Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Linear absorption spectra of solvated thiouracils resolved at the hybrid RASPT2/MM level

(Article begins on next page)

28 April 2024

Linear Absorption Spectra of Solvated Thiouracils Resolved at the Hybrid RASPT2/MM level

Artur Nenov[1,§], Irene Conti[1,§], Rocio Borrego-Varillas[2], Giulio Cerullo[2], Marco Garavelli[*,1]

[1]Dipartimento di Chimica Industriale, Università degli Studi di Bologna, Viale del Risorgimento 4, I-40136 Bologna, Italy.
[2]IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy
[§]These authors contributed equally

Abstract

On the example of 2-, 4- and 2,4-thiouracil we demonstrate the performance of the RASPT2/RASSCF protocol in reproducing the spectral positions and line shapes of linear absorption spectra that have been recorded and documented in this work. Through a QM/MM scheme coupled to a room-temperature Wigner sampling we simulate condensed phase spectra, permitting to compare our results against experimental data observed in aqueous solution. In particular, we discuss the sensitivity of the simulations to: a) the active space size by pushing the limits beyond the full-valence active spaces; b) the consideration of "dark" $n\pi^*$-states in the state averaging; c) the flavor of RASPT2 technique; d) the basis set. The benchmarking demonstrates that full-π valence active spaces tend to red-shift the absorption band. Increasing the active space rectifies the problem and we obtain near-quantitative agreement between our experiments and calculations. We, furthermore, demonstrate that the choice of RASPT2 flavor has to be made through rigorous benchmarking.

Introduction

The hybrid quantum mechanics (QM)/ molecular mechanics (MM) approach has established itself as a powerful tool to efficiently incorporate the interactions of a chemical compound of interest (treated at the QM level) with its native environment, i.e. solvent, crystal, protein cavity, etc. (treated at the MM level), trading thereby quantum-mechanical accuracy for a reduced computational cost [1-5]. QM/MM facilitates the development of realistic multiscale models capable of reproducing condensed phase experimental observables. Over the years one has witnessed an increasing level of sophistication regarding the interplay between the quantum and classical layers, going from mechanical [6-8] through electrostatic [9-11] over to polarized embedding schemes [12-15]. The more advanced schemes have revealed weaknesses of their predecessors often responsible for the over- or under-estimation of solvatochromic shifts. The feedback from the QM on the MM layer has been demonstrated to be of importance for the reliable description of photochemical and photophysical processes as the absorption of photons distorts the electronic density (to a new quantum state) which in turn polarizes the environment[12-15].

Next to the QM/MM coupling scheme, the computational level at which the QM layer is treated represents another essential parameter which can affect the accuracy of the QM/MM simulations. QM computational accuracy will be in the focus of the present article. In the field of photochemistry and photophysics, in particular, the description of the multiconfigurational nature of excited states is a challenge, which requires strong approximations in order to facilitate computing. Unwise use could impair severely the quality of the theoretical predictions. Complete active space self-consistent field (CASSCF)[16,17] and its more general formulation restricted active space self-consistent field (RASSCF)[18] augmented with multiconfiguration second order perturbation theory (CASPT2 and RASPT2)[19-21] have found broad acclaim thanks to their versatility and, in many cases, the near-

quantitative agreement with experimental data at comparably low cost (compared to higher order multi-configurational approaches). The RASSCF/RASPT2 protocol allows to limit the active space (AS) to the relevant (mostly valence) orbitals necessary to describe the electronic nature (leading configurations) of the excited states of interest while treating dynamic correlation effects perturbatively. Thereby, active space size and composition represent crucial parameters that influence the final result. Flavors of the RASPT2 scheme, i.e. single-state (SS)[19], multi-state (MS)[22] and extended multi-state (XMS)[23] offer various ways to incorporate dynamic correlation, each one coming with its benefits and weaknesses. Finally, different semi-ab-initio scaling factors (real[24,25] or imaginary shift[26], IPEA shift[27,28]) have been introduced to cope with intruder states or to properly describe dissociation.

