can be compared to the experimental IR and Raman spectra, and therefore all the bands can be assigned.

iv. To correct the systematic error of the theoretical methods, accurate scaling procedures are presented and used in the calculated spectra of the molecules under study. The use of the LSE scaling equation procedure

significantly reduced the error in the calculated wavenumbers and it is the best for this purpose [55]. Thus, we recommend its use instead of the single overall factor procedure utilized in the literature.

v. With the help of DFT calculations, vibrational spectroscopy was used as a tool for the identification of the tautomers present in the isolated state. As an example, the study carried out on the 5-bromocytosine molecule is described. In the isolated state, at least the C1 and C2b tautomers are presented. The wavenumbers corresponding to these tautomers were identified and assigned in the IR experimental spectrum reported in an Ar matrix.

vi. Another example of this utility in the identification of tautomers is shown in the cytosine molecule. The scaled wavenumbers of the different tautomers were compared to the experimental values reported in Ar/Ne matrices. However, in this case it was not possible to identify the characteristic bands corresponding to *enol* or *imino* tautomers, Thus, only the *keto* form appears clearly in the spectrum of this phase, in contrast to that reported using other techniques.

vii. Computational methods were also used to simulate the crystal unit cell of 5-chlorocytosine and 5-bromocytosine nucleobases, with the consequent interpretation of their vibrational spectra. With this simulation, a remarkable improvement was reached in the accuracy for the assignment of their spectra in the solid state; in particular in the vibrations involved in intermolecular H-bonds. In 5-bromocytosine, two tetramer forms were calculated. The scaled wavenumbers of these tetramer forms appeared in accordance with the experimental IR and Raman data in the solid state, and thus all the normal modes were identified and discussed. Vibrational spectroscopy indicated that in the solid state the biomolecule 5-bromocytosine exists only in the amino-oxo form.

viii. The possibility of tautomerization of cytosine is calculated to be much more likely than that of 5-bromocytosine. The effect of the Br atom is to prevent tautomerization from taking place, and it is higher in 5-bromocytosine than in 5-bromouracil. All the tautomers in the cytosine molecule appear with much lower relative energies than their counterparts in the uracil molecule.

Acknowledgements

MAP wish to thank to BSCH-UCM PR26/16 for financial support

This article is also available in: Ramasami, Computational Sciences. De Gruyter (2017), isbn 978-3-11-046536-5.

References

[1] Alcolea Palafox M. The prediction of vibrational spectra: The use of scale factors. Recent Res. Devel. in Physical Chem. Transworld Research Network: Trivandruw, India. Vol. 2 1998:213–232.

[2] Thirunarayanan S, Arjunan V, Marchewka MK, Mohan S, Atalay Y. Characterisation of 1,3-diammonium propylselenate monohydrate by XRD, FT-IR, FT-Raman, DSC and DFT studies. J Molec Struct. 2016;1107:220–230.

[3] Farasat M, Reza Shojaei SH, Maqsood Golzan MM, Farhadi K. Theoretical study of the potential energy surface and electric dipole moment of aniline. J Molec Struct. 2016;1108:341–346.

[4] Szafran M, Komasa A, Anioła M, Katrusiak A, Dega-Szafran Z. Structure of the complex of dimethylphenyl betaine with dichloroacetic acid studied by X-ray diffraction, DFT calculations, infrared and Raman spectra. Vibrat Spectrosc. 2016;84:92–100.

[5] Agrawal M, Deval V, Gupta A, Sangala BR, Prabhu SS. Evaluation of structure-reactivity descriptors and biological activity spectra of 4-(6-methoxy-2-naphthyl)-2-butanone using spectroscopic techniques. Spectrochim Acta A. 2016;167:142–156.

[6] Pan H, Wang W, Ma Z, Xu L, Zhang Z, Zhao Y, et al. Structures and spectroscopic properties of three [RuCl(2mqn)$_2$NO] (H2mqn = 2-methyl-8-quinolinol) isomers: an experimental and density functional theoretical study. Polyhedron. 2016;118:61–69.

[7] Watson JD, Crick FHC. Molecular structure of nucleic acids – A structure for deoxyribose nucleic acid. Nature. 1953;171(4356):737–738.

[8] Alcolea Palafox M, Iza N, Gil M. The hydration effect on the uracil frequencies: an experimental and quantum chemical study. J Molec Struct (Theochem). 2002;585(1-3):69–92.

[9] Alcolea Palafox M. Scaling factors for the prediction of the frequencies of the ring modes in benzene derivatives. J Phys Chem A. 1999;103:11366–11377.

[10] Kawai T, Ikegami M, Kawai K, Majima T, Nishimura Y, Arai T. Recognition of substituted cytosine derivatives by the base pairing with guanine connected to pyrene. Chem Phys Lett. 2005;407(1-3):58–62.

