JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

The Absorption Spectrum of Cytosine Tautomers: Beyond the Static Approach

Alex Domingo,*[,†] Antonio Rodríguez-Fortea,[†] and Coen de Graaf*[,†,‡]

[†]Departament de Química Física i Inorgànica, Universitat Rovira i Virgili, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

[‡]Institució Catalana de Recerca i Estudis Avançats, Pg. Lluís Companys 23, 08010 Barcelona, Spain

ABSTRACT: The absorption spectrum of cytosine in water has been studied by combining Car−Parrinello molecular dynamics (MD) with a multiconfigurational perturbation theory treatment of the electronic structure. The MD simulations were performed for four different tautomeric forms of cytosine in a unit cell with 60 water molecules. The relative energies and transition dipole moments of a large number of excited states have been calculated on a representative sample of conformations along the MD trajectories. In this way, the broad experimental peaks can be decomposed, and the effect of the distortions on the nature of the excited states can be assessed. The loss of planarity of the molecule is significant, and hence, the excited states can no longer be defined as pure $n \rightarrow \pi^*$ or $\pi \rightarrow \pi^*$ excitations. We propose an analysis to assign the different transitions according to the main contribution. The keto N1H form turns out to be the most stable one, and the calculated spectra of this tautomer show good agreement with experimental measurements. The mixed $n\pi^*/\pi\pi^*$ character of some states leads to a significant increase of intensity in spectral regions dominated by the dark $n\pi^*$ transitions considering a planar structure.

1. INTRODUCTION

The sensibility of the DNA molecule to UV light absorption can lead to mutations of the genetic material and eventually carcinogenesis. However, the DNA molecule has effective mechanisms of deactivation to avoid genetic damage. The constituent nucleobases of DNA have a prominent role in this process because not only do all four bases (adenine, thymine, guanine, and cytosine) have large absorption coefficients in the UV range but also the monomers and base pairs show intrinsic molecular fast decays upon electronic excitation.[1,2]

Even though the four nucleobases are relatively simple molecules, the understanding of their electronic structure is complex. The DNA base cytosine is known to have multiple decay pathways that involve various excited states. The UV radiation of cytosine produces primarily dipole allowed transitions to singlet $\pi\pi^*$ states, the so-called bright states, and dipole forbidden transitions to singlet $n\pi^*$ states at a lower rate, the so-called dark states. Theoretical and experimental research on the potential energy surfaces (PES) of the ground state and the lowest excited states of cytosine have found many plausible radiationless deactivation mechanisms for this chromophore. Nowadays, the proposed decay pathways for cytosine include multiple internal conversion via conical intersections (CI) at different regions of the hyper-PES between various $^1\pi\pi^*$ and $^1n\pi^*$ states and the ground state.[3−7] A relaxation mechanism through an intersystem crossing (ISC) to the lowest $^3\pi\pi^*$ state has been also suggested.[8] Furthermore, some decay paths involving conformational distortions[9] or excited tautomerization[10] pointed out the significant role of the structural flexibility of cytosine. These results prove the large versatility and high efficiency of the cytosine molecule to decay back to the ground state but, additionally, the high complexity of its electronic structure.

At first sight, the absorption spectra of cytosine are relatively simple. They show basically two intense and broad absorption bands with some additional minor features depending on the experimental conditions.[11−14] A brief summary of available spectroscopic data for cytosine is shown in Table 1. The main bands of the absorption spectra of cytosine correspond to transitions I (∼ 4.5 eV) and IV (∼ 6.0 eV), which are transitions in the molecular plane assigned to $\pi\pi^*$ excitations.[15] Transitions II and III (5−6 eV) are usually masked by the more intense $\pi\pi^*$ bands, with their characterization by means of experimental polarization data and transition moments becoming difficult.[16−18]

The theoretical research of the electronic structure has put some light into these weak absorptions. Early calculations of the cytosine molecule using an all valence electrons self consistent field molecular orbital with configuration interaction (SCF-MO−CI) resulted in detailed information about the characteristics of both the $\pi\pi^*$[20] and $n\pi^*$[21] transitions, identifying transition II at 5.2 eV as an $n\pi^*$ excitation from the nonbonding orbital of the carbonyl group. Moreover, Matos and Roos used the complete active space self consistent field (CASSCF) method to treat the complete π electronic system of cytosine. They found that the lone pairs of the O and N atoms have a large effect on the electron correlation of the molecule and calculated the three lowest $\pi\pi^*$ transition energies to be 5.6, 6.9, and 8.1 eV.[17] Subsequent calculations by Fülscher and Roos included $\sigma-\pi$ polarization through the second-order perturbation theory using a CASSCF reference wave function (CASPT2) to treat the dynamical correlation of σ electrons. They obtained more accurate results with this method and calculated the transition energies of four $\pi\pi^*$ states at 4.4, 5.4, 6.2, and 6.7 eV, a $n\pi^*$ transition as the second excited state at 5.0 eV, and a possible higher $n\pi^*$ transition at 6.5 eV.[22] Petke et al. performed multi-reference CI calculations of cytosine that led to similar conclusions for the lowest transitions but pointed at the complexity of the higher energy region due to spectral congestion above 5.2 eV.[18]

Received: October 3, 2011
Published: December 05, 2011

Table 1. Summary of Spectroscopic Data for Cytosine

		transition energy (eV)					
spectrum medium	ref	I	II	III	IV	V	VI
trimethyl phosphate	11	4.5	5.2		6.1, 6.7		
vapor	12	4.3					
water (pH = 7)	13	4.7					
water (pH = 2)	13	4.5			5.9		
crystal monohydrate	14	4.6					
water[a]	16	4.7	5.3	5.6	6.2	7.7	8.1
ethanol	19	4.5	5.2		6.0		

[a] Transitions I−IV adjusted from crystal polarized reflection spectra. Transitions V−VI estimated from thin film spectrum.

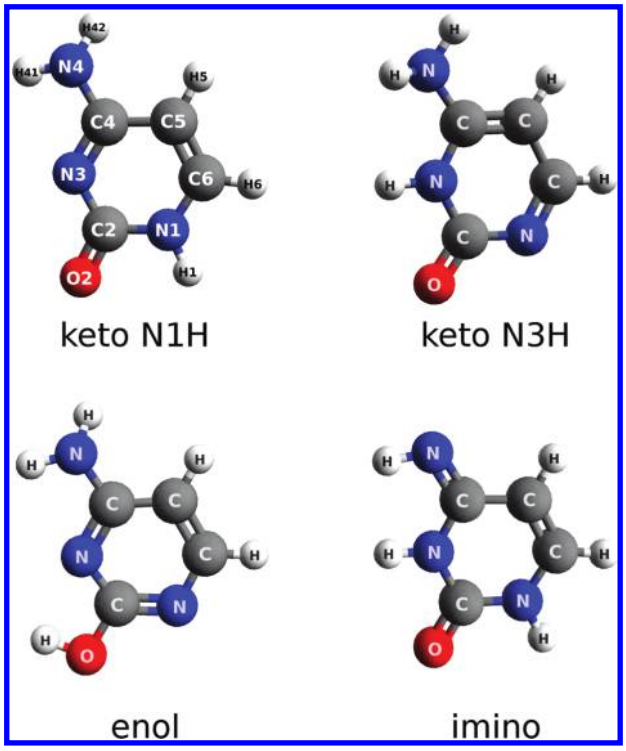

Figure 1. Structures of four tautomeric forms of cytosine.