In this report we demonstrate the performance of the RASPT2/RASSCF protocol within a QM/MM framework with electrostatic embedding in reproducing linear absorption (LA) spectra of thiobases, nucleobases where the oxygen atom has been replaced by sulfur. The aim of the present work is to define a good protocol to reproduce the spectral positions and the line shapes of these nucleobases derivatives, showing an important biological relevance and an increasing amount of applications [29,30]. Thereby, we discuss the sensitivity of the simulations to:

a) the AS size by pushing the limits beyond the full-valence AS and examining the balance between correlation treated variationally (i.e. at the RASSCF level) and perturbatively (i.e. at the RASPT2 level);

b) the presence of lone pairs in the AS, i.e. the inclusion of "dark" $n\pi^*$-states in the state averaging;

c) the RASPT2 flavor by comparing the performance of SS- and MS-RASPT2;

d) the basis set through comparison of the small basis set ANO-L[321,21][31] (i.e. double-ζ augmented with a set of d-functions to describe orbital polarization effects) to the large basis set ANO-L[4321,321] (i.e. triple-ζ augmented with a two sets of d-functions and a set of f-functions).

At this point we note that the RASSCF/RASPT2 protocol is only one of a wealth of methods which exist for computing the excited state electronic structure: a) other multi-configurational methods such as configuration interaction singles (CIS), double (CISD) and higher orders, multi-reference configuration interaction (MRCI); b) method based on a single determinant as reference such as time-dependent density functional theory (TDDFT), algebraic diagrammatic construction ADC2 and ADC3 and coupled cluster singles (CCS), doubles (CCSD) and higher orders or in the equation of motion (EOM) formalism; c) semi-empirical methods like OM2-MRCI. Compared to other multi-configurational methods, the RASSCF/RASPT2 protocol is computationally less demanding for reaching comparable level of accuracy. Furthermore, it offers a straightforward scheme for improving accuracy to a near-quantitative agreement with experiment – active space increase[32-35] – as evidenced by recently published linear and nonlinear spectroscopy simulations [36,37]. In contrast to all other methods (except MRCI) the user is left with the choice of the configurational space and the challenge is to find the correct balance between electronic correlation treated variationally at CASSCF level and perturbatively at CASPT2 level, thereby avoiding artifacts due to the limited active space size while adhering to reasonable computational timings.

We note that another widely used protocol for describing solute-solvent interaction is the polarizable continuum model (PCM), a simple, yet efficient approach to treat solvation [38-40]. It has been shown to give very good results in combination with linear response methods like TDDFT [41,42]. PCM is also integrated within the RASSCF/RASPT2 scheme in Molcas, however, in our experience its application with this framework is not straightforward. On the one hand it requires state-specific calculation for each state of interest (equilibrium computation for the ground state, a non-equilibrium computation for each excited state), thus increasing the computational cost by a factor equal to the number of states of interest. On the other hand, MS-CASPT2 cannot be performed with PCM. Most

importantly, though, due to excited state energetic proximity and variational nature of the CASSCF procedure the state-specific excited computation often exhibits convergence problems or changes wave function character upon convergence.

The studied systems are the 2-, 4- and 2,4-thiouracil, where one (or two) oxygen atom has been replaced by one (or two) sulfur. In spite of presenting a similar structure to that of canonical bases, the thionation induces dramatic changes in the photophysics: (i) red-shifted linear absorption (LA) spectrum to the UVA range which spans 315-400 nm; (ii) relaxation pathways via the population of triplet states through sub-picosecond intersystem crossing (ISC)[43], in contrast to the ultrafast internal conversion to the ground state mediated by conical intersections observed for the nucleobases[44-46]; (iii) triplet quantum yield near unity. Their particular photochromic properties have indeed demonstrated potential uses in DNA-based nanotechnology, photolabelling or photo-chemotherapy of skin cancer. It is thus not surprising that thionalogues of nucleobases have become a subject of great interest in the last few years, both from the experimental [43,47-59] and computational sides[43,47,49,50,55,60-67].

