When Tautomers Matter: UV-Vis Absorption Spectra of Hypoxanthine in Aqueous Solution from Fully Atomistic Simulations

Sara Gómez*,[a] Chiara Cappelli[a]

The UV-Vis spectrum of the solvated purine derivative Hypoxanthine (HYX) is investigated using the Quantum Mechanics/Fluctuating Charges (QM/FQ) multiscale approach combined with a sampling of configurations through atomistic Molecular Dynamics (MD) simulations. Keto 1H7H and 1H9H tautomeric forms of HYX are the most stable in aqueous solution and form different stable complexes with the surrounding water molecules, ultimately affecting the electronic absorption spectra. The final simulated spectrum resulting from the combination of the individual spectra of tautomers agrees very well with most of the characteristics in the measured spectrum. The importance of considering the effect of the solute tautomers and, in parallel, the contribution of the different solvent arrangements around the solute when modeling spectral properties, is highlighted. In addition, the high quality of the computed spectra leads to suggesting an alternative way for acquiring tautomeric populations from combined computational/experimental spectra.

Introduction

There has always been an interest in studying the photophysical and photodynamical properties and behavior of the canonical nucleobases, and their numerous variations and analogues. Hypoxanthine (also denoted as 6-oxopurine, here HYX), displayed in Figure 1, is a purine analogue [1] formed from the exocyclic incorporation of an oxo group at position C6 of purine, and it is closely related to adenine and guanine. HYX can be formed from the deamination of guanine. The biological importance of HYX is due to its occasional presence as a minor purine base in transfer RNA [2], and besides, it is also a plausible prebiotic precursor [3–5]. The excited-state dynamics of HYX in aqueous solution has attracted the attention of many researchers from experimental [6–8] and theoretical perspectives [9–11] and a high degree of photostability has been associated with the fact that its excited state lifetime is the shortest ever observed among canonical and derivative purine bases. [4,6,12]

Because of the structural similarity, as happens for free purine bases, HYX may exist as a mixture of tautomers in solution, which can in turn hamper the interpretation of some experimental results. [13] For HYX, both prototropic

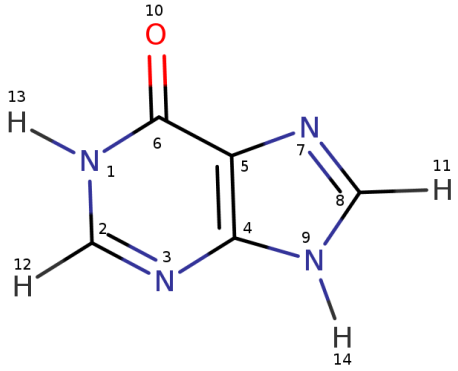

Figure 1. HYX in one of its ketonic tautomeric forms and atom numbering adopted in this work.

(relocation of a hydrogen atom) or keto-enol forms of tautomerism can occur. Therefore, the diverse HYX tautomers have been the subject of many calculations and experiments. Energetics and structural properties, [14–17] electronic structure, properties associated with electrostatic interactions and redox processes [18,19], interactions with water, [20] electronic transitions, excited-state properties and geometries [21,22], solvatochromic effects on the vertical excitation energies [10] are the main investigated aspects. Summarizing the literature, HYX has two major ketonic tautomers, 1H7H and 1H9H, that differ in which of the N7 and N9 positions carries an H atom. They inter-convert to each other or any of the enol forms via proton transfer transition states (TS), that may be water-mediated [23], or induced by either UV irradiation [24] or by the application of a laser on the surface of a bare noble nanomaterial. [25] As a matter of fact, hydration has an appreciable effect on the relative population of different tautomers. [18,19,23] As demonstrated by Shukla and coworkers [23] [26], explicit solvation slightly favors the stability of the keto-1H9H form over the keto-1H7H one, and the enol tautomers become largely destabilized. TSs for the proton transfer from the keto to the enol tautomeric form in isolated HYX or assisted by one or two water molecules have been computationally investigated in both ground and excited states [23,27]. The authors concluded that a high activation energy characterizes keto-enol proton transfer, but it is facilitated by the presence of water molecules [23]. Mechanisms of the intramolecular proton transfers between the prototropic forms of HYX-related compounds have been proposed by Ren and coworkers [28–30] but are scarce for HYX.

The distribution of the HYX tautomers has also been studied with spectroscopic methods. ^{13}C NMR results sug-

[a] Dr. Sara Gómez*, Prof. Dr. Chiara Cappelli
Scuola Normale Superiore, Classe di Scienze, Piazza dei Cavalieri 7, 56126, Pisa, Italy
E-mail: sara.gomezmaya@sns.it

gest that the populations of the tautomeric forms 1H7H and 1H9H are approximately the same.[31] Synchrotron measurements of X-ray absorption spectroscopy and ^{1}H NMR data helped determine that both in solid state and dissolved in an organic solvent, HYX is present in its 1H9H keto form.[32] On the other hand, UV spectroscopic studies in aqueous media indicate the dominance of the keto-1H9H form over the keto-1H7H form.[33] Both dominant neutral tautomers of HYX have also similar ultrashort lifetimes.[7,8,34] It is worth noticing that some spectra have been measured in the gas phase or solution and interpreted in terms of the contributions of several tautomers, using for example computed vibrational frequencies and intensities for the assignment of the tautomers recorded in the experimental setups.[35] For instance, the IR vibrational spectra of all the ketonic and enolic tautomeric forms of neutral HYX have been vastly studied in the gas phase and solution for the isolated forms[18,19,36] and for the monohydrated H-bonded complexes[37]. Fernandez-Quejo *et al.*[36] suggested that to reach a fair comparison with the experimental values of the condensed phase, the right tautomers and a model of solvation like the polarizable continuum model (PCM)[38] must be considered in the calculations. In the same line, experimental and theoretical X-ray studies indicate that simulations reproduce the XPS experimental spectral profiles of HYX just when the two most stable tautomers (keto 1H7H and 1H9H) are accounted for in the modeling.[39] Furthermore, other techniques belonging to the spectroscopic family (e.g. UV Resonance Raman measured at 260 nm[40], photoelectron spectroscopy[41], etc) have been used to unveil the behavior of neutral HYX and recently of HYX clusters.[42] In the last decade, the interest has been focused on the physicochemical behavior and interactions between DNA nucleobases -and derivatives like HYX- and nanoparticles. Among some works, the experimental and theoretical surface-enhanced Raman spectroscopy (SERS) of hypoxanthine/gold[43], the real-time monitoring of the HYX proton transfer using also SERS[25], the UV–visible absorption measurements of the interaction of HYX with colloidal ZnS nanoparticles[44], and the electrochemical analyses and coordination behavior of HYX-Au(III) ion[45] can be mentioned.