The influence of different solvents to the electronic structure of cytosine has been another important subject of research. The absorption spectrum of cytosine depends on the dielectric constant of the solvent (Table 1) and, in the case of water, on the pH value.[13,23] Furthermore, cytosine has various tautomeric forms that can be classified on the basis of their functionalization in three basic structures: keto, enol, and imino (Figure 1). The population of these tautomers varies depending on the characteristics of the solvent, and additionally, they have quite different electronic properties, complicating even more the interpretation of spectroscopic data.[24] Even though various tautomeric structures can be present for a given solvating condition, the enol form is the most stable in apolar solvents and the gas phase, while the keto form is the most stable in polar solvents.[19,25−29]

There has been much effort to incorporate aqueous solvent effects into the theoretical models of cytosine and other organic molecules to achieve biologically relevant data.[30] To simulate the electrostatic effect of the solvent in the computational model, one commonly includes some kind of continuum reaction-field around the quantum system.[31−35] Additionally, some explicit water molecules can be included in the electronic structure calculation to achieve a more accurate description of the solute−solvent interactions.[36−39] The main known effect of aqueous solvent on the electronic structure of cytosine is a blue shift of the $n-\pi^*$ transitions due to stabilization of the electron lone pairs, which pushes the lower dark excitations to the spectral congestion region.

The cytosine tautomer found in the Watson and Crick structure of DNA is the keto form.[40] However, the large capability of the nucleobases to transfer protons and the charge transfer phenomena through the nucleobases stacks in the DNA chain[41,42] can lead to tautomerization of the DNA bases. It has been suggested that the tautomerization of cytosine occurs primarily through ionic structures[25] and that photoinduced tautomerization could offer extra pathways for deactivation.[10,36] Moreover, work on other biologic systems like urocanic acid showed the role that different tautomers can have on the absorption spectrum.[43] Thus, it becomes important to consider not only the keto form of cytosine but also its other tautomeric structures to understand the electronic properties of cytosine related biologic systems.

The main drawback of the theoretical models described so far is that they are limited to a static description of the system. Although it is possible to treat different tautomers and include solvent effects in the calculation, only ideal structures of the molecule have been studied so far. However, cytosine is known to be very flexible and easily loses its planarity. Molecular dynamics simulations using the Car−Parrinello scheme explored the conformational space of the DNA bases showing a wide range of accessible distortions for this molecule.[44] Thus, once a cytosine molecule deviates from planarity, it is not strictly possible to talk about π and n states and $\pi\pi^*$ and $n\pi^*$ transitions, because σ, π, and n MOs start to mix, and the interpretation of the single contributions to the absorption spectra becomes difficult to perform. Nevertheless, the inclusion of the vibrorotational degrees of freedom on the theoretical study of these compounds is required to achieve a good understanding of their electronic structure. Nowadays, it is becoming a rule to combine molecular dynamics simulations with QM/MM calculations to study biologic systems,[45−49] and the cytosine molecule is no exception.[6,7,50,51] Hence, this theoretical framework incorporates the conformational evolution of the system at a given temperature plus the effect of the solvent and sophisticated electronic structure calculations.

The aim of the present work is to continue developing the *ab initio* theoretical description of the cytosine molecule by combining Car−Parrinello molecular dynamics (Car−Parrinello MD)[52] with the complete active space self-consistent field second-order perturbation theory (CASPT2).[53,54] We study the conformational space of four tautomeric forms of cytosine in water (Figure 1) by means of Car−Parrinello MD simulations. Subsequently, on a representative sample of the resulting trajectories, we perform CASPT2 calculations to obtain an accurate description of the electronic structure of each conformation. The solvent effects are incorporated on the CASPT2 step through two solvent models, the polarizable continuum model (PCM)[55] and a modification of PCM including some explicit water molecules. The resulting data set for each tautomer is combined to generate its absorption spectrum in water. We offer a detailed

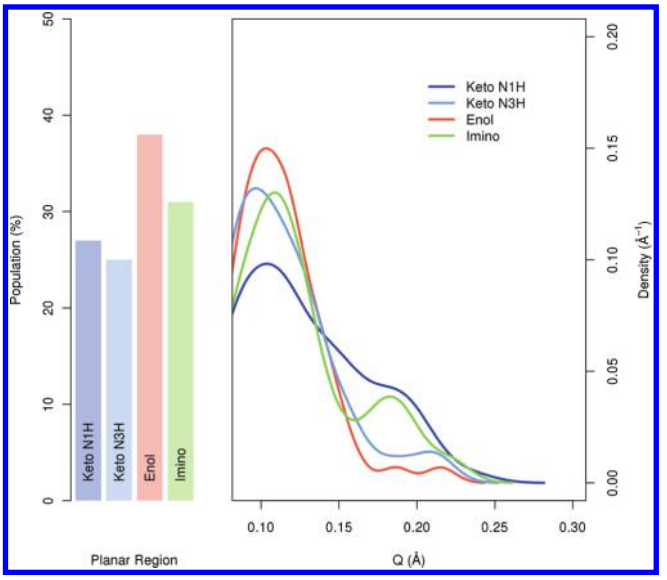

Figure 2. Degree of puckering of the cytosine aromatic ring for each tautomeric form. Left: Proportion of planar structures. Right: Value of the Q parameter for nonplanar structures.

view of the underlying structure of the absorption spectrum of cytosine, comparing the two solvent models and analyzing the large mixing observed between the different transition types.

2. COMPUTATIONAL SCHEME

In contrast to classical molecular dynamics simulations, first-principles (also called *ab initio*) molecular dynamics methods compute the forces acting on the nuclei from electronic structure calculations (typically using the density functional theory) that are performed as the molecular dynamics trajectory is generated ("on the fly"). Consequently, those systems with electronic structures that change significantly (i.e., formation/breaking of bonds) during the dynamics can be easily handled by molecular dynamics. The method developed by Car and Parrinello[52] uses fictitious dynamics to develop the electronic orbital functions, which are only minimized at the beginning of the dynamics, preventing the need for a costly self-consistent iterative minimization at each time step. Details about the Car−Parrinello MD method and first-principles molecular dynamics in general can be found in ref 56 and references therein.

We performed Car−Parrinello MD simulations for each tautomeric form of cytosine considered in this work, namely the keto N1H, keto N3H, enol, and imino forms (Figure 1). All simulations were performed in a box with side lengths of 12.5487 Å containing 60 water molecules and one cytosine molecule. The molecular dynamics conditions were set according to the previous work by Isayev et al.[44] on cytosine in a vacuum. The core electrons are described by Troullier−Martins normconserving pseudopotentials.[57] The electronic potential was calculated by means of the BLYP density functional[58,59] using a plane waves basis set with a cutoff of 80 Ry. The system was first equilibrated at 300 K and then maintained at a steady temperature with a Nosé−Hoover chain thermostat.[60,61] All H atoms are deuterium isotopes, to avoid the high frequency hydrogen stretchings and be able to use a larger time step of 0.121 fs with a fictitious electron mass μ = 700 au to improve the computational cost.[62] For each tautomeric form, 100 conformational structures were extracted from a trajectory of 2 ps. The MD simulations were carried out with the CPMD software version 3.11.1.[63]

Table 2. Classification in the Six Symmetrical Distorted and Planar Forms of the 100 Conformational Structures of the Four Cytosine Tautomers

	relative proportion (%)			
symmetrical form	keto N1H	keto N3H	enol	imino
planar	27	25	38	31
chair	2	3	1	5
half-chair	13	5	7	12
envelope	14	17	13	14
screw-boat	18	19	12	12
boat	13	14	20	16
twist-boat	13	17	9	10