In this context, major efforts have been concentrated in understanding the dependence of the absorption and ISC properties on the position and number of sulfur-substitutions. A clear example of the rich amount of information provided by linear spectroscopy simulations can be found in the thio-analogues of thymine. Bai et al.[65] calculated the electronic structure of thiothymines in gas phase at the MS-CASPT2 level and simulated LA spectra using the semi-empirical density functional theory/multi-reference configuration interaction (DFT/MRCI) method. They compared the spectra to their parent monomer thymine and proposed a 4-electrons/4-orbital minimum model to explain the experimental spectral features[52]. Specifically, they demonstrated by analyzing the DFT Kohn-Sham orbitals and configuration coefficients, that the wave functions associated with the individual transitions responsible for the absorption bands in the LA spectrum are each dominated by a single configuration state function involving HOMO-1 (π_4), HOMO ($\pi_{S(2|4)}$), LUMO (π^*_1) and LUMO+1 (π^*_2) (see also **Tables 1-3**). Using the MS-CASPT2/SA-CASSCF protocol the lowest-energy absorption band of 4-thiothymine was assigned to the overlap of two $\pi\pi^*$ transitions, a more intense $\pi_{S4}\rightarrow\pi^*_1$ (i.e. HOMO→LUMO) and the less intense $\pi_4\rightarrow\pi^*_1$, both preceded by a dark state with $n\pi^*$ character[48,65]. Solvent effects have been studied within the QM/MM framework, thereby utilizing either the CASPT2/CASSCF[49,67] protocol or the semi-empirical floating occupation molecular orbital configuration interaction (FOMO/CI) method[61], concluding in all cases that the vertical excitation energy of the first spectroscopically bright excited singlet state decreases in aqueous solution with respect to the gas phase.

The thioanalogues of cytosine and guanosine have also attracted considerable attention. MS-CASPT2 calculations on 2-thiocytosine have revealed that the most intense transitions are the $S_4(\pi_s\pi^*)$ and $S_6(\pi\pi^*)$ states, with energies of 4.29 eV and 5.59 eV[43]. Reichardt et al. reported vertical excitation energies for the first bright state $S_2(\pi\pi_{cs}^*)$ of 6-thioguanosine at the DFT level in vacuum, acetronile and water[50]. A gas-phase MS-CASPT2/CASSCF study reported the presence of a low lying dark $n\pi_{cs}^*$ state and identified the nature of the second bright transition as $\pi\pi_{cc}^*$ [61].

Finally, we briefly review previous work on thiouracils, which is the subject of the present manuscript. The absorption spectra of the three systems above 230 nm (i.e. below ~5.5 eV) have a single-peak structure in the case of 2- and 4-thiouracil and a two-peak structure in the case 2,4-thiouracil, all three being red-shifted to a different degree with respect to the parent molecule uracil[51]. Among the sulfur-substituted uracil bases, 2-thiouracil is the molecule that has been studied the most. Mai et al. calculated the absorption spectrum of isolated 2-thiouracil at the MS-CASPT2/SA-3-CASSCF(12,9) level of theory[64] and compared to the experimental gas-phase spectrum by Khvorostov [68]. Due to the limited number of excited states considered the simulations resolved well the intense shoulder at about 290 nm (4.3 eV) arising due to the $\pi_{S2}\rightarrow\pi^*_1$ transition (maximum of the simulated LA spectrum 295 nm), while not capturing the maximum at around 270 nm (4.6 eV) due to the second bright

transition of $\pi_{S4}\rightarrow\pi^*_2$ nature as shown by computations comprising more excited states at the equilibrium geometry [59]. The LA spectrum in polar solvents was found to be shifted only to a minor degree, which is consistent with the dipole moments[51,53]. This was also confirmed by Jiang et al. who computed the electronic transition energies of 2-thiouracil in acetonitrile[55]. As to 4-thiouracil, computations relying on the TD-DFT method assign the intense absorption band around 330 nm to the second excited singlet state S_2 of $\pi_{S4}\rightarrow\pi^*_1$ nature state and the weaker band at 240 nm to the S_6 state of $\pi_{S4}\rightarrow\pi^*_2$ nature [47].

The aims of this work are to assess how the QM/MM scheme performs in reproducing the spectral positions and bandwidths of 2-, 4- and 2,4-thiouracil in its native aqueous environment as a function of the various QM parameters and to dissect the LA spectra, thereby highlighting the contributing electronic transitions.