Regarding excitations, for the complete interpretation of the spectroscopic results found in experiments, researchers can perform, in principle, surface-hopping nonadiabatic Molecular Dynamics (MD) simulations, static explorations of the excited-state potential energy surfaces (PES) of the representative tautomers[35], or configurational sampling with classical MD and then conventional DFT calculations on the ground and/or excited states. Here, we focus on the latter strategy given the unique photostability of HYX that quickly brings the molecule back to the ground state after excitation with UV radiation. In other words, HYX exhibits a fast decay or relaxation mechanism from the excited states, with an estimated lifetime < 1 ps, $\simeq$100 fs.[6–10] Additionally, the combination of MD simulations with the Quantum Mechanics/Fluctuating Charges (QM/FQ) multiscale approach has constituted a suitable computational protocol to accurately study a plethora of spectroscopies and properties of diverse systems in aqueous solution and other complex environments.[46–48] The advantages and drawbacks of continuum approaches when compared to atomistic ones to model the spectroscopy of solvated systems have been recently discussed.[49] We emphasize here that the inclusion of the solvent is a critical step in the calculations. In this respect, based on the study of the spectroscopic properties of key nucleic acid bases in water using the combined quantum mechanical/molecular mechanics (QM/MM) and cluster-continuum computational protocol, Li and coworkers[11] reported that "*the hydrogen bond network can remarkably modify their absorption spectra*"[11] In addition, solvent effects or the influence of hydrogen bonds (HB) on excited states can be rationalized from synergistic computational and experimental data[12,50]

In a previous paper[51] we discussed the conformational preferences, dynamical behavior, hydration patterns, and electronic absorption spectra of methylated xanthines, namely, paraxanthine (1,7-dimethylxanthine), theophylline (1,3-dimethylxanthine and caffeine (1,3,7-trimethylxanthine). In this contribution, we exploit the same strategy to investigate HYX, which is a more challenging molecule, for which some peculiar features seen in various spectroscopies could be related to the very distinct molecular structure of its tautomers. In particular, we focus on the electronic absorption spectra of the two biologically relevant HYX keto-1H7H and keto-1H9H tautomers in aqueous solution. To the best of our knowledge, atomistically computed UV-Vis electronic spectra for solvated HYX are reported here for the first time along with an assessment of alternative solvation models. Moreover, we introduce the derivation of tautomeric populations from a mixed strategy that combines simulated spectra and experimental data.

Methodology

Hypoxanthine neutral tautomers

HYX exists as a mixture of tautomeric forms. After an extensive search of its structural isomers, the B3LYP/6-311++(d,p) minimum-energy ground-state geometries of the HYX keto (also named oxo) and enol (also named hydroxy) motifs are depicted in Figure 2 and listed in Table S1 in the Supporting Information (SI). In such an early step, solvation effects were accounted for by the implicit PCM approach.[38] As a side note, geometries and populations obtained with other DFT methods were virtually identical. Our results suggest that in solvent HYX can exist mainly in two keto tautomeric forms, 1H7H and 1H9H, the latter being the dominant species with 69% of Boltzmann population at 298 K, coming from the 0.3 kcal/mol relative energy that separates the tautomers with the lowest energies. In line with the findings of Ref. 8, the respective 1H3H and other possible tautomeric structures are significantly higher in energy. It has also been reported that hydration increases the population of the keto-1H9H form.[23] The reader is advised to notice that those Boltzmann populations computed with PCM are providing early indications about what is happening for HYX in solution. Notwithstanding, as will be shown below, such populations can be extracted from more rigorous methods, such as MD-based QM/MM spectral calculations. Hence, starting from the optimized structures of the two prototropic forms 1H9H and 1H7H of the molecule, MD runs were carried out.

Molecular Dynamics simulations

MD simulations at 298 K were performed with the GROMACS software (version 2020.3)[52] for one molecule of each

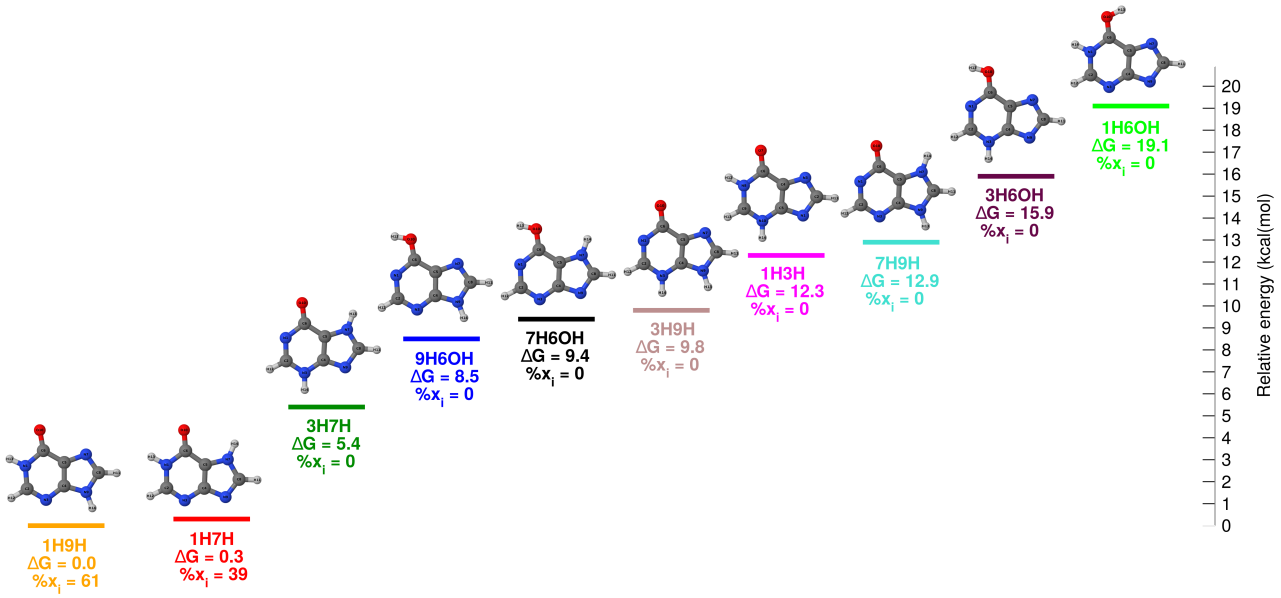

Figure 2. Relative energies (kcal/mol) for keto and enol HYX tautomers. All calculations use the B3LYP/6-311++(d,p) level of theory and the PCM implicit solvation approach. Rotamers of the hydroxy forms (depending on the orientation of the OH group) can also be considered. The relative stabilities between tautomers and consequently their populations can change when specific water$\cdots$HYX interactions are accounted for (see subsection From single tautomer populations to final spectra and viceversa)

tautomeric form in aqueous solution. The Antechamber set of programs allowed the establishment of the General Amber Force Field (GAFF) parameters [53,54] for both tautomers. In the simulations, a cubic cell was used with a side length of 56 Å, and the initial HYX edge distance was equal to 25 Å. TIP3P water molecules [55] were added to fill the corresponding box. Energy minimization was performed with the steepest descendent algorithm. An initial NVT equilibration was conducted with the velocity rescaling method [56] with the coupling constant τ=1.0 ps. Then, a second equilibration was conducted for 1 ns in the NPT ensemble with the same thermostat and the Parrinello-Rahman barostat [57] with a coupling constant τ=5.0 ps (isotropic). In the production stage, the simulations were run for 50 ns, with a time step of 2 fs. Periodic boundary conditions were applied in all directions. Further technical details of the dynamics simulations can be found in a previous study. [51]

MD trajectories were analyzed with the TRAVIS package [58,59], extracting, in particular, the Radial Distribution Functions (RDFs). Finally, a set of 250 uncorrelated snapshots was selected from the last 40 ns of the MD trajectory of each HYX tautomer in water. Such snapshots were cut in a sphere shape of radius 17 Å centered on the solute molecule and they were used in the subsequent QM/FQ spectral calculations. [46]