The 400 conformers selected from the Car−Parrinello MD simulations are the ones used in the electronic structure calculations of the excited states of cytosine. To ensure the sample representativeness of this set of structures, we analyzed the degree of puckering of their aromatic ring in terms of the Cremer−Pople (CP) parameters,[64] which offer a systematic method to measure the distortion of each conformer and classify them following Boeyen's scheme.[65] All structures are classified in symmetrical nonplanar conformations based on their CP parameters Q, θ, and ϕ. The Q parameter serves as a measure of ring planarity, the more planar structures being the ones with a smaller Q. The conformers with a weighted average torsion angle smaller than $\pm 5°$ are considered planar, which corresponds to Q values smaller than 0.1 Å. Figure 2 shows the relative population of planar structures and the pronounced loss of planarity of the four tautomeric forms of cytosine. Only three conformations out of 10 can be considered planar. The degree of distortion of the four tautomers is very similar, and all of them have significantly smaller Q values in water than in the gas phase.[44] Furthermore, the tautomeric forms with a protonated N3 ring atom show slightly larger average distortions. This behavior could be ascribed to a larger steric repulsion caused by the N3 hydrogen atom on the out-of-ring heteroatoms. The other two parameters θ and ϕ are used to identify the type of distortion of the conformer and assign them to one of the six symmetrical forms of a molecule with a six-membered ring. Table 2 analyzes the proportion of the symmetrical forms between the different tautomers of cytosine. The percentage of chair like forms is low for all tautomers compared with the planar and boat like forms, which are the most abundant. The presented analysis was performed using the PLATON package.[66]

The aqueous solvent effects in the electronic structure calculations were computed by means of a PCM to reproduce the electrostatic interaction of water with the cytosine molecule. The electronic response of the solvent upon absorption is partially defined by the ground state properties of the solute, the so-called slow component, and the excited final state, the so-called fast component. Moreover, we performed a second set of calculations for all of the structures including in the wave function some explicit water molecules surrounding the cytosine molecule, in addition to the PCM wrapping. Hence, the effect of hydrogen bonding between the solute and closer water molecules can be incorporated into the electronic structure. Those explicit water molecules come directly from the Car−Parrinello MD, maintaining their relative orientation

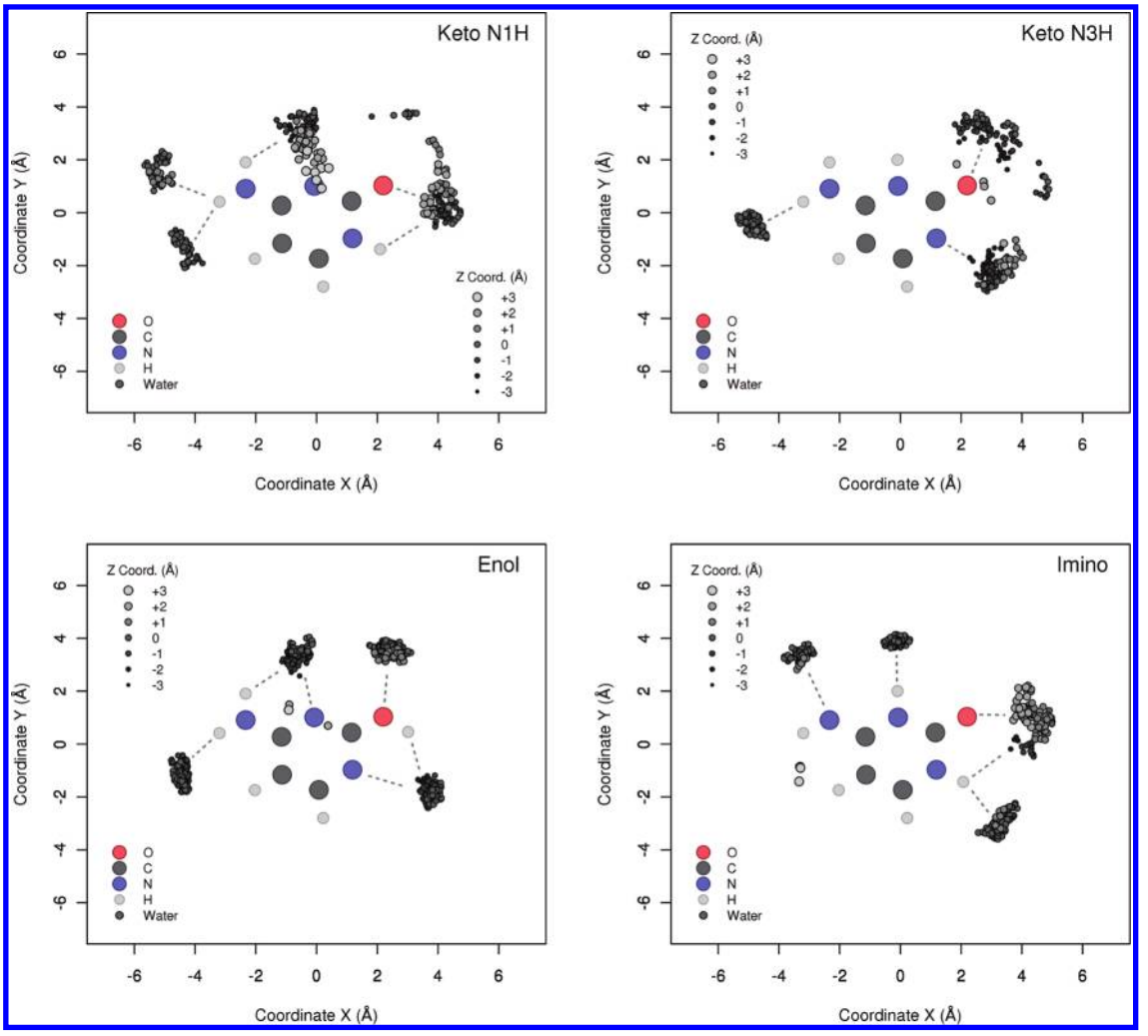

Figure 3. Relative orientation of the selected water molecules on each tautomeric form of cytosine: keto N1H (top left), keto N3H (top right), enol (bottom left), and imino (bottom right). Each graph is a superposition of the position on the moleclar plane of all water molecules chosen along the 100 conformations (dark dots). The H-bonding formed by the solvent molecules is qualitatively represented by dashed lines to the respective ideal tautomeric structure.

to the cytosine conformer and being selected on the basis of their distance to the electronegative atoms of the molecule.

The two keto tautomers have three water molecules added, whereas the enol and imino tautomers have four explicit water molecules. This choice is based on the number of available sites for H-bonding. We consider the interaction of water with the functional groups and the heteroatoms of the ring. Therefore, the keto tautomers have three main interaction sites with the solvent, the carbonyl and amine groups and one lone pair of a N atom on the ring. In the case of the enol and imino tautomers, we add an additional water to maintain the balance between the two N atoms of the ring. Figure 3 shows the relative orientation of the selected explicit water molecules on the cytosine tautomers. Even though each cytosine conformer has its closest water molecules particularly oriented, they are found primarily positioned near the lone pairs and the various functional groups, as expected. The superposition of the water molecules extracted for all conformers shows how the polarity of the tautomer affects the mobility of the first solvating shell. The water molecules are significantly more fixed along the MD simulation of the enol tautomer than for the less polar keto forms.