In order to allow for a proper simulation-to-experiment mapping one has to go beyond electronic structure computations at equilibrium and consider vibrational dynamics and temperature effects to reproduce the spectral line shapes. We utilize Wigner sampling to describe the probabilistic character of the quantum-mechanical wave function and discuss the effect of the sampling on the spectral line shapes.

Computational Details

Classical molecular dynamics

The starting geometry for each of three uracil thioanalogues was obtained following a classical molecular dynamics simulation of the water solvated system, using the AMBER 12 suite [69,70] with the ff10 force field. Periodic boundary conditions were applied to a 35 x 34 x 31 Å (2-thiouracil) 40 x 39 x 35 Å (4-thiouracil) 35 x 34 x 31 Å (2,4-thiouracil) periodic box containing 1229, 1210 and 1255 water molecules, respectively, described by the TIP3P force field [71]. The hydrogen-containing bonds were restrained by means of the SHAKE algorithm [72] while the water geometry was rigidized by the SETTLE scheme [73]. Non-bonding and electrostatic interactions were assessed with a cutoff of 9 Å, making use of the method for quantification of the long-range electrostatics. The time step was set to 0.5 fs and integration of the equations of motion was done with the leap-frog algorithm. Thermalization of the systems was reached by heating of the pre-optimized system to 300 K in 15 ps steps of 100 K. A production run at 300 K and 1 atm was carried out for 500 ps. The lowest energy structure along the dynamics was selected as a starting point for subsequent QM/MM calculations refinement (see the following section).
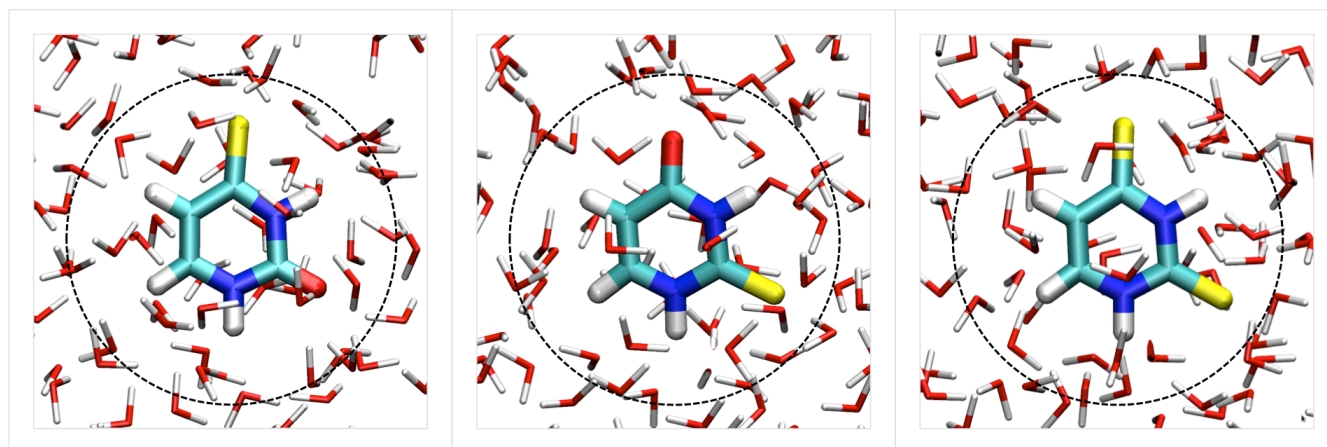

Figure 1. High/Medium/Low layer paritioning of 2- (left), 4- (middle) and 2,4-(right) thiouracil in the QM/MM protocol. The High layer comprises the thiouracil, the Medium layer comprises all waters in 3 Å distance of the QM region (encircled with a dashed line). The rest of waters are included in the Low layer. Note, that for the purpose of LA spectroscopy simulations the differentiation between Middle and Low layers is irrelevant.