Excitation energies and absorption spectra

Vertical excitation energies and spectra were computed on each droplet. To choose the level of theory to be used in the time-dependent density functional theory (TD-DFT) spectral calculations, several functionals including B3LYP, CAM-B3LYP, M06, M06-2X, and PBE0 along with a 6-311++(d,p) basis set [60,61] were initially tested using PCM to explore bulk solvent effects. Table S2 in the SI lists computed wavelengths and oscillator strengths for the main electronic transitions of 1H9H HYX. Based on those results, we chose B3LYP as the functional to be employed in the QM/MM computations. The keto-1H7H or keto-1H9H were included in the QM region while the MM region composed of the solvent was modeled by the FQ force field. It is worth mentioning that B3LYP/6-311++(d,p)/FQ calculations have already been successfully applied to three solvated methyl-xanthines systems and theoretical vertical excitation energies do not deviate from experimental values by more than 3%. [51] Water FQ parameters were taken from Ref. 62, FQ^a, and from Ref. 63, FQ^b. TD-DFT calculations were performed using a locally modified version of Gaussian16 [64]. In all cases, 12 vertical excitation energies per snapshot were obtained. Absorption spectra were further fitted to Gaussian functions with full width at half maximum (FWHM) of 0.6 eV. The final spectra are the result of averaging over the MD snapshots. For the sake of comparison, we also calculated the UV-Vis spectra using the non-polarizable TIP3P force field [55] to describe the solvent molecules. This will be called QM/TIP3P in what follows. Additionally, we have considered the water molecules in the first solvation shell of the HYX and computed the spectra including them in the QM portion. That approach will be denoted as the QM/QM$_w$/FQ^a solvation model. Each QM/PCM, QM/TIP3P, QM/FQ^a, QM/FQ^b, and QM/QM$_w$/FQ^a approach is deeply explained in Refs. 51,65.

Natural Bond Orbitals (NBO) calculations [66,67] were performed on one representative snapshot of each trajectory and served two purposes: first, they provide a conceptual basis to estimate the strength of the HB networks formed between solute and solvent [68], and second, they help in the assignment of the nature of the electronic transitions via the Canonical Molecular Orbital (CMO) decomposition. NBO calculations were done with the NBO7 software [69].

We used the computed absorption spectra to suggest an alternative way for acquiring tautomeric populations by combining computational and experimental spectra. An in-

depth explanation of the fitting procedure is provided in the SI, section 3.

Results and Discussion

As mentioned above, the analysis of the tautomeric preferences for the neutral form of HYX in water led to the finding that in solution the keto-1H7H and 1H9H forms are favored, but the tautomeric equilibrium is displaced toward the keto-1H9H form (see relative ground-state energies in Figure 2 and Table S1). Relevant to this work, according to Ref. 33 the UV spectra of HYX in water allowed identify the keto-1H9H form as the dominant motif in solution. In this section, the results of MD runs and spectral computations of the two tautomers performed following a previously described procedure [51] are analyzed. An alternative way to estimate tautomeric populations from joint experimental and computational results is also proposed.

HYX$\cdots$water interactions from MD simulations

The absence of the methyl groups and a single carbonyl bond in the purine skeleton confer less conformational freedom to the HYX structure when compared to methylated xanthines. Still, it gives it higher versatility in the formation of HBs with the solvent. 2 and 4 hydrogen bond donor and acceptor sites, respectively, have been proposed for HYX, [12] but no correlation was found between the vibrational cooling rate and the total number of HBs for a series of xanthines in solution. Hydration patterns from the MD simulations of HYX in water are shown in the top panel of Figure 3 for both tautomers. They were analyzed employing the RDFs between the following potentially HB-involved atoms: the two nitrogen atoms and the carbonyl oxygen in HYX with H atoms in solvent molecules, and the hydrogen atoms linked to nitrogens in HYX with the oxygen atoms in water molecules. The distances for the peak maxima and running coordination numbers (RCN) are summarized in Table 1. RDFs show that both the 1H7H and 1H9H tautomers have almost identical dynamics in solution, but the total RCNs that can be interpreted as the number of waters in the first solvation shell of HYX indicates a slightly preferential interaction between the 1H9H with the solvent. The two selected hydrogen atoms of each tautomer feature a well-defined peak around 1.9 Å, and a coordination number of one water molecule is found after integrating their $g(r)$ curves until the first RDF minimum. In the same way, the carbonyl oxygen atom presents values of O$\cdots$ H$_w$ distances characteristic of standard HBs (1.9 Å), and two water molecules are expected to interact with its lone pairs. From Figure 3 is clear that the nitrogen labeled as N3 is less prone to form an HB with the solvent (in fact, this is a pyridinic nitrogen and no clear peaks are detected in the RDF), while for N7 does appear a peak at about 2.1 Å stating another HB interaction. Thus, about five water molecules are part of the first solvation shell of HYX and they were included in the QM region in the QM/QM$_w$/FQ^a case. Beyond those interactions, secondary HBs formed with the H atoms bonded to carbons have been reported in joint Quantum Chemical and Monte-Carlo studies of the neutral HYX tautomers in water. [20]

To obtain an in-depth insight into the interaction of HYX with the solvent molecules, we selected one snapshot from each trajectory and used NBO to estimate the strengths of individual HBs. Figure 3, bottom panel and the last two columns of Table 1 unveil that there is an interplay between HBs in HYX-water interactions. In the two cases, 1H7H and 1H9H, charge transfers from a lone pair, n to an antibonding, σ^* orbital, are associated with the formation of such HBs and are of the type $n_{\rm O} \rightarrow \sigma^*_{\rm O-H}$, $n_{\rm O} \rightarrow \sigma^*_{\rm N-H}$ or $n_{\rm N} \rightarrow \sigma^*_{\rm O-H}$. The perturbation energies, $E^{(2)}_{d\rightarrow a}$ for such contacts, present typical values for HBs that are not charge-assisted. [70–72] It is important to mention that simultaneous interactions of a water molecule with two sites of HYX can also take place, as in the case of the water near H14 in HYX 1H9H that forms also an HB with the N3 atom, of considerable strength, $E^{(2)}_{d\rightarrow a}$ = 9.94 kcal/mol. Broadly speaking, $E^{(2)}_{d\rightarrow a}$ values obtained for snapshots computed with the QM/QM$_w$/FQ^b parametrization are higher than those provided by QM/QM$_w$/FQ^a.

Excitation energies and absorption spectra

Various models of solvation with distinct levels of refinement can be used to describe the aqueous environment of HYX when computing excitation energies. Herein, the non-polarizable QM/TIP3P and polarizable QM/FQ multiscale methods are applied to perform TD-DFT calculations on multiple snapshots of each solvated HYX tautomers taking the ground state geometries extracted from the MD simulations. The implicit QM/PCM approach and the QM/QM$_w$/FQ^a that consists in including water molecules in the QM portion were also tested. Final spectral profiles using the Boltzmann populations or spectral-derived tautomeric populations are compared with available experimental data. Convergence of the spectra was achieved with 250 snapshots (see Figure S2 in the SI).

Figure 4 depicts the stick-spectra of 1H7H and 1H9H tautomers of HYX in aqueous solution along with their convoluted spectra computed with QM/PCM and QM/FQ^b models. The fact that the diverse arrangements of the water molecules are taken into account in the QM/FQ^b approach, gives a lot of variability in the position and intensities of the different sticks, turning out in an automatic broadening for the bands. Notice that the distribution of the sticks is not similar for the two tautomeric forms, (the 1H9H sticks are more spread), for which band shifts and unequal relative intensities are to be expected.