The scope of this work is restricted to the lowest 9 eV region of the absorption spectrum of cytosine, not only to cover the largely studied first and second intense broad bands but also to explore the higher energy features observed above 7 eV.[16] Thus, we performed CASPT2 calculations of the 12 lowest singlet excited states of each conformer of cytosine with the two solvent models proposed, totalling 800 electronic structure calculations. The active space (CAS) of these calculations is formed by 14 electrons in the eight π orbitals plus two nonbonding σ orbitals. This CAS allows the description of the $^1\pi\pi^*$ and $^1n\pi^*$ transitions that contribute most to the absorption spectra of cytosine with an accurate treatment of the correlation from the lone pair electrons[17] and the $\sigma-\pi$ polarization.[22] The basis set for the cytosine molecule is of the atomic natural orbitals (ANO) type including scalar relativistic effects[67] to obtain an optimum treatment of correlation and polarization. The C, N, and O atoms of cytosine have (4s,3p,2d,1f) contracted basis functions and the H atom a (2s,1p) contracted set. The contracted basis set employed for the explicit water molecules is (3s,2p,1d) for the O atom and (2s,1p) for the H atom. These basis sets are large enough to describe all considered valence states of cytosine. We performed the

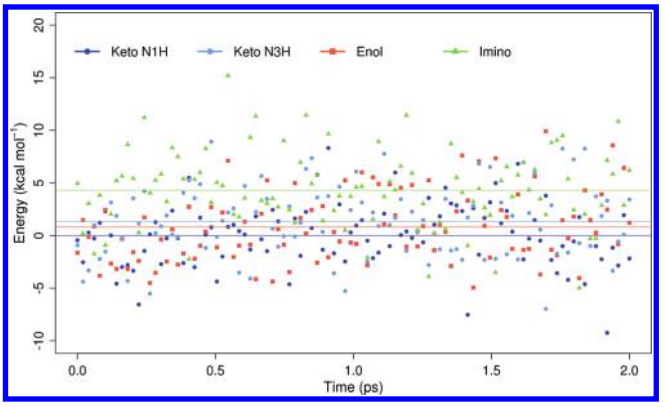

Figure 4. Ground state energy of the 400 conformers extracted from the Car−Parrinello MD simulations. Results from CASPT2 calculations using a PCM solvent model for water. The reference energy has been set to the average energy of the keto N1H tautomer. Horizontal lines represent the average energy of each tautomeric form.

CASPT2 calculations with a shift factor of 0.20 au,[68] which we previously tested to be the minimum value to eliminate possible intruder states. The intensity of the 11 lowest electronic transitions was computed by state interaction calculations between the ground state and all of the excited states at the CASPT2 level, inserting the transition dipole operator into the Hamiltonian.[69] The electronic structure calculations were carried out with the MOLCAS 7 package.[70,71] The resulting data from the 800 calculations were statistically treated with the R package[72] to generate the corresponding spectra by means of kernel density estimations with Gaussian kernel functions of 0.13 ± 0.01 eV of bandwidth (bw), depending on the characteristics of the data set.

Computer codes have been developed to automate the extraction of the target molecule(s) from the MD trajectories and to transform this information into a readable format for the electronic structure code. Moreover, some scripts were made to process the large amount of data generated in the computational procedure.

3. RESULTS

We obtained the absorption spectrum of the keto N1H, keto N3H, enol, and imino tautomers in water. For each tautomer, we show two spectra, one using a PCM solvent model and another with a PCM including some explicit water molecules. Both spectra are compared and analyzed in the next sections. Additionally, to check the molecular description obtained by the combination of molecular dynamics with CASPT2 calculations, we show the energetic evolution of each tautomer along the Car−Parrinello MD simulation. The energy values of spectroscopic properties are given in electronvolts, and the analysis of the relative energy between the four tautomers is done in kilocalories per mole to ease the comparison with previous studies.

3.1. Relative Stability of Cytosine Tautomers. Figure 4 shows the CASPT2 energies of the 400 conformers selected from the Car−Parrinello MD simulations of the considered cytosine tautomers. We compared the relative energy of the electronic ground state of all of the structures using a PCM solvent model of water. We did not use the solvent model including explicit water molecules, in order to make easier the analysis. A rigorous study of the relative stability of the cytosine tautomers in water is out of the scope of the present work, because it is much more complex than what we can obtain from the present calculations. Nonetheless, this straightforward comparison served to check that the description of the different tautomers offered by the combination of Car−Parrinello MD with CASPT2 is correct.

The tautomer with the lowest energy is the keto N1H, followed by the enol at 0.8 kcal mol^{-1} higher energy, the keto N3H at 1.3 kcal mol^{-1}, and the imino form being the most energetic at 4.3 kcal mol^{-1}. These results are in good qualitative agreement with previous theoretical studies about the relative stability of cytosine tautomers in water.[19,32,33,36] The most stable form in water is known to be the keto N1H, followed by the enol tautomer. The keto N3H and imino forms are both more unstable, but their difference in energy varies depending on the theoretical model employed. Our results show that the energy between the four tautomers ranges from 1 to 5 kcal mol^{-1}. These values are in good agreement with other theoretical studies that obtained differences in energy smaller than 10 kcal mol^{-1} using diverse methods. Such small energetic differences suggest that the barriers of the tautomerization processes of cytosine can be easily overcome at room temperature.[73] Actually, the energetic fluctuations of these tautomers during the MD simulation is larger than 5 kcal/mol, each of the four tautomers being the lowest-energy form at some point of the simulation.

3.2. Absorption Spectra of Cytosine Tautomers. The spectra of cytosine tautomers are shown in Figures 5−9. The global absorption spectrum has been decomposed into its individual contributions for each transition type, the $\pi\pi^*$, $n_O\pi^*$, and/or $n_N\pi^*$. Since the wave function describes the complete electronic system involved in this kind of excitation, we know in detail the characteristics of those excited states. However, the classification of transitions between the $\pi\pi^*$ and the $n\pi^*$ is not straightforward. Once the molecule becomes distorted, it is not strictly possible to talk about the symmetric features of the planar structure. The MOs can mix formal aromatic π orbitals with σ orbitals to some degree. Moreover, the MO approach employed in these calculations is intrinsically delocalized, which can result in nonbonding σ orbitals spanning over two atoms with lone electron pairs. Therefore, the breakdown of the total absorption spectrum has been done by thoroughly analyzing the contributions to the orbitals in the CAS and by a population analysis of the atoms in the excited states.

We started by identifying all of the active orbitals of each conformer by means of their atomic orbital coefficients. We focused on finding the nonbonding orbitals of the corresponding O and N atoms with lone pairs. No attempts were made to classify the remaining π orbitals. The deviations from the strict π symmetry of the planar geometry are significant, making it too complicated to discriminate between the $\pi\pi^*$ transitions. The next step consisted of distinguishing those electronic transitions that significantly change the population of the nonbonding orbitals. We set a minimum criterion for the nonbonding orbitals of 0.4 electrons for the difference between the ground state and the excited state to accept those transitions as $n\pi^*$ excitations. The rest of the transitions were assigned to $\pi\pi^*$.

It must be noted that in the nonideal conformers taken from the simulation, the composition of the 12 lowest states is not equal for each structure of a single tautomer. The proportion of $n\pi^*$ and $\pi\pi^*$ transitions varies from one structure to another, resulting in a total number of different transitions found along the series of conformers to be larger than 11. Depending on the structural distortions of the conformer, it can show either a complete lack of $n\pi^*$ transitions or more $n\pi^*$ than $\pi\pi^*$ excitations.