QM/MM calculations: geometry refinement and electronic structure computations

QM/MM calculations were performed with the COBRAMM suite developed in our group interfacing various QM codes with AMBER (presented in detail elsewhere[10,74-76]). Prior to High/Medium/Low layer partitioning a spherical droplet centered at thiouracil with a radius of 12 Å (containing 260 waters in 2-thiouracil, 245 waters in 4-thiouracil and 264 waters in 2,4-thiouracil) was cut from the cubic box. In each of the three systems the QM region (**Figure 1**) comprises the thionucleobase. The water molecules in 3 Å distance were included in the MM movable Medium layer (**Figure 1**). The remaining water molecules were kept fixed in the Low layer. Ground state geometry refinement was done using the optimizer of Gaussian 09 [77] at the Møller-Plesset second order perturbation theory (MP2) level as implemented in the Molcas 8 package [78] through its interface with COBRAMM. The generally contracted ANO-L basis set was employed [31], adopting contractions 4s3p2d1f on sulfur, 3s2p1d on carbon, oxygen and nitrogen and 2s1p on hydrogen. The stationary nature of the optimized structures was verified by frequency computations. The electronic spectrum of the three systems was computed at the RASSCF/RASPT2 level as implemented in Molcas 8 through its interface with COBRAMM. In the case of 2- and 4-thiouracil the complete active space comprises the sulfur lone pair and the valence π-orbitals to a total of 12 electrons in 9 orbitals (Tables 1 and 2). The active space is further extended by four extra-valence virtual π-orbitals which are treated at RASSCF level allowing for up to two excitations. In the following we use the shorthand notation RAS(12,9|2,4) when referring to this AS. In the case of 2,4-thiouracil the complete active space comprises both sulfur lone pairs thus resulting in 14 electrons in 10 orbitals (Table 3). Upon extension with four extra-valence virtual π-orbitals (two σ and two σ^* orbitals on the sulfur atoms) we obtain a RAS(2,2|14,10|2,2). The excited state computations were averaged over six states in the case of 2- and 4-thiouracil, while 9 states were taken into account in the case of 2,4-thiouracil. The CASPT2 computations were performed both in the single-state and multi-state flavor. Thereby, the ionization-potential-electron-affinity (IPEA) shift [27,28] was set to 0.0 and an imaginary shift [26] of 0.2 a.u. was used throughout. Transition dipole moments (TDMs) were calculated both at the SA-CASSCF level and at the MS-CASPT2 level using the RASSI routine [79] of the Molcas code. The Cholesky decomposition was used to accelerate the computations of two-electron integrals [80]. The generally contracted ANO-L basis set was employed [31], adopting contractions discussed above, relative to the ground state minimum.

Solution and gas phase vertical excitation energies documented in Table 1, 2 and 3 are calculated on the ground state minimum of the solvated structure obtained from the QM/MM optimization described above and on the minimum calculated at the same level, but without waters, concerning the gas phase values.

Linear absorption spectroscopy: sampling and electronic structure computations

For each system 500 initial conditions were generated using the Wigner sampling technique through an interface with a stand-alone script, part of the quantum molecular dynamics program JADE [81], which considers temperature effects and mode-specific excitation. Consequently, sampling was performed at room temperature (300K). High frequency modes, belonging to C-H and N-H stretchings, were excluded from the sampling. Additionally, in the case of 4- and 2-,4-thiouracil one normal mode with a very low frequency (< 100 cm^{-1}) was excluded from the sampling as it was asserted through preliminary sampling that displacement along such modes breaks the harmonic approximation and distorts excessively the geometry. The sampled geometry is subjected to a 100 ps long MM equilibration run in which the QM region is kept fixed through harmonic constraints. This strategy allows the mobile MM region to adapt to each sampled geometry, thereby removing the bias of the water arrangement of the initial snapshot. The water coordinates of the last snapshot of the MM equilibration run are then combined with the sampled thiouracil geometry to form a snapshot used to compute the electronic properties.

The electronic spectra of the three systems were computed at the RASSCF/RASPT2 level with the