In terms of sticks, the case of methylated xanthines in our preceding work [51] showed that a single electronic transition to the first excited state, S_1, of the kind HOMO→LUMO, is responsible for the first maximum (energy-wise) in the UV-Vis spectra. However, for HYX, we found a combination of several transitions to generate that band as can be seen in Figure 4 and Figure S1 in the SI. For instance, according to PCM results of the 1H9H tautomer, such a band stems from the sticks at 254.19 nm and 233.73 nm, whose oscillator strengths are 0.1614 and 0.2190, respectively. It should be noted that the relationship between those two electronic transitions and the appearance of the lowest energy band remains valid when other functionals are used in the TD-DFT calculations, as can be consulted in Table S2 in the SI. Figure S1 in the SI also shows the complex nature of the bands of the absorption spectra of HYX in the case of QM/FQ^b, when they are examined in the light of the separated excitations.

The first two columns of Figure 5 compare the averaged

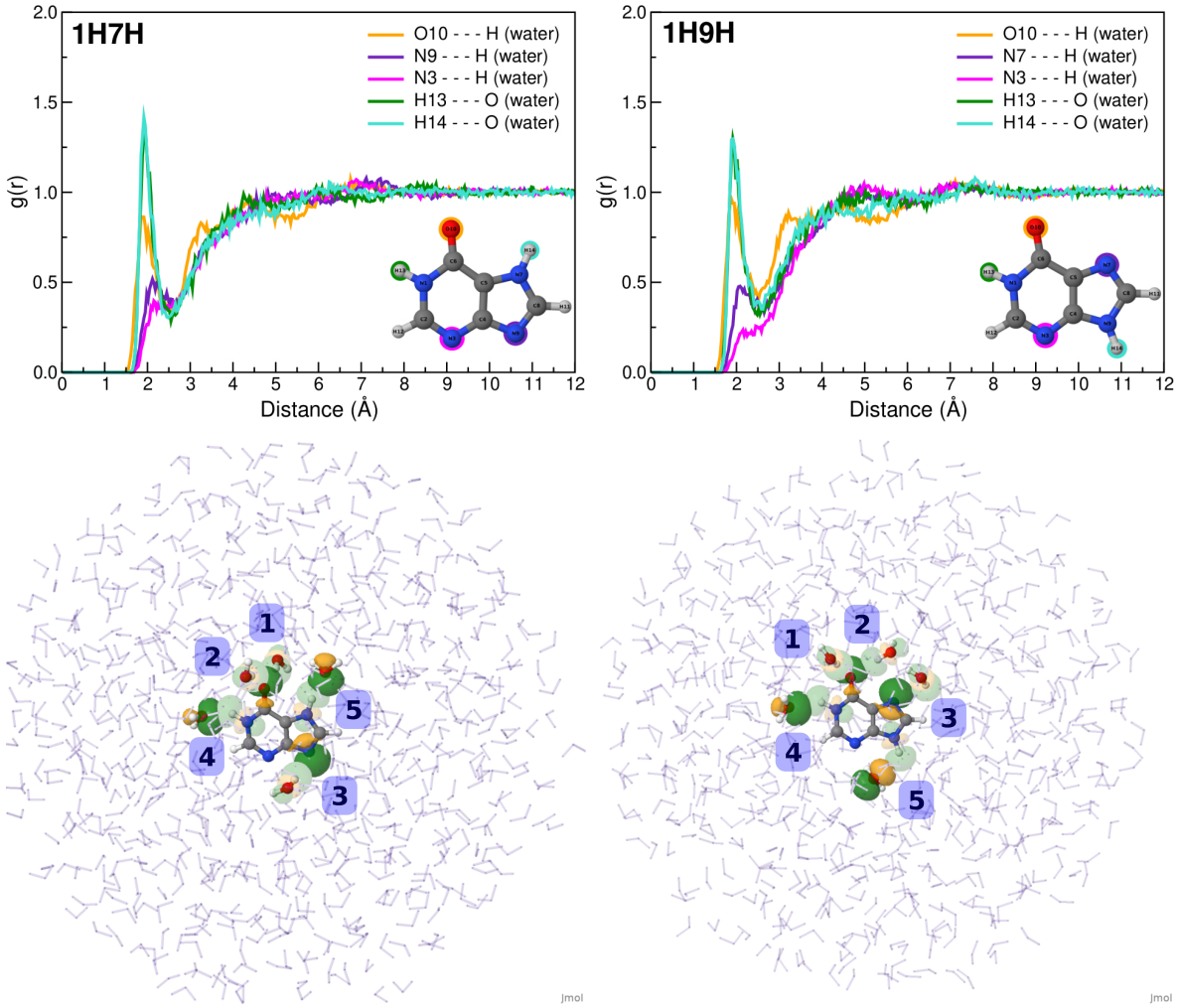

Figure 3. Top panel: Radial distribution function between selected sites of HYX and water molecules. Bottom panel: NBO orbitals involved in the main hydrogen bond interactions of HYX with water. HYX 1H7H on the left and 1H9H on the right. Stabilization energies for the labeled interactions can be read in Table 1. See Figure 1 for atom labeling.

Table 1. Characteristics of the main hydrogen bonds formed when each 1H7H and 1H9H HYX tautomer is solvated with water molecules. d_{RDFmax} (in Å) and the RCN denote the position of the first peak and the running coordination number in the RDF profiles of Figure 3. $E^{(2)}_{d\to a}$ is the stabilization energy related to the interaction between the NBO orbitals (see labels in Figure 3, bottom panel) in a QM/QM$_w$/FQ^a treatment of the system.

label	atoms		d_{RDFmax} (Å)	RCN	NBO orbitals	$E^{(2)}_{d\to a}$ (kcal/mol)
1H7H						
1	O10	H_w	1.9	1.9	$n_{\mathrm{O10}} \to \sigma^*_{\mathrm{O-H}}$	0.82
2						3.77
3	N9	H_w	2.1	1.2	$n_{\mathrm{N9}} \to \sigma^*_{\mathrm{O-H}}$	1.24
	N3	H_w	–	–		
4	H13	O_w	2.0	1.1	$n_{\mathrm{O_w}} \to \sigma^*_{\mathrm{N-H13}}$	10.28
5	H14	O_w	1.9	1.0	$n_{\mathrm{O_w}} \to \sigma^*_{\mathrm{N-H14}}$	11.46
1H9H						
1	O10	H_w	1.9	2.0	$n_{\mathrm{O10}} \to \sigma^*_{\mathrm{O-H}}$	1.59
2						8.63
3	N7	H_w	2.1	1.4	$n_{\mathrm{N7}} \to \sigma^*_{\mathrm{O-H}}$	0.10
	N3	H_w	–	–		
4	H13	O_w	1.9	1.0	$n_{\mathrm{O_w}} \to \sigma^*_{\mathrm{N-H13}}$	0.30
5	H14	O_w	1.9	1.2	$n_{\mathrm{O_w}} \to \sigma^*_{\mathrm{N-H14}}$	4.35

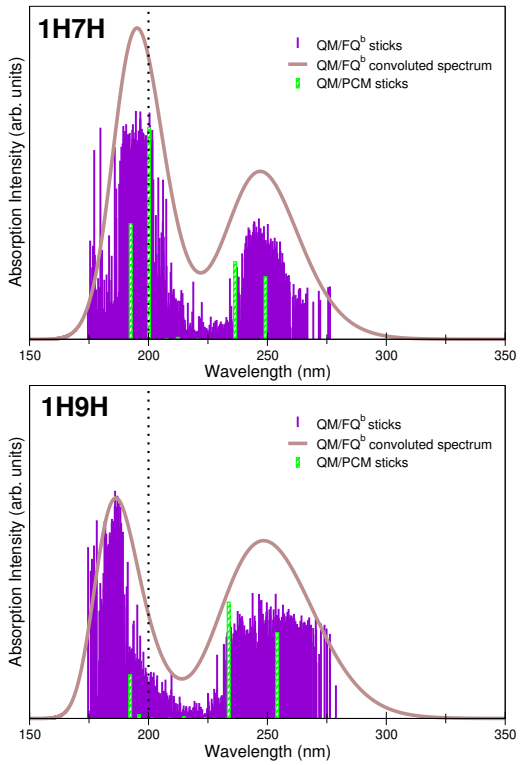

Figure 4. QM/FQ^b calculated data for hypoxanthine in water reported as stick spectrum and convoluted with a Gaussian band shape (FWHM = 0.66 eV). QM/PCM stick spectrum is also reported in green bars. The vertical dotted lines recall that there is no experimental information below 200 nm.