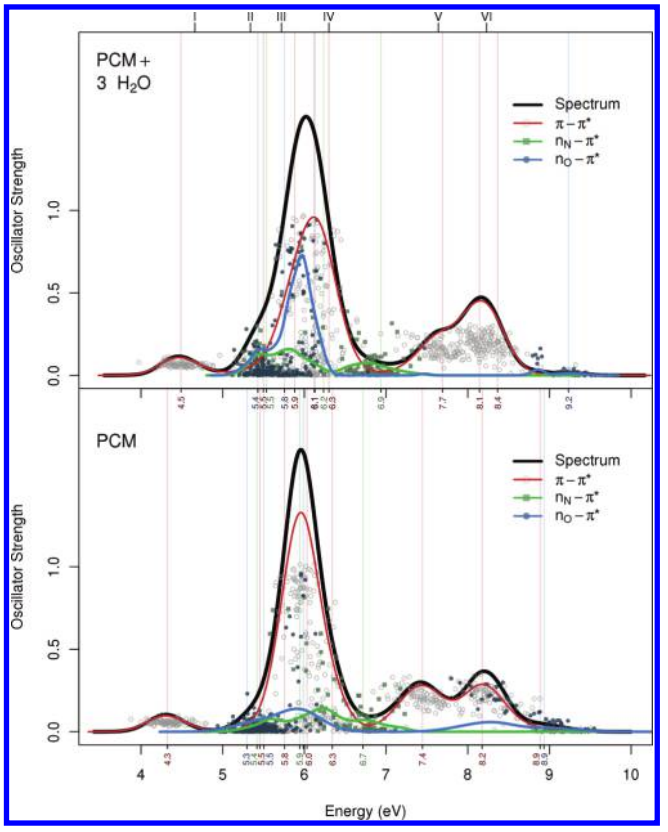

Figure 5. Computed absorption spectra of cytosine keto N1H tautomer. Result from 100 conformers using a PCM solvent model (bottom) and, additionally, with three explicit water molecules (top). Density estimation with Gaussian kernel functions (bw = 0.14 eV). Roman numbers on top mark the experimentally observed features.[16]

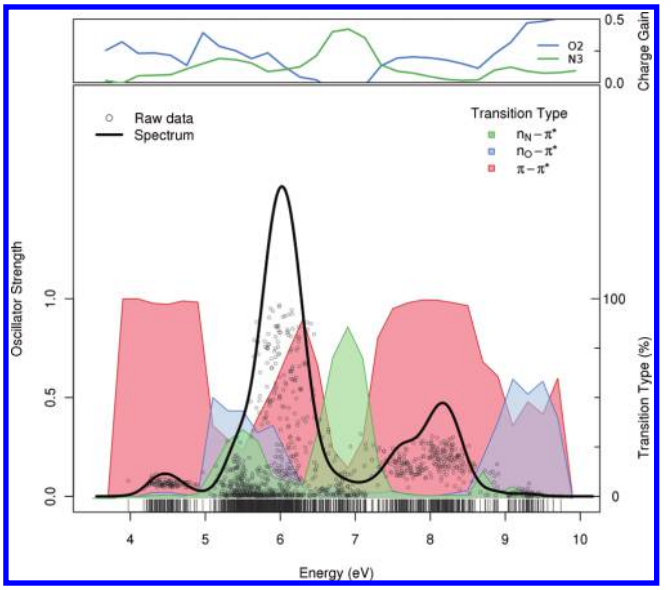

Figure 6. Analysis of the absorption spectra of cytosine keto N1H tautomer with three explicit water molecules. The colored graphs behind the spectrum show the percentage in intervals of 0.2 eV of each type of transition, and the graph on the upper part shows the LoProp population analysis on the O and N atoms with lone pairs.

Keto N1H Tautomer. Figure 5 shows the computed absorption spectra of the cytosine keto N1H tautomer in water. We have identified eight $\pi\pi^*$, four $n_O\pi^*$, and three $n_N\pi^*$ transitions along all conformers of this tautomer. The number of these transitions is independent of the solvent model employed. The explicit water molecules cause a slight blue shift of ~0.2 eV on all transitions except for a few $\pi\pi^*$ of the high absorption band that remain stable. The major effect of the inclusion of the water molecules in the calculation is a large increase on intensity of some $n_O\pi^*$ excitations. These $n_O\pi^*$'s are located in the region of spectral congestion around ~6 eV, where the most intense $\pi\pi^*$ transitions occur. This characteristic suggests that the water molecules could favor the π character of these transitions, increasing the overlap between the involved states and, consequently, increasing their absorption. Additionally, the three water molecules improve the description of the highest excited states, reducing the mixing between the $\pi\pi^*$ and $n_O\pi^*$ transitions observed at 9 eV in the PCM spectrum.

There are four pure $\pi\pi^*$ bands in the spectrum of the keto N1H, the first absorption band located at 4.5 eV and the two peaks at 7.7 and 8.1 eV, which mask a fourth transition at 8.4 eV. The other spectral features that can be unequivocally assigned to single excitations are two optically nearly inactive transitions, a $n_N\pi^*$ at 6.9 eV and a $n_O\pi^*$ at 9.2 eV. The remaining electronic transitions appear in the high absorption band at ~6 eV, which is formed by a mix of various absorptions. The transitions observed can be grouped into three main bands that correlate rather well with the bands identified by Zaloudek et al.[16] Transition II has contributions of all three types of excitations. Transition III is quite intense and formed equally by $\pi\pi^*$ and $n_O\pi^*$ absorptions, and transition IV is raised basically by two $\pi\pi^*$, which seem to correlate to the E_{1u} state of benzene, as suggested by Tinoco and Clark.[11] Furthermore, transition I clearly corresponds to the first $\pi\pi^*$ band at 4.5 eV. In the higher energy region, we observe three $\pi\pi^*$ transitions that match very accurately in energy and band morphology to transitions V, formed by one $\pi\pi^*$, and transition VI, formed by two $\pi\pi^*$. Even though the computed spectrum is much more complicated than previously expected, it has very good agreement with the experimental data of this tautomer.

We performed a second analysis of the absorption spectrum of the keto N1H tautomer to confirm the results obtained by the classification of transitions previously made. We calculated in intervals of 0.2 eV the percentage of absorptions assigned to each type of electronic transition considered for the keto N1H form, the $\pi\pi^*$, $n_O\pi^*$, and $n_N\pi^*$. Additionally, we computed the population of the O and N atoms with lone electron pairs, namely the O2 and N3 atoms, by means of the LoProp partitioning scheme.[74] Figure 6 shows a graphical representation of both data for comparison. The charge variation on the N atom correlates well with the regions of the spectrum with large contributions of $n_N\pi^*$ transitions, the region between 5 and 6 eV, and the $n_N\pi^*$ transition at 6.9 eV. Moreover, the agreement between the charge gains of the O atom and the assignments of $n_O\pi^*$ transitions performed is also good. The larger charge gains for O are found primarily in the 5−6 eV region and at energies higher than 9 eV, where the $n_O\pi^*$ bands have been identified. The nonzero charge gain for O in some regions of the spectrum dominated by $\pi\pi^*$ transitions is due to the π orbital of the O atom, which usually forms a MO localized on the carbonyl group. Therefore, the $\pi\pi^*$ transitions involving that π MO will produce significant changes in the charge of the O atom. This is not the case for the N atom, which will delocalize the charge to some part