Molcas 8 software package through its interface with COBRAMM. For the purpose of AS benchmarking the minimal full valence-π active space comprising 10 electrons in 8 orbitals (i.e. CAS(10,8)) was extended by four and eight extra-valence virtual π-orbitals which were treated at RASSCF level allowing maximally two simultaneous excitations. In the following we will refer to these active spaces as RAS(10,8|2,4) and RAS(10,8|2,8). The excited state computations were averaged over four states in the case of 2- and 4-thiouracil, while five states were taken into account in the case of 2,4-thiouracil. Additionally, the CAS(10,8) and RAS(10,8|2,4) were augmented by the lone pair(s) of the sulfur(s) to disentangle the effect of the nπ* transitions on the LA spectra, thus giving CAS(12,9) and RAS(12,9|2,4) for 2- and 4-thiouracil, CAS(14,10) and RAS(14,10|2,4) for 2,4-thiouracil. The excited state computations were averaged over six states in the case of 2- and 4-thiouracil, while 10 states were taken into account in the case of 2,4-thiouracil. All CASPT2 and RASPT2 computations were performed both in the single-state and multi-state flavor. All calculations were performed with the generally contracted basis set ANO-L adopting contractions 4s3p2d1f on sulfur, 3s2p1d on carbon, oxygen and nitrogen and 2s1p on hydrogen (denoted as ANO-L$_S$). Additionally, for 2- and 4-thiouracil the SA-6-CAS(12,9) computations were repeated with larger contractions 5s4p2d1f on sulfur, 4s3p2d1f on carbon and nitrogen and 3s2p1d on hydrogen (denoted as ANO-L$_L$) to test the dependency on the basis set size.

For each system the spectra were computed utilizing the transition energies and oscillation strengths computed at each snapshot. Thereby, each transition was convoluted with a Gaussian function with a FWHM of 0.1 eV. State-specific spectra were generated by grouping transitions according to the leading configuration of the ES wave function (i.e. $n_{S2}{\rightarrow}\pi^*_1$, $\pi_{S2}{\rightarrow}\pi^*_1$, $\pi_4{\rightarrow}\pi^*_1$, etc.) by analyzing the CI vectors. This was made possible as wave function analysis performed at the equilibrium geometry showed that each transition is dominated by a single configuration state function, the sole exception being the nπ*-states of 2,4-thiouracil ($n_{S24}{\rightarrow}\pi^*_1$, $n_{S24}{\rightarrow}\pi^*_2$, $n_{S42}{\rightarrow}\pi^*_1$, $n_{S42}{\rightarrow}\pi^*_2$) which were considered together. For the purpose of the following discussion we use the terms *configuration* and *state* interchangeably.

Experimental methods

4-Thiouracil (97% purity), 2-thiouracil (99% purity) and 2,4-dithiouracil (98% purity) were purchased from Sigma-Aldrich and used as received. A phosphate buffer saline (PBS) solution was prepared by dissolving 0.15 g of sodium dihydrogen phosphate and 0.27 g of sodium hydrogen phosphate in 200 ml of ultrapure water to obtain a pH 7.4 and a concentration of 16 mM. The 4-thiouracil, 2-thiouracil and 2,4-dithiouracil solutions were prepared in a concentration of 0.85, 1.5 and 1.8 mM respectively.

The steady-state absorption spectrum was measured in a 1-mm cuvette at room temperature using a V-570 spectrophotometer. It was background corrected by subtracting the solvent spectrum measured under the same experimental conditions. The solutions displayed an absorbance of ~1.8 OD at peak wavelengths of 330 nm (4-thiouracil), 275 nm (2-thiouracil) and 360 nm (2,4-dithiouracil).

Results

2-thiouracil

Table 1 summarizes the electronic structure of 2-thiouracil in gas phase and in aqueous solution. The lowest $\pi\pi^*$ transitions are the $\pi_{S2}\rightarrow\pi^*_1$, $\pi_{S2}\rightarrow\pi^*_2$ and the $\pi_4\rightarrow\pi^*_1$ absorbing at 4.14 eV (300 nm), 4.54 eV (273 nm) and 5.32 eV (233 nm), respectively, at the MS-CASPT2 level. Two dark $n\pi^*$ transitions involving the sulfur lone pair appear in the window above 230 nm, the $n_{S2}\rightarrow\pi^*_1$ at 4.04 eV (307 nm) and the $n_{S2}\rightarrow\pi^*_2$ at 4.94 eV (251 nm). We emphasize the small energy gap between the lowest $\pi\pi^*$ and $n\pi^*$ states (0.10 eV) as well as between the lowest two $\pi\pi^*$ states (0.40 eV). Notably, the two lowest bright states $\pi_{S2}\rightarrow\pi^*_1$ and $\pi_{S2}\rightarrow\pi^*_2$ interchanged intensities at the SS-CASPT2 level. Furthermore, we observe $\pi_{S2}\rightarrow\pi^*_2/n_{S2}\rightarrow\pi^*_2$ state mixing due to the energetic proximity of two states. As a consequence the $n_{S2}\rightarrow\pi^*_2$ gains oscillator strength. MS-CASPT2 resolves this issue. Gas phase calculations show a strong resemblance to the computations in water, the $n_{S2}\rightarrow\pi^*_1 /\pi_{S2}\rightarrow\pi^*_1$ gap growing to 0.4 eV. The solvatochromic shifts vary in the range of +0.20 eV (e.g. $n_{S2}\rightarrow\pi^*_1$) to -0.20 eV (e.g. $\pi_{S2}\rightarrow\pi^*_2$).