simulated spectra of the two HYX tautomers as obtained with all the approaches employed to model solvent effects. Unlike previous cases[51], notorious differences between the five exploited models are seen in the spectra, in particular in the band shapes and relative intensities. By focusing on the results of the 1H9H tautomer (first column of Figure 5) two observations can be drawn: (*i*) the center of each band slightly shifts with the model, although for QM/TIP3P, QM/FQ^a, and QM/QM$_w$/FQ^a is difficult to pinpoint a center since the first band is very broad up to the limit of recognizing the various involved transitions, (*ii*) the relative intensities change with any chosen approach, overall the second band is lower in intensity than the first one, except in the spectrum provided by the QM/FQ^b model. Conversely, for the 1H7H tautomer (second column of Figure 5), the less expanded distribution of the sticks (see Figures 4 and S1) gives a more consistent picture of the absorption spectra, with almost an unaltered position of the bands and a lower intensity for the main band, irrespective of the exploited approach.

Experimental spectra reproduced from Refs. 6,44 are also included in Figure 5 to assess the quality of our modeling. The measured spectrum for solvated HYX is characterized by a main broad band located at about 250 nm[6,73] and another band near in intensity at about 202 nm[44]. While simulations are relatively successful in getting the band positions, achieving a relative intensity like the one in the experiment is indeed a more difficult task. Overall, PCM results are blue-shifted by ≈0.20 eV with respect to the experimental value, and the relative intensity of the second peak (energy units) can be overestimated or underestimated depending on the tautomer employed in the simulation. By comparing the results of the atomistic approaches with the measured spectrum it is clear that when it comes to the first band (lowest energy band), all models predict a good position for the maximum wavelength with unparalleled similarities in the 1H7H case. Close resemblances are also found for the 1H9H tautomer with the QM/FQ^b model. Nonetheless, band shapes are not always correctly reproduced, most notably in the 1H9H case. Yet, the reproduction of the relative intensities between the bands is a challenge if the spectra of individual tautomers are checked. Notice that the spectrum obtained for HYX 1H9H by using the QM/FQ^b model is the only one keeping the proper relative intensities between the two bands without the need to insert explicit water molecules in the QM portion.

As discussed before, rather different results are obtained when using an implicit or an explicit model to treat solvated HYX. It means that the modeling of the specific solute-solvent interactions, (those confirmed by the RDFs in Figure 3), can affect the simulated spectroscopic signatures. Moreover, given that, to a certain extent, quite similar spectra are built at the QM/TIP3P level of theory and with the other QM/MM polarizable models, polarization effects are not indispensable in the prediction of some characteristics of the spectra of this molecule, for example, the vertical excitation energy for the first transition. Nevertheless, in the thorough simulation of a spectrum and in general of electronic properties, the treatment of the solvent employing dynamic charges could become an important factor.[46,48] This is in concordance with the attained results for caffeine, theophylline, and paraxanthine reported in Ref. 51.

The analysis of the molecular orbitals directly involved in the electronic transitions allows for discerning the type of transition behind the main bands.[51,68,74–76] The identification of the nature of the first transitions of HYX in water is a more intriguing case than for its similar methylated molecules caffeine, paraxanthine, and theophylline. The largest coefficients in the CI expansion of the two excited states giving rise to the first band denote transitions from HOMO to LUMO, and from HOMO to LUMO + 1. According to the NBO analysis, the HOMO of HYX is an assortment of π-character NBOs, with $\pi_{C4=C5}$, $\pi_{N7=C8}$, and $\pi_{C2=N3}$ exhibiting the main contributions, with percentages of 33%, 21% and 19%, respectively. Dominant NBO contributions reported by Farrokhpour and Fathi[41] also stress that the HOMO of HYX is a π molecular orbital, with $\pi_{C4=C5}$ having the most contribution. Charge-receptor molecular orbitals, LUMO and LUMO + 1, admit some mixes as well. First, the LUMO has an approximate composition of 50% $\pi^*_{C2=N3}$ and 17% $\pi^*_{C4=C5}$, with weaker (9% and 7%) contributions from the nitrogen lone pairs, n_{N1} and n_{N9}. Notice that in contrast to the results for prior methylxanthines, this molecular orbital receiving electron charge from the HOMO does not include the $\pi^*_{C=O}$ NBO which was the leading antibond orbital in the description of the electronic transitions for such molecules. Second, a π^* mixture ($\pi^*_{C6=O10}$, $\pi^*_{N7=C8}$, $\pi^*_{C4=C5}$) is associated to the composition of the LUMO + 1 canonical molecular orbital. Thus, it can be stated that the general type of transition is $\pi \rightarrow \pi^*$. Graphical depictions of HOMO, LUMO, and LUMO + 1 are presented in Figure 6 for the 1H9H tautomer.

As pointed out earlier in the literature[73,77] when two oxo groups are present in a general xanthine there is a red-