[11] Steenken S, Reynisson J. DFT calculations on the deprotonation site of the one-electron oxidised guanine–cytosine base pair. Phys Chem Chem Phys. 2010;12(31):9088–9093.

[12] Tripathi J, Tripathi PK, Singh L, Soni RK, Alcolea Palafox M, Rastogi VK. Tautomeric and electronic properties of biomolecule 2-thiocytosine. Int J Curr Chem. 2010;1(2):119–132.

[13] Alcolea Palafox M, Rathor SK, Rastogi R, Bhat D, Jothy BV, Rastogi VK. Asian J Phys. 2015;24(1):33–59.
[14] Alcolea Palafox M, Rastogi VK. The biomolecule of 5-chlorocytosine: geometry of the six main tautomers in the isolated state by DFT calculations, and interpretation of the FT-IR and FT-Raman spectra in the solid state. Asian Chem Lett. 2015;19(1):1–25.
[15] Raczyńska ED, Sapula M, Zientara-Rytter K, Kolczyńska K, Stępniewski TM, Hallmann M. DFT studies on the favored and rare tautomers of neutral and redox cytosine. Struct Chem. 2016;27:133–143.
[16] Yang ZB, Rodgers MT. Theoretical studies of the unimolecular and bimolecular tautomerization of cytosine. Phys Chem Chem Phys. 2004;6:2749–2757.
[17] Morris SM. The genetic toxicology of 5-fluoropyrimidines and 5-chlorouracil. Mutat Res. 1993;297(1):39–51.
[18] Kang JI, Burdzy A, Liu PF, Sowers LC. Synthesis and characterization of oligonucleotides containing 5-chlorocytosine. Chem Res Toxicol. 2004;17:1236.
[19] Krishnakumar V, Balachandran V. Normal coordinate analysis of 5-fluoro, 5-chloro and 5-bromo-cytosines. Ind J Pure Appl Phys. 2001;39(10):623–627.
[20] Krishnakumar V, Balachandran V. Analysis of vibrational spectra of 5-fluoro, 5-chloro and 5-bromo-cytosines based on density functional theory calculations. Spectrochim Acta A. 2005;61(5):1001–1006.
[21] Piacenza M, Grimme S. Systematic quantum chemical study of DNA-base tautomers. J Comput Chem. 2004;25(1):83–98.
[22] Fogarasi G. Relative stabilities of three low-energy tautomers of cytosine: A coupled cluster electron correlation study. J Phys Chem A. 2002;106:1381–1390.
[23] Alparone A, Millefiori A, Millefiori S. Non-planarity and solvent effects on structural and polarizability properties of cytosine tautomers. Chem Phys. 2005;312(1-3):261–274.
[24] Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman HM. Geometric parameters in nucleic acids: nitrogenous bases. J Am Chem Soc. 1996;118(3):509–518.
[25] Seminario JM, Politzer P. Modern density functional theory: a tool for chemistry Vol. 2. Amsterdam: Elsevier, 1995.
[26] March NH. Electron density theory of atoms and molecules. London: Acad. Press, 1991.
[27] Alcolea Palafox M, Rastogi VK. Quantum chemical predictions of the vibrational spectra of polyatomic molecules. The uracil molecule and two derivatives. Spectrochim Acta A. 2002;58(3):411–440.
[28] Becke AD. Density-functional thermochemistry. 3. The role of exact exchange. J Chem Phys. 1993;98(7):5648–5652.
[29] Lee C, Yang W, Parr RG. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron-density. Phys Rev. 1988;B37(2):785–789.
[30] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 09, Revision D.01. Wallingford CT: Gaussian, Inc., 2009.
[31] Scott AP, Radom L. Harmonic vibrational frequencies: An evaluation of Hartree—Fock, Møller—Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors. J Phys Chem. 1996;100(41):16502–16513.
[32] Pulay P, Saebo S, Malagoli M, Baker J. Accuracy and efficiency of atomic basis set methods versus plane wave calculations with ultrasoft pseudopotentials for DNA base molecules. J Comput Chem. 2005;26:599–605.
[33] Radchenko ED, Plokhitnichenko AM, Ivanov AY, Sheina GG, Blagoy YP. Ketoenole tautomeria of guanine and isocytosine molecules. Biofizika. 1986;31(3):373–381.
[34] Nowak MJ, Lapinski L, Fulara J. Matrix isolation studies of cytosine: the separation of the infrared spectra of cytosine tautomers. Spectrochim Acta A. 1989;45(2):229–242.
[35] Szczesniak M, Szczepaniak K, Kwiatkowski JS, KuBulat K, Person WB. Matrix isolation infrared studies of nucleic acid constituents. 5. Experimental matrix-isolation and theoretical ab initio SCF molecular orbital studies of the infrared spectra of cytosine monomers. J Am Chem Soc. 1988;110(25):8319–8330.
[36] Radchenko ED, Sheina GG, Smorygo NA, Blagoi YP. Experimental and theoretical studies of molecular structure features of cytosine. J Mol Struct. 1984;116(3-4):387–396.
[37] Aamouche A, Ghomi M, Coulombeau C, Jobic H, Grajcar L, Baron MH, et al. Neutron inelastic scattering, optical spectroscopies and scaled quantum mechanical force fields for analyzing the vibrational dynamics of pyrimidine nucleic acid bases. 1. Uracil. J Phys Chem. 1996;100(13):5224–5234.
[38] Szczepaniak K, Person WB, Leszczynski J, Kwiatkowski JS. Matrix isolation and DFT quantum mechanical studies of vibrational spectra of uracil and its methylated derivatives. Polish J Chem. 1998;72(2):402–420.
[39] Chandra AK, Nguyen MT, Zeegers-Huyskens T. Theoretical study of the interaction between thymine and water. Protonation and deprotonation enthalpies and comparison with uracil. J Phys Chem A. 1998;102:6010–6016.
[40] Harsányi L, Császár P, Császár A, Boggs JE. Interpretation of the vibrational spectra of matrix-isolated uracil from scaled ab initio quantum mechanical force fields. Int J Quantum Chem. 1986;29(4):799–815.
[41] Ivanov AY, Plokhotnichenko AM, Radchenko ED, Sheina GG, Blagoi YP. FTIR spectroscopy of uracil derivatives isolated in Kr, Ar and Ne matrices: matrix effect and fermi resonance. J Mol Struct. 1995;372(2-3):91–100.
[42] Colarusso P, Zhang K, Guo B, Bernath PF. The infrared spectra of uracil, thymine, and adenine in the gas phase. Chem Phys Lett. 1997;269(1-3):39–48.
[43] Lés A, Adamowicz L, Nowak MJ, Lapinski L. The infrared spectra of matrix isolated uracil and thymine: an assignment based on new theoretical calculations. Spectrochim Acta A. 1992;48(10):1385–1395.
[44] Császár P, Harsányi L, Boggs JE. Vibrational frequencies and assignments for some isotopomers of uracil using a scaled ab initio force field. Int J Quantum Chem. 1988;33(1):1–17.
[45] Rastogi VK, Singh C, Jain V, Alcolea Palafox M. FTIR and FT-Raman spectra of 5-methyluracil (thymine). J Raman Spectrosc. 2000;31(11):1005–1012.
[46] Rastogi VK, Jain V, Yadav RA, Singh C, Alcolea Palafox M. Fourier transform Raman spectrum and ab initio and density functional computations of the vibrational spectrum, molecular geometry, atomic charges and some molecular properties of the anticarcinogenic drug 5-fluorouracil. J Raman Spectrosc. 2000;31(7):595–603.