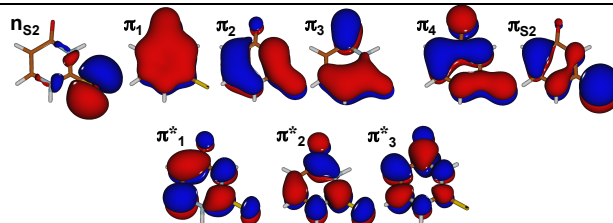|	In solution		Gas phase		orbitals					
				2-THIOURACIL						
Transition	Energy (eV)	f	Energy (eV)	f						
$n_{S2}\rightarrow\pi^*_1$	4.04(3.91)	0.01(0.00)	3.87(3.80)	0.00(0.00)						
$\pi_{S2}\rightarrow\pi^*_1$	4.14(4.15)	0.36(0.09)	4.25(4.25)	0.40(0.19)						
$\pi_{S2}\rightarrow\pi^*_2$	4.54(4.37)	0.15(0.28)	4.76(4.54)	0.12(0.36)						
$n_{S2}\rightarrow\pi^*_2$	4.94(4.55)	0.00(0.23)	5.06(4.94)	0.00(0.00)						
$\pi_4\rightarrow\pi^*_1$	5.32(5.15)	0.12(0.13)	5.15(4.99)	0.06(0.06)						

Table 1. Electronic spectra and active space orbitals of 2-thiouracil ground state minimum (see Computational Details section) at the MS- and SS-RASPT2/SA-6-RASSCF(12,9|2,4)/ANO-L$_S$ level in gas phase and in water. SS-RASPT2 energies and oscillator strengths are given in parenthesis.

Figure 2 shows the LA spectra of 2-thiouracil obtained by semi-classical sampling (see Computational Methods) at room temperature at the SS- and at the MS-CASPT2 levels of theory. A minimal full valence-π CAS(10,8) (red solid line), as well as two active spaces augmented by four and eight extra-valence virtual orbitals, denoted as RAS(10,8|2,4) (green solid line) and RAS(10,8|2,8) (blue solid line) are compared to the experimental spectrum (black solid line).
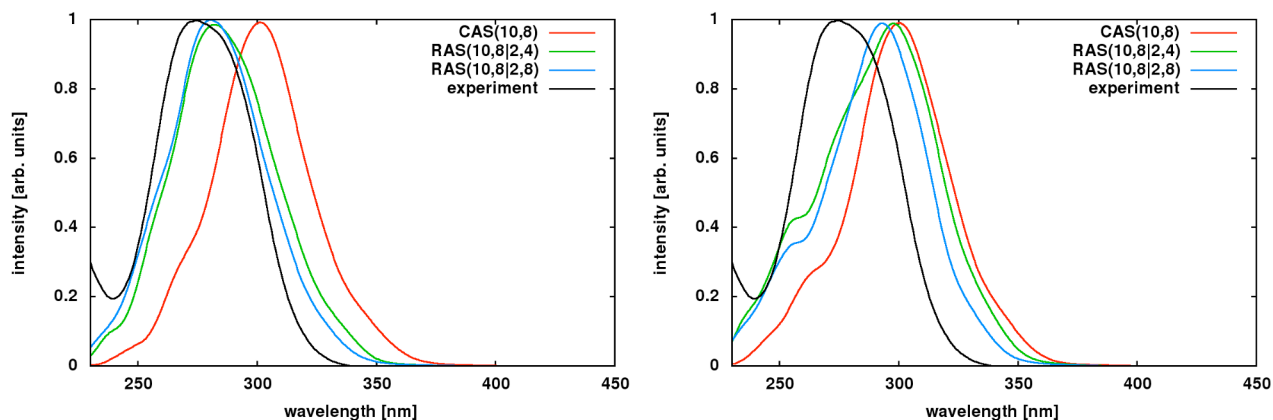