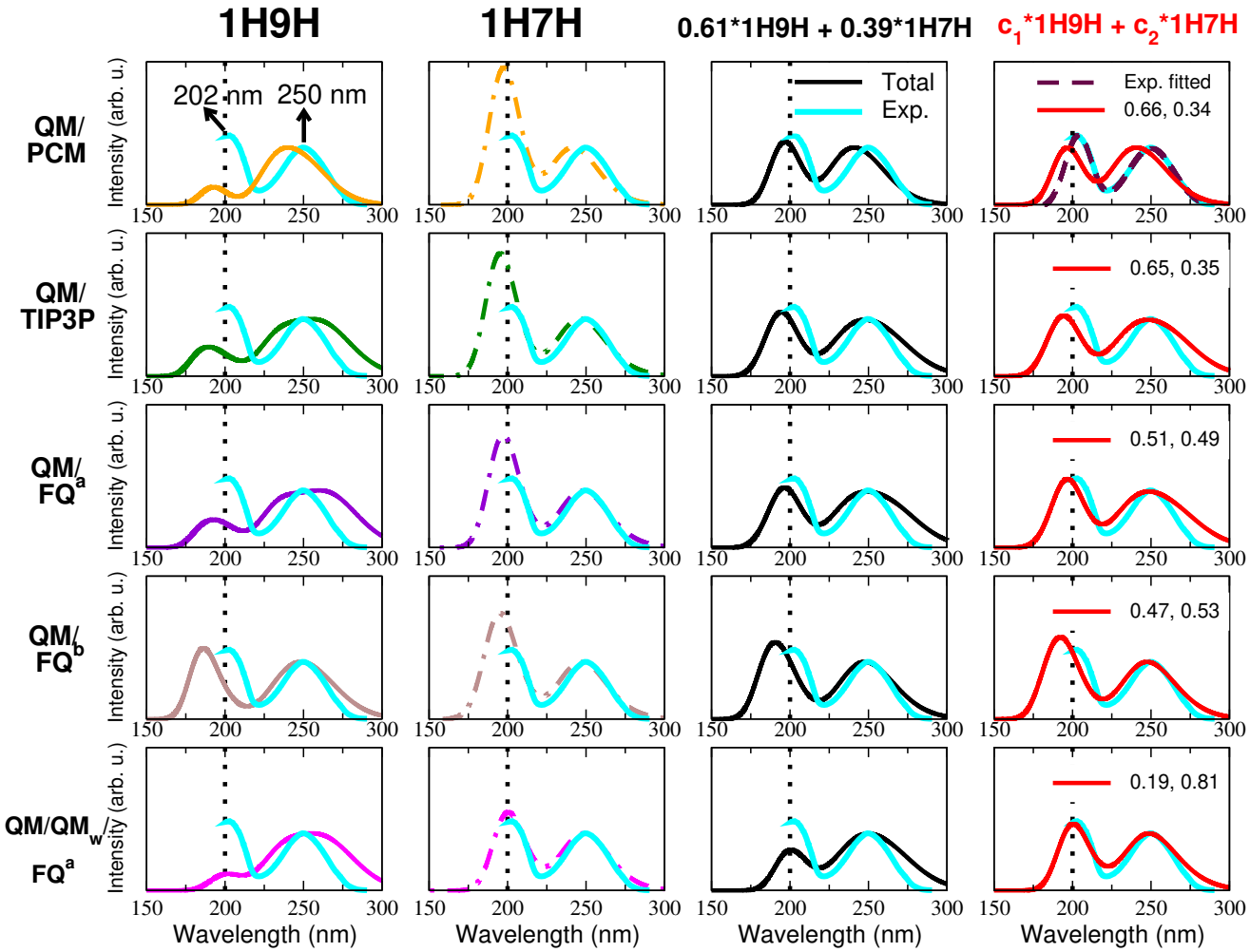

Figure 5. QM/PCM, QM/TIP3P, QM/FQ ([a]Parametrization from Ref. 62 and [b]Parametrization from Ref. 63), QM/QM$_w$/FQ^a, and experimental UV-Vis spectra of HYX in aqueous solution. Experimental data are taken from Refs. 6,44. On average, 5 water molecules were treated at the DFT level in the QM/QM$_w$/FQ^a solvation model. 250 snapshots, extracted from 50 ns Molecular Dynamics runs, were used in the QM/MM calculations. TD-DFT calculations were carried out at B3LYP/6-311++G(d,p) level of theory. Total spectra using the Boltzmann populations (solid black curves) or spectral-derived tautomeric populations (solid red curves) are plotted in the third and fourth columns. For the latter, the coefficients, c_1 and c_2, are written in the legend of the subplots. The vertical dotted lines recall that there is no experimental information below 200 nm.

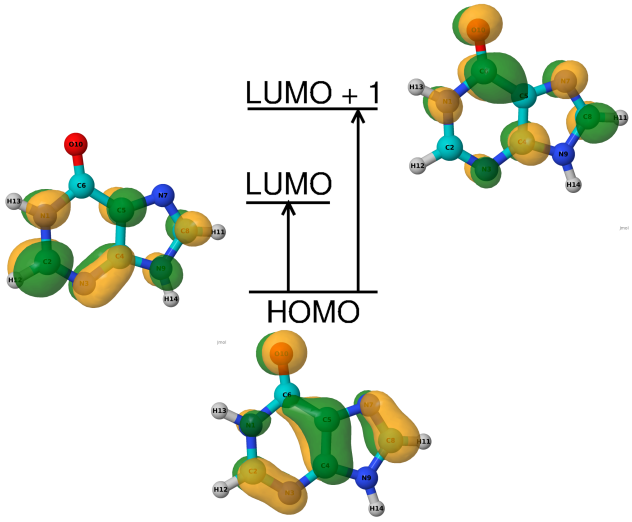

Figure 6. 1H9H-HYX molecular orbitals involved in the double transition that gives rise to the main band in absorption spectra of hypoxanthine in solution.

shift in λ_{max}, which is the case for the methylxanthines [51]. Conversely, for hypoxanthine the second carbonyl group is missing, altering the LUMO coefficients in such a way that this orbital is rather spread out over the ring without contributions of the $\pi^*_{C6=O10}$ orbital. The unique carbonyl antibonding orbital, $\pi^*_{C=O}$ becomes part of the LUMO +1, being much more involved in the charge reception at a shorter wavelength ($\approx$ 233.73 nm in PCM), which shifts the entire band because of having the highest oscillator strength, as listed in Table S2 in the SI. It is relevant to note that Shukla and coworkers [21] performed investigations of the excited-state properties and geometries of the dominant form of HYX (keto-1H9H form) at the CASSCF level and studied the interaction of that tautomer with three water molecules in the excited state. Our transition energies match with the values reported by them, and they also argued that the HOMO and LUMO orbitals contribute the most and are crucial to the excitation processes. On the other hand, the solvatochromic effects on the vertical excitation energies of the separate keto-1H9H and keto-1H7H have been computationally studied. [10].

Building upon observations reported in other works and their experimental UV spectrum of HYX, Röttger *et al.* [8] pointed out that some spectral aspects arise from a superposition of the response of a couple of tautomers [8] that are predicted to coexist in equilibrium in aqueous solution. [33,36,40] However, establishing the precise ratio that tautomers have in the mixture is not a straightforward procedure. Therefore, in the following section, we suggest a different approach to determine populations by combining data from computational and experimental spectra.

From single tautomer populations to final spectra and viceversa

Estimating the proportion of tautomers in a mixture is an extremely challenging task and a major limitation in tautomerism. Despite this, tautomers can be related to the total spectrum of a molecule in different ways employing, for instance, theoretical predictions. The classical approach consists of taking the spectra of each separated component and combining them based on their contributions (populations), which can be either computed or collected from other sources when available. Ref. 78 is a recent example of this technique. To illustrate this road, i.e. single tautomer populations $\mapsto$ final spectrum, in a purely computational approach, Boltzmann populations obtained for our test case, HYX, with PCM, namely, 39% for 1H7H and 61% for 1H9H (see Figure 2 and Table S1) were used to build a total spectrum in the case of each solvation model. The final spectra are shown as solid black curves in the third column of Figure 5. In view of the fact that spectra of HYX 1H7H overestimate the intensity of the band at 202 nm, whereas spectra of HYX 1H9H underestimate them, weighted spectra (black curves) compensate by offering a balance to improve the intensity of that band. It should be noted that in general, all the total spectra present almost the same profiles however, the absorption peaks remain in slightly different positions. Clearly, the inclusion of the two tautomeric forms can properly explain all the intricacies (relative intensities, shapes, positions) of the spectral profile of HYX in an aqueous media. Following a similar procedure, examples of the importance of accounting for the leading tautomers when reproducing spectra of molecules have been reported in the literature for the anticancer agent Doxorubicin (DOX). [79–81]