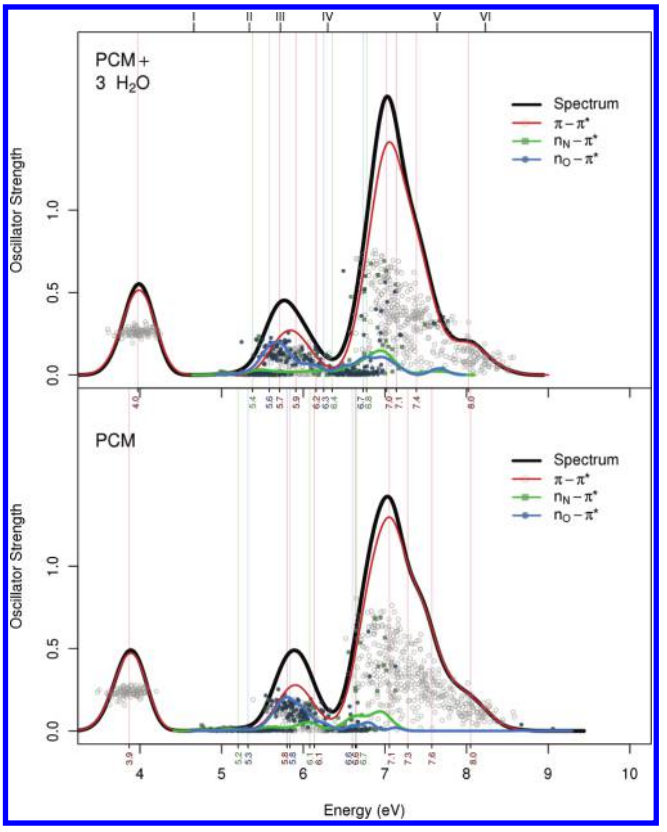

Figure 7. Computed absorption spectra of cytosine keto N3H tautomer. Result from 100 conformers using a PCM solvent model (bottom) and, additionally, with three explicit water molecules (top). Density estimation with Gaussian kernel functions (bw = 0.12 eV). Roman numbers on top mark the experimentally observed features.[16]

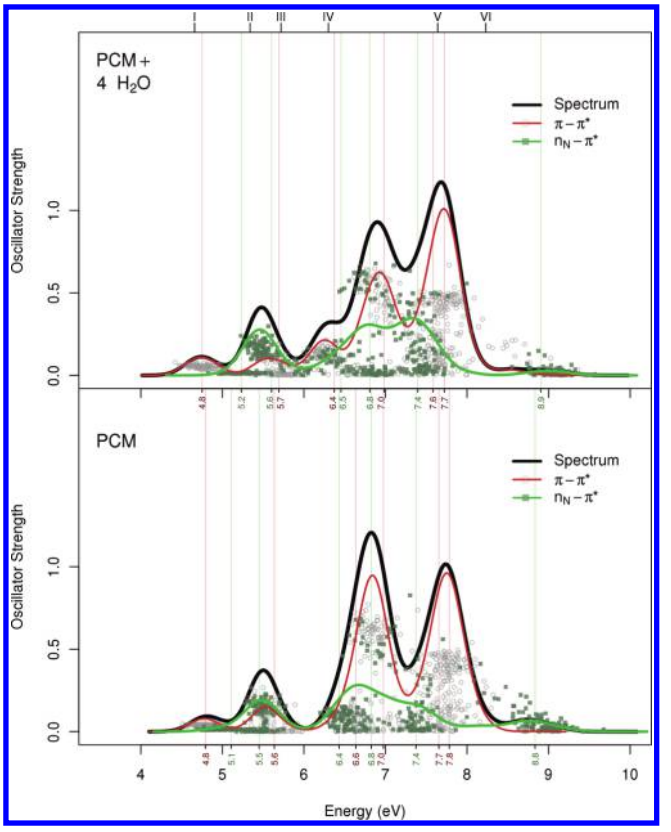

Figure 8. Computed absorption spectra of the cytosine enol tautomer. Result from 100 conformers using a PCM solvent model (bottom) and, additionally, with four explicit water molecules (top). Density estimation with Gaussian kernel functions (bw = 0.14 eV). Roman numbers on top mark the experimentally observed features.[16]

of the pyrimidine ring. This comparison shows that the breakdown performed on the absorption spectra is valid.

Keto N3H Tautomer. Figure 7 shows the computed absorption spectra of the cytosine keto N3H tautomer in water. This tautomer has one fewer $n_O\pi^*$ transition compared with the keto N1H. We have identified eight $\pi\pi^*$, three $n_O\pi^*$, and three $n_N\pi^*$ transitions in total. The number of these transitions is independent of the solvent model employed. The inclusion of explicit water molecules in the calculations of this tautomer does not affect much its electronic structure. The $n\pi^*$ transitions are slightly blue-shifted by ~0.3 eV, while the $\pi\pi^*$ transitions are in general lowered by ~0.2 eV. Qualitatively, the main effect of treating the hydrogen bonds is a larger differentiation between the different types of transitions that form the second band of the spectrum and a small shrinking of the main absorption band. However, we do not observe a high increase in intensity of any $n\pi^*$ transition, as was the case for the keto N1H form; this is ascribed to the small amount of $n\pi^*$ excitation in the main $\pi\pi^*$ band region.

The first band on the spectrum of the keto N3H tautomer is formed solely by transitions of the $\pi\pi^*$ type. The second observed band shows a more complex structure, composed primarily by one $n_O\pi^*$ with relatively high intensity, mixed with two $\pi\pi^*$ transitions. On the red edge of the second band, there is a $n_N\pi^*$ transition that does not contribute to the absorption. Similarly, in the optically inactive region between the second and third band, there are dark transitions of all types. The high absorption band of this spectrum has a maximum plus two shoulders at higher energies of decreasing intensity; all of these features are caused by intense $\pi\pi^*$ transitions.

Enol Tautomer. Figure 8 shows the computed absorption spectra of the cytosine enol tautomer in water. The spectral analysis of this tautomeric form is simpler compared to the keto and imino forms. There are no $n_O\pi^*$ transitions, and only one type of $n_N\pi^*$ can be considered because the lone pairs of the two N atoms are indistinguishable. Accordingly, six $\pi\pi^*$ and six $n_N\pi^*$ transitions are identified along all enol conformers. The inclusion of explicit water molecules produces small blue shifts on some $n_N\pi^*$ transitions and small red shifts of some $\pi\pi^*$ transitions, in line with the results of the previous tautomers. The largest effect is found in the third $\pi\pi^*$ transition at the beginning of the high absorption bands which forms a new visible band at 6.4 eV. Moreover, the $n_N\pi^*$ transitions located in regions of the spectrum with a large number of intense $\pi\pi^*$ transitions, like the second band at 5.5 eV and the two high absorption bands found between 6.5 and 8.0 eV, experience a significant gain of intensity upon inclusion of the water molecules in the CASPT2 calculation. Similarly to the other tautomers, this phenomenon arises from an increase of the π character of these transitions, increasing the overlap between the ground state and the involved excited state.

The peaks of absorption of the enol tautomer spectrum are well-defined and have contributions from few transitions, easing their assignments. The first and fifth absorption bands,

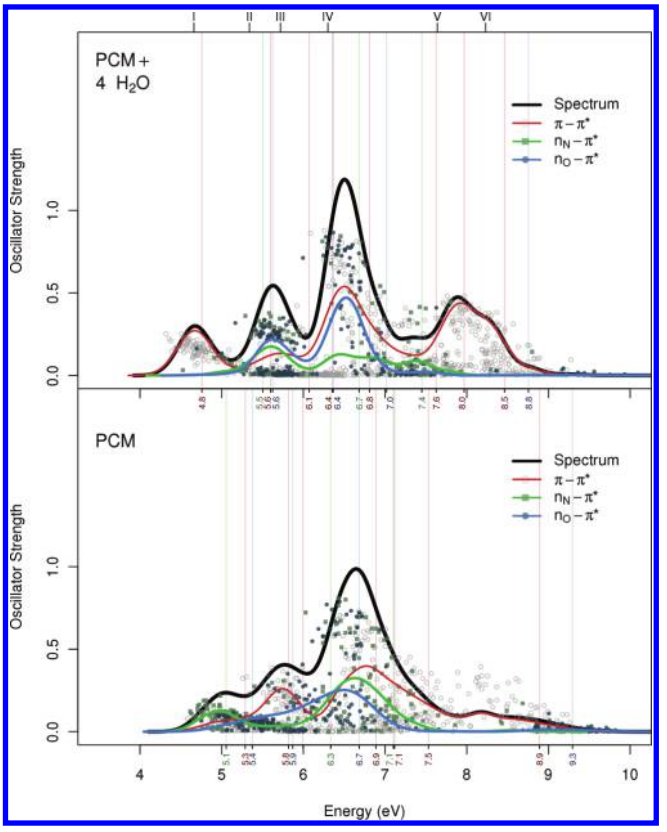

Figure 9. Computed absorption spectra of cytosine imino tautomer. Result from 100 conformers using a PCM solvent model (bottom) and, additionally, with four explicit water molecules (top). Density estimation with Gaussian kernel functions (bw = 0.14 eV). Roman numbers on top mark the experimentally observed features.[16]