[47] Ferro D, Bencivenni L, Teghil R, Mastromarino R. Vapour pressures and sublimation enthalpies of thymine and cytosine. Thermochim Acta. 1980;42(1):75–83.
[48] Florian J, Baumruk V, Leszczynski J. IR and Raman spectra, tautomeric stabilities, and scaled quantum mechanical force fields of protonated cytosine. J Phys Chem. 1996;100(13):5578–5589.
[49] Aamouche A, Ghomi M, Grajcar L, Baron MH, Romain F, Baumruk V, et al. Neutron inelastic scattering, optical spectroscopies and scaled quantum mechanical force fields for analyzing the vibrational dynamics of pyrimidine nucleic acid bases: 3. Cytosine. J Phys Chem A. 1997;101:10063–10074.
[50] Ten GN, Baranov VI. Analysis of electronic-vibrational spectra of uracil, thymine, and cytosine. Opt Spectrosc. 2004;97(2):195–203.
[51] Gould IR, Vincent MA, Hillier I, Lapinski L, Nowak MJ. A new theoretical prediction of the infrared spectra of cytosine tautomers. Spectrochim Acta A. 1992;48(6):811–818.
[52] Jaworski A, Szczesniak M, Szczepaniak K, Kubulat K, Person WB. Infrared spectra and tautomerism of 5-fluorocytosine, 5-bromocytosine and 5-iodocytosine. Matrix isolation and theoretical AB initio studies. J Mol Struct. 1990;223:63–92.
[53] Barker DL, Marsh RE. Crystal structure of cytosine. Acta Crystallogr. 1964;17(12):1581.
[54] Alcolea Palafox M, Rastogi VK, Kumar S, Joe H. The biomolecule of 5-bromocytosine: FT-IR and FT-Raman spectra and DFT calculations. Identification of the tautomers in the isolated state and simulation the spectra in the solid state. Spectrochim Acta A. 2013;111:104–122.
[55] Alcolea Palafox M. Scaling factors for the prediction of vibrational spectra. I. Benzene molecule. Int J Quantum Chem. 2000;77:661–684.