A second way of looking at the problem of tautomer proportion is an opposite passage to the above-mentioned one and deals with extracting the contribution of each tautomer from the total spectrum, i.e. final spectrum $\mapsto$ single tautomer populations. In simple words, starting from an experimental measurement, the recorded signal is fitted with the computed spectral curves addressing the individual contributions of each separated spectrum, and, therefore, the tautomeric populations. To apply such a mixed strategy to our system, we first fitted the experimental spectrum using one Gaussian function per band to extend the spectral range of the reported data. The top right panel of Figure 5 and Figure S3 in the SI compare the original (solid cyan) and fitted (dashed maroon) experimental spectrum. Then, we employed the Non-Linear Least-Squares Minimization and Curve-Fitting for Python, (*lmfit*) package [82] to decompose the fitted experimental data into a combination of the two simulated spectral curves. For example, denoting $\mathbf{e}$ as the data vector containing the experimental spectrum, $\mathbf{e}$ can be expressed as the combination of several individual spectra $\mathbf{s_i}$ with coefficients c_i that represent their contributions. In the case studied here, ($\mathbf{e}_{\mathrm{HYX}} = c_{\mathrm{1H9H}} \cdot \mathbf{s}_{\mathrm{1H9H}} + c_{\mathrm{1H7H}} \cdot \mathbf{s}_{\mathrm{1H7H}}$). After fitting with the least squares method, *lmfit* provides the coefficients corresponding to the best fit, and ultimately the tautomer populations. Total spectra resulting from this scheme are plotted in the fourth column of Figure 5. c_{1H9H} and c_{1H7H} are written as a legend ($c_{\mathrm{1H9H}}, c_{\mathrm{1H7H}}$) in each subplot and listed in Table S3 in the SI along with the correlation coefficients for the fitting procedure. Other details of the fitting can be found in the SI, section 3. In the QM/PCM and QM/TIP3P cases, the extracted populations agree very well with the Boltzmann ones, while all QM/FQ approaches predict a comparable population (51:49 and 47:53) for the two tautomeric forms of HYX in aqueous solution. Instead, QM/QM$_w$/FQ^a explicitly favors the dominance of the 1H7H tautomer, with a correlation coefficient, R^2=0.87.

This simple methodology is adaptable to different numbers of tautomers and enables us to obtain new relevant information on the contribution of individual structures to

total spectra when populations are not known.

Conclusion

UV-Vis electronic absorption spectra of the two most stable hypoxanthine tautomers (keto-1H7H and keto-1H9H) have been investigated in aqueous solution through MD-based QM/MM simulations. Solvent effects have been included hierarchically with either implicit or explicit QM/MM approaches including the use of fluctuating charges (FQ) in the MM portion to better describe the mutual polarization between the target and the environment.

The results presented in this work indicate that the atomistic description of the solvent always improves the position of the two bands that are reported in experiments and arise from $\pi \rightarrow \pi^*$ transitions. Such findings are indicative of the specific solute-solvent interactions in the form of hydrogen bonds that take place during the solvation of HYX and that dynamically change in response to the different arrangements of the water molecules around HYX. The average strength of such contacts estimated from the RDFs and with NBO suggests that they are moderately strong but not as strong as the charge-assisted ones.

Simulated spectra of the individual tautomers are not able to grasp both the relative intensities and positions of the bands of the single substituted purine HYX in aqueous solution, but our computations do replicate all the features of the UV-Vis experimental spectral profiles only when the theoretical modeling takes into consideration the two most stable tautomers of HYX.

As demonstrated in this paper, our robust computational protocol, which takes advantage of a comprehensive exploration of the configurational phase space of the system, can also retrieve the contribution of the individual components of a sample serving to extract populations of tautomers by fitting computed separated profiles to the recorded spectrum of a given system.

Acknowledgements

We gratefully acknowledge the Center for High-Performance Computing (CHPC) at SNS for providing the computational infrastructure. CC acknowledges the support of the European Union by the Next Generation EU project ECS00000017 'Ecosistema dell'Innovazione' Tuscany Health Ecosystem (THE, PNRR, Spoke 4: Nanotechnologies for diagnosis and therapy).

Conflict of Interest

There are no conflicts to declare.

Keywords: Tautomers • Hypoxanthine • MD • UV-Vis spectra • Quantum Mechanics/Fluctuating Charges

References

[1] S. Matsika, *Modified Nucleobases*, pages 209–243, Springer International Publishing, Cham **2015**.

[2] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science* **1965**, *147*, 1462.

[3] M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, J. P. Dworkin, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13995.

[4] G. Gate, A. Williams, M. R. Haggmark, N. D. Svadlenak, G. Hill, M. S. de Vries, Nucleobases as Molecular Fossils of Prebiotic Photochemistry: Excited-state Dynamics of C2 and C6 Substituted Purines, in *Prebiotic Photochemistry: From Urey–Miller-like Experiments to Recent Findings*, The Royal Society of Chemistry **2021**.

[5] S. J. Hoehn, N. E. Caldero-Rodríguez, C. E. Crespo-Hernández, Photochemistry of RNA, RNA Monomers, and Plausible Prebiotic Precursors, in *DNA Photodamage: From Light Absorption to Cellular Responses and Skin Cancer*, The Royal Society of Chemistry **2021**.

[6] J. Chen, B. Kohler, *Phys. Chem. Chem. Phys.* **2012**, *14*, 10677.

[7] J. P. Villabona-Monsalve, R. Noria, S. Matsika, J. Peón, *J. Am. Chem. Soc.* **2012**, *134*, 7820.

[8] K. Röttger, R. Siewertsen, F. Temps, *Chem. Phys. Lett.* **2012**, *536*, 140.

[9] X. Guo, Z. Lan, Z. Cao, *Phys. Chem. Chem. Phys.* **2013**, *15*, 10777.

[10] X. Guo, Y. Zhao, Z. Cao, *Phys. Chem. Chem. Phys.* **2014**, *16*, 15381.

[11] J. Li, X. Guo, Y. Zhao, Z. Cao, *Int. J. Quantum Chem.* **2015**, *115*, 680.

[12] Y. Zhang, J. Chen, B. Kohler, *J. Phys. Chem. A* **2013**, *117*, 6771.

[13] C. E. Crespo-Hernández, L. Martínez-Fernández, C. Rauer, C. Reichardt, S. Mai, M. Pollum, P. Marquetand, L. González, I. Corral, *J. Am. Chem. Soc.* **2015**, *137*, 4368.

[14] M. Nonella, G. Hänggi, E. Dubler, *J. Mol. Struct.: THEOCHEM* **1993**, *279*, 173.

[15] I. Kondratyuk, S. Samijlenko, D. Hovorun, et al., *J. Mol. Struct.* **2000**, *523*, 109.

[16] S. X. Tian, K. Z. Xu, *Chem. Phys.* **2001**, *264*, 187.

[17] M. E. Costas, R. Acevedo-Chávez, *J. Solution Chem.* **2012**, *41*, 864.

[18] B. Hernández, F. J. Luque, M. Orozco, *J. Org. Chem.* **1996**, *61*, 5964.

[19] M. E. Costas, R. Acevedo-Chávez, *J. Phys. Chem. A* **1997**, *101*, 8309.

[20] M. L. San Román-Zimbrón, M. E. Costas, R. Acevedo-Chávez, *J. Mol. Struct.: THEOCHEM* **2004**, *711*, 83.

[21] M. Shukla, J. Leszczynski, *J. Phys. Chem. A* **2003**, *107*, 5538.

[22] M. Zhang, Y. Guo, X. Feng, X. Jin, L. Qiu, L. Zhu, S. Cui, Y. Sun, Y. Ma, X. Ma, et al., *J. Clust. Sci* **2021**, *32*, 93.

[23] M. Shukla, J. Leszczynski, *J. Phys. Chem. A* **2000**, *104*, 3021.

[24] A. Gerega, L. Lapinski, M. J. Nowak, H. Rostkowska, *J. Phys. Chem. A* **2006**, *110*, 10236.

[25] B. Zhou, S. Li, X. Tang, P. Li, X. Cao, B. Yu, L. Yang, J. Liu, *Nanoscale* **2017**, *9*, 12307.

[26] M. K. Shukla, J. Leszczynski, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 637.

[27] M. Shukla, J. Leszczynski, *Int. J. Quantum Chem.* **2005**, *105*, 387.