at 4.8 and 7.7 eV, respectively, can be assigned to pure $\pi\pi^*$ bands. Instead, the second band is caused primarily by $n_N\pi^*$ excitations. The complexity of this spectrum increases for the third and fourth bands, at 6.4 and 6.9 eV, which appear as a mix with similar weights of both types of transition. Furthermore, there are some features which are masked in the spectrum. Located at 7.3 eV, in the valley between the two high absorption peaks of the fourth and fifth bands, there is a $n_N\pi^*$ band of lower intensity but that is optically active. The rest of transitions do not contribute to absorption, like the excitations found at ∼9 eV.

Imino Tautomer. Figure 9 shows the computed absorption spectra of the cytosine imino tautomer in water. We have identified eight $\pi\pi^*$, three $n_O\pi^*$, and three $n_N\pi^*$ transitions along all conformers. This tautomer shows the largest change upon the addition of explicit water molecules in the solvent model. The basic morphology of the spectra is formed by three peaks of increasing intensity, which are present in both solvent models. However, the inclusion of water molecules into the imino tautomer forms a fourth band in the higher energy region of the spectra. This band appears by the concentration of $\pi\pi^*$ transitions that otherwise are found as a disperse cloud in the PCM spectrum. There is a relatively large blue shift of a $\pi\pi^*$ excitation from 7.1 eV in PCM to 8.0 eV including four waters. Moreover, there is also a red shift of another intense $\pi\pi^*$ transition from 8.9 to 8.5 eV. These intense absorptions are the main peak and shoulder of the fourth band. Additionally, the inclusion of water molecules to the imino form has two effects that we already observed in the other tautomers. The first one is the increase in absorption of some $n_O\pi^*$ excitations located in the most intense band, due to an increase of their π character that leads to a larger overlap between the involved states. The second one is a higher differentiation between the transitions that contribute to each absorption band, which in this case is observed along all of the spectrum.

The first band on the absorption spectrum of the imino tautomer is a pure $\pi\pi^*$ band. However, the source of these excitations is not the π MO on the carbonyl group as for the two keto tautomers. Instead, they involve a π MO mainly localized on the amine group. These transitions are identified as $n_N\pi^*$ excitations in the spectrum without explicit water molecules. The hydrogen bonds formed with surrounding solvent molecules stabilize the lone pair in the N atom, increasing the $\pi\pi^*$ character of the lowest electronic excitations and pushing the first $n_N\pi^*$ transitions into the second absorption band. This band is formed by a mix of three transitions, each one of a different type, that contribute in the same proportion to the total absorption. Subsequently, an optically almost inactive $\pi\pi^*$ transition is identified at 6.1 eV. The third band is the most intense feature in the imino tautomer spectrum. One $\pi\pi^*$ and one $n_O\pi^*$ transition are the main excitations behind this band. Additionally, there is a weak $\pi\pi^*$ transition that produces a shoulder on the end of the third band. There are two dark transitions masked by the intense absorption, a $n_N\pi^*$ at 6.7 eV and a $n_O\pi^*$ at 7.0 eV. Furthermore, a small peak appears in the valley between the third and fourth bands caused by a $n_N\pi^*$ transition at 7.4 eV. The high end of the spectrum shows a low absorption produced by a $n_O\pi^*$ transition.

4. CONCLUSIONS

We have successfully combined the Car−Parrinello MD method with CASPT2 calculations to describe the absorption spectrum of four cytosine tautomers (Figure 1). We explored the conformational space of these tautomeric forms in water, extracted a representative sample of 100 conformers for each tautomer and studied their electronic structure with an efficient treatment of electron correlation. The wave function based calculations included the solvent effects through a PCM for water and some additional explicit water molecules. From the resulting data set, we built the spectrum of each tautomer and analyzed their underlying structure.

The computed absorption spectrum of the keto N1H tautomer has a very good agreement with previous experimental data (Figure 5). The combination of Car−Parrinello MD and CASPT2 generates a theoretical spectrum of cytosine that reproduces not only the energetics but also the bandwidths and intensities of the spectral features. The analysis of the calculated excitations makes it possible to identify them by their characteristics and describe the different individual contributions to the spectrum. The six observed transitions in the experimental spectrum are formed by a total of 15 transitions, which are grouped to give rise to the six experimental absorptions. The difference in energy between the calculated and the experimental bands is very low, on the order of 0.1 eV. A comparison with the LoProp population of the O and N atoms with lone electron pairs proved the classification of the different transition contributing to the spectrum to be correct (Figure 6). Additionally, following the same procedure performed on the keto N1H tautomer, we computed and analyzed the absorption spectra of the other three

tautomers. All of them show similar characteristics to the keto N1H tautomer but with significant differences on their spectrum composition.

The results obtained confirm the complexity of studying the cytosine molecule. Even though it is a relatively small molecule, its flexibility and tautomerization phenomena impose the use of dynamical techniques to cover vibro-rotational degrees of freedom that are not accessible from static approaches. Moreover, the increased complexity also affects the electronic structure and its analysis. Each tautomer shows different electronic properties which require an individualized study of them. The absorption spectra are formed by various transition types, which we classified in $n_O\pi^*$, $n_N\pi^*$, and $\pi\pi^*$ excitations. These archetypical transitions can mix due to the conformational distortions that break the planar symmetry of the ideal cytosine molecule.

The effect on the absorption spectrum of the hydrogen bonds between cytosine and solvent molecules is to shift both $n\pi^*$ and $\pi\pi^*$ transitions. However, there is not a clear trend on the change produced in the absorption energies. The $n\pi^*$ transitions are usually blue-shifted, but the effect on $\pi\pi^*$ transitions is very variable. We observe that the hydrogen bonds combined with the structural distortions of the conformer can favor even more the mix between excitations involving a nonbonding MO with those involving a π MO. Some $n\pi^*$ excitations experience a large increase of intensity due to their increased π character. This behavior shows that once the conformational distortions are incorporated into the theoretical model, it is no longer possible to get pure transitions of one single type. Thus, there is no simple description of the electronic structure of cytosine, even though its molecular structure and absorption spectra can look simple.

AUTHOR INFORMATION

Corresponding Author

*E-mail: alex.domingo@urv.cat; coen.degraaf@urv.cat.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Financial support has been provided by the Spanish Ministry of Science and Innovation (Projects CTQ2008-06644-C02-01 and CTQ2008-06549-C02-01), the Generalitat de Catalunya (Project 2009SGR462 and *Xarxa d'R+D+I en Química Teòrica i Computacional*, XRQTC), and COST Action CODECS, CM1002.

REFERENCES

(1) Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernández, C. E.; Kohler, B. *Annu. Rev. Phys. Chem.* **2009**, *60*, 217−239. PMID: 19012538.

(2) Crespo-Hernández, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. *Chem. Rev.* **2004**, *104*, 1977−2020. PMID: 15080719

(3) Merchán, M.; Serrano-Andrés, L. *J. Am. Chem. Soc.* **2003**, *125*, 8108−8109. PMID: 12837073.