[28] H. Ren, W. Wang, W. Wang, N. Wei, *Chin. J. Chem.* **2010**, *28*, 1027.
[29] H.-J. Ren, K.-H. Su, Y. Liu, X.-J. Li, J. Xiao, Y.-L. Wang, *J. Mol. Model.* **2013**, *19*, 3279.
[30] H. J. Ren, *Adv. Mater. Research* **2013**, *690*, 1418.
[31] M. T. Chenon, R. J. Pugmire, D. M. Grant, R. P. Panzica, L. B. Townsend, *J. Am. Chem. Soc.* **1975**, *97*, 4636.
[32] J. Reid, T. Bond, S. Wang, J. Zhou, A. Hu, *Powder Diffr.* **2015**, *30*, 278.
[33] D. Lichtenberg, F. Bergmann, Z. Neiman, *Isr. J. Chem.* **1972**, *10*, 805.
[34] S. Boldissar, M. S. de Vries, *Phys. Chem. Chem. Phys.* **2018**, *20*, 9701.
[35] G. Gate, R. Szabla, M. R. Haggmark, J. Šponer, A. L. Sobolewski, M. S. de Vries, *Phys. Chem. Chem. Phys.* **2019**, *21*, 13474.
[36] M. Fernandez-Quejo, M. De La Fuente, R. Navarro, *J. Mol. Struct.* **2005**, *744*, 749.
[37] R. Ramaekers, A. Dkhissi, L. Adamowicz, G. Maes, *J. Phys. Chem. A* **2002**, *106*, 4502.
[38] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.
[39] O. Plekan, V. Feyer, R. Richter, A. Moise, M. Coreno, K. C. Prince, I. L. Zaytseva, T. E. Moskovskaya, D. Y. Soshnikov, A. B. Trofimov, *J. Phys. Chem. A* **2012**, *116*, 5653.
[40] S. Gogia, A. Jain, M. Puranik, *J. Phys. Chem. B* **2009**, *113*, 15101.
[41] H. Farrokhpour, F. Fathi, *J. Comput. Chem.* **2011**, *32*, 2479.
[42] D. Barreiro-Lage, G. Mattioli, C. Nicolafrancesco, P. Rousseau, A. Milosavljevic, S. Diaz-Tendero, *J. Phys. B: At. Mol. Opt. Phys.* **2024**.
[43] W. Huang, J.-Z. Jiang, L. Chen, B.-Q. Zhang, S.-F. Deng, J. J. Sun, W.-K. Chen, *Electrochim. Acta* **2015**, *164*, 132.
[44] D. Saha, S. Diamai, W. Warjri, D. P. Negi, *J. Mol. Liq.* **2019**, *278*, 460.
[45] L. Jin, J.-Q. Yang, F.-Z. Yang, D.-Y. Wu, Z.-Q. Tian, *J. Electrochem. Soc.* **2020**, *167*, 022511.
[46] T. Giovannini, F. Egidi, C. Cappelli, *Chem. Soc. Rev.* **2020**, *49*, 5664.
[47] T. Giovannini, F. Egidi, C. Cappelli, *Phys. Chem. Chem. Phys.* **2020**, *22*, 22864.
[48] S. Gómez, T. Giovannini, C. Cappelli, *ACS Phys. Chem. Au.* **2023**, *3*, 1.
[49] T. Giovannini, C. Cappelli, *ChemComm* **2023**, *59*, 5644.
[50] G.-J. Zhao, K.-L. Han, *Acc. Chem. Res.* **2012**, *45*, 404.
[51] S. Gómez, T. Giovannini, C. Cappelli, *Phys. Chem. Chem. Phys.* **2020**, *22*, 5929.
[52] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1*, 19.
[53] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
[54] J. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247.
[55] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
[56] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[57] M. Parrinello, A. Rahman, *J. Appl. Phys.* **1981**, *52*, 7182.
[58] M. Brehm, B. Kirchner, *J. Chem. Inf. Model* **2011**, *51*, 2007.
[59] M. Brehm, M. Thomas, S. Gehrke, B. Kirchner, *J. Chem. Phys.* **2020**, *152*, 164105.
[60] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.
[61] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.
[62] S. W. Rick, S. J. Stuart, B. J. Berne, *J. Chem. Phys.* **1994**, *101*, 6141.
[63] T. Giovannini, P. Lafiosca, B. Chandramouli, V. Barone, C. Cappelli, *J. Chem. Phys.* **2019**, *150*, 124102.
[64] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16 Revision A.03 **2016**, gaussian Inc. Wallingford CT.
[65] T. Giovannini, M. Macchiagodena, M. Ambrosetti, A. Puglisi, P. Lafiosca, G. Lo Gerfo, F. Egidi, C. Cappelli, *Int. J. Quantum Chem.* **2019**, *119*, e25684.
[66] F. Weinhold, C. R. Landis, *Discovering Chemistry with Natural Bond Orbitals*, Wiley-VCH, Hoboken NJ, 319pp **2012**.
[67] F. Weinhold, C. Landis, E. Glendening, *Int. Rev. Phys. Chem.* **2016**, *35*, 399.
[68] S. Gómez, M. Ambrosetti, T. Giovannini, C. Cappelli, *J. Phys. Chem. B* **2024**, *128*, 2432.
[69] E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, P. Karafiloglou, C. R. Landis, F. Weinhold, NBO 7.0 **2018**, theoretical Chemistry Institute, University of Wisconsin, Madison, WI.
[70] S. Gómez, N. Rojas-Valencia, S. A. Gómez, C. Cappelli, G. Merino, A. Restrepo, *Chem. Sci.* **2021**, *12*, 9233.
[71] J. David, S. Gómez, D. Guerra, D. Guerra, A. Restrepo, *ChemPhysChem* **2021**, *22*, 2401.
[72] S. Gómez, S. Gómez, J. David, D. Guerra, C. Cappelli, A. Restrepo, *Molecules* **2022**, *27*, 8665.
[73] H. M. Kalckar, M. Shafran, et al., *J. Biol. Chem.* **1947**, *167*, 429.
[74] L. Uribe, S. Gómez, T. Giovannini, F. Egidi, A. Restrepo, *Phys. Chem. Chem. Phys.* **2021**, *23*, 14857.
[75] S. Gómez, N. Rojas-Valencia, T. Giovannini, A. Restrepo, C. Cappelli, *Molecules* **2022**, *27*, 442.
[76] N. Rojas-Valencia, S. Gómez, T. Giovannini, C. Cap-

pelli, A. Restrepo, F. Núñez-Zarur, *J. Phys. Chem. B* **2023**, *127*, 2146.
[77] M. M. Stimson Sr, M. A. Reuter Sr, *J. Am. Chem. Soc.* **1943**, *65*, 153.
[78] A. Patra, A. I. Krylov, S. Mallikarjun Sharada, *J. Chem. Phys.* **2023**, *159*, 064101.
[79] M. Jia, X. Song, Q. Zhang, D. Yang, *J. Clust. Sci.* **2018**, *29*, 673.
[80] E. Florêncio e Silva, E. S. Machado, I. B. Vasconcelos, S. A. Junior, J. D. L. Dutra, R. O. Freire, N. B. da Costa, *J. Chem. Inf. Model* **2019**, *60*, 513.
[81] P. Lafiosca, S. Gómez, T. Giovannini, C. Cappelli, *J. Chem. Theory Comput.* **2022**, *18*, 1765.
[82] M. Newville, T. Stensitzki, D. B. Allen, A. Ingargiola, LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python **2015**.