(4) Serrano-Andrés, L.; Merchán, M. *THEOCHEM* **2005**, *729*, 99–108. Proceedings of the 30th International Congress of Theoretical Chemists of Latin Expression, The 30th International Congress of Theoretical Chemists of Latin Expression.

(5) Merchán, M.; González-Luque, R.; Climent, T.; Serrano-Andrés, L.; Rodríguez, E.; Reguero, M.; Peláez, D. *J. Phys. Chem. B* **2006**, *110*, 26471−26476. PMID: 17181307.

(6) González-Vázquez, J.; González, L. *ChemPhysChem* **2010**, *11*, 3617–3624.

(7) Barbatti, M.; Aquino, A. J. A.; Szymczak, J. J.; Nachtigallova, D.; Lischka, H. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6145–6155.

(8) Merchán, M.; Serrano-Andrés, L.; Robb, M. A.; Blancafort, L. *J. Am. Chem. Soc.* **2005**, *127*, 1820−1825. PMID: 15701017.

(9) Blancafort, L. *Photochem. Photobiol.* **2007**, *83*, 603–610.

(10) Kosma, K.; Schröter, C.; Samoylova, E.; Hertel, I. V.; Schultz, T. *J. Am. Chem. Soc.* **2009**, *131*, 16939−16943, PMID: 19874018.

(11) Clark, L. B.; Tinoco, I. *J. Am. Chem. Soc.* **1965**, *87*, 11–15.

(12) Clark, L. B.; Peschel, G. G.; Tinoco, I. *J. Phys. Chem.* **1965**, *69*, 3615–3618.

(13) Voelter, W.; Records, R.; Bunnenberg, E.; Djerassi, C. *J. Am. Chem. Soc.* **1968**, *90*, 6163–6170.

(14) Lewis, T. P.; Eaton, W. A. *J. Am. Chem. Soc.* **1971**, *93*, 2054–2056.

(15) Callis, P. R.; Simpson, W. *J. Am. Chem. Soc.* **1970**, *92*, 3593–3599.

(16) Zaloudek, F.; Novros, J. S.; Clark, L. B. *J. Am. Chem. Soc.* **1985**, *107*, 7344–7351.

(17) Matos, J. M. O.; Roos, B. O. *J. Am. Chem. Soc.* **1988**, *110*, 7664–7671.

(18) Petke, J. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Phys. Chem.* **1992**, *96*, 6992–7001.

(19) Alyoubi, A. O.; Hilal, R. H. *Biophys. Chem.* **1995**, *55*, 231–237.

(20) Hug, W.; Tinoco, I. *J. Am. Chem. Soc.* **1973**, *95*, 2803–2813.

(21) Hug, W.; Tinoco, I. *J. Am. Chem. Soc.* **1974**, *96*, 665–673.

(22) Fülscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 2089–2095.

(23) Billinghurst, B. E.; Oladepo, S. A.; Loppnow, G. R. *J. Phys. Chem. B* **2009**, *113*, 7392−7397. PMID: 19438283.

(24) Tomić, K.; Tatchen, J.; Marian, C. M. *J. Phys. Chem. A* **2005**, *109*, 8410−8418. PMID: 16834234.

(25) Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, *98*, 6338–6349.

(26) Szczesniak, M.; Szczepaniak, K.; Kwiatkowski, J. S.; KuBulat, K.; Person, W. B. *J. Am. Chem. Soc.* **1988**, *110*, 8319–8330.

(27) Brown, R. D.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1989**, *111*, 2308–2310.

(28) Feyer, V.; Plekan, O.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K. C.; Trofimov, A. B.; Zaytseva, I. L.; Schirmer, J. *J. Phys. Chem. A* **2010**, *114*, 10270–10276.

(29) Bazso, G.; Tarczay, G.; Fogarasi, G.; Szalay, P. G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6799–6807.

(30) Pedone, A.; Biczysko, M.; Barone, V. *ChemPhysChem* **2010**, *11*, 1812–1832.

(31) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.

(32) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811–6821.

(33) Sambrano, J. R.; de Souza, A. R.; Queralt, J. J.; Andrés, J. *Chem. Phys. Lett.* **2000**, *317*, 437–443.

(34) Mercier, Y.; Santoro, F.; Reguero, M.; Improta, R. *J. Phys. Chem. B* **2008**, *112*, 10769−10772. PMID: 18700794.

(35) Improta, R.; Barone, V. *THEOCHEM* **2009**, *914*, 87–93. Time-dependent density-functional theory for molecules and molecular solids.

(36) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 11338–11346.

(37) Broo, A.; Holmén, A. *J. Phys. Chem. A* **1997**, *101*, 3589–3600.

(38) Kowalski, K.; Valiev, M. *J. Phys. Chem. A* **2008**, *112*, 5538−5541. PMID: 18505240.

(39) Brancato, G.; Rega, N.; Barone, V. *Chem. Phys. Lett.* **2010**, *500*, 104–110.

(40) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738.

(41) Olaso-González, G.; Roca-Sanjuán, D.; Serrano-Andrés, L.; Merchán, M. *J. Chem. Phys.* **2006**, *125*, 231102.

(42) Kostko, O.; Bravaya, K.; Krylov, A.; Ahmed, M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2860–2872.

(43) Barbatti, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 4686–4692.

(44) Isayev, O.; Furmanchuk, A.; Shishkin, O. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2007**, *111*, 3476−3480. PMID: 17388492.

(45) Schyman, P.; Laaksonen, A.; Hugosson, H. W. *Chem. Phys. Lett.* **2008**, *462*, 289–294.
(46) Lawson Daku, L. M.; Linares, J.; Boillot, M.-L. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6107–6123.
(47) Manzoni, V.; Lyra, M. L.; Gester, R. M.; Coutinho, K.; Canuto, S. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14023–14033.
(48) Barone, V.; Bloino, J.; Monti, S.; Pedone, A.; Prampolini, G. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10550–10561.
(49) Barone, V.; Bloino, J.; Monti, S.; Pedone, A.; Prampolini, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2160–2166.
(50) Valiev, M.; Kowalski, K. *J. Chem. Phys.* **2006**, *125*, 211101.
(51) Barbatti, M.; Aquino, A. J. A.; Lischka, H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4959–4967.
(52) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
(53) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.
(54) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. *Chem. Phys.* **1980**, *48*, 157–173.
(55) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Chem. Phys.* **2001**, *114*, 5691–5701.
(56) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics*; Cambridge University Press: Cambridge, U.K., 2009.
(57) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.
(58) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
(59) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(60) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(61) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
(62) Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. *J. Chem. Phys.* **2004**, *120*, 300–311.
(63) *CPMD*; IBM Corp.: Endicott, NY, 1990; MPI für Festkörperforschung: Stuttgart, Germany, 1997.
(64) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
(65) Boeyens, J. C. A. *J. Chem. Crystallogr.* **1978**, *8*, 317–320.
(66) Spek, A. L. *Acta Crystallogr., Sect. D* **2009**, *65*, 148–155.
(67) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. *J. Phys. Chem. A* **2004**, *108*, 2851–2858.
(68) Forsberg, N.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **1997**, *274*, 196–204.
(69) Malmqvist, P.-Å.; Roos, B. O. *Chem. Phys. Lett.* **1989**, *155*, 189–194.
(70) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222–239. Proceedings of the Symposium on Software Development for Process and Materials Design.
(71) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T. B.; Pitoňák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224–247.
(72) Team, R. D. C. *R: A Language and Environment for Statistical Computing*. 2009. ISBN 3-900051-07-0.
(73) Kessler, H. *Angew. Chem., Int. Ed.* **1970**, *9*, 219–235.
(74) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–4500.