Figure 2. *Dependence of the LA spectrum of 2-thiouracil on the size of the active space size in the SS (left) and MS (right) flavors. Experimental spectrum in aqueous solution is given as reference.*

The LA spectrum of 2-thiouracil above 230 nm is characterized by a single intense band peaking at 270 nm with a faint shoulder at 290 nm. At SS-CASPT2 level we notice that the minimal CAS(10,8) red-shifts the absorption band by ca. 25 nm (~0.40 eV at this wavelength). Increasing the AS by four extra-valence virtual orbitals corrects the red-shift to give a very good agreement with the experiment (8 nm red-shift, ~0.10 eV at this wavelength). The addition of further orbitals affects the spectrum only marginally. The MS-CASPT2 spectra red-shifts the position of the absorption band by 25 nm (0.40 eV at this wavelength). Interestingly, increase of the AS size induces only a weak blue-shift of the band maximum from 300 nm with CAS(10,8) to 293 nm with RAS(10,8|2,8). The shift is accompanied by a broadening of the spectrum in the blue. Overall, SS-CASPT2 shows a much better agreement to the experimental data.
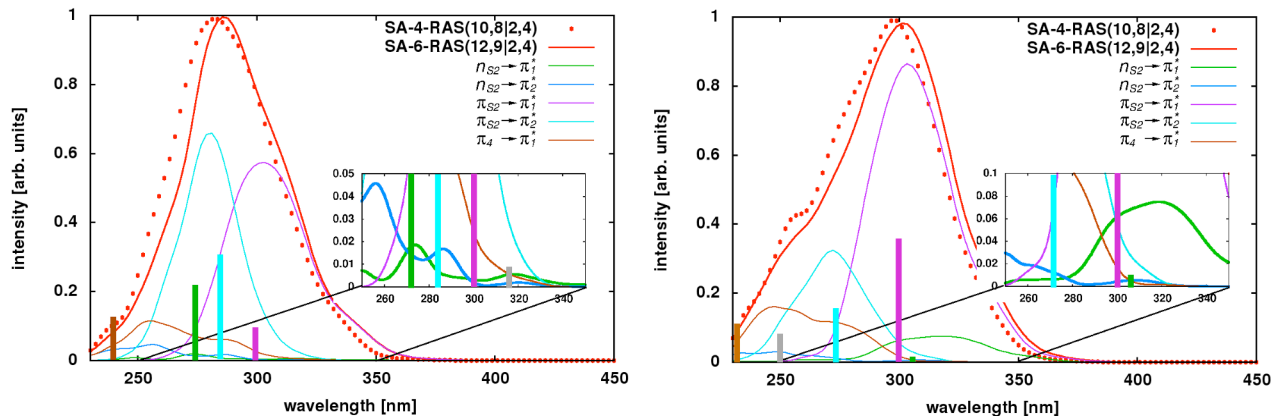

Figure 3. *State assignment in 2-thiouracil at the RASPT2/SA-6-RASSCF(12,9|2,4) level in the SS (left) and MS (right) flavors. Values at the equilibrium geometry (Table 1) are shown as stick spectra.*

Next we focus on the individual contributions to the spectrum. For the purpose we expand the RAS(10,8|2,4), shown above to produce spectra in good agreement with experiment at the SS-CASPT2 level, by including the sulfur lone pair thus obtaining a RAS(12,9|2,4) capable of resolving the relative positions of the bright and dark bands. As expected, **Figure 3** demonstrates that the inclusion of the lone pair does not affect the LA spectrum (compare solid (RAS(12,9|2,4)) and dotted (RAS(10,8|2,4)) red lines). Two configurations, $\pi_{S2} \to \pi^{*}_{1}$ (purple) and $\pi_{S2} \to \pi^{*}_{2}$ (cyan), are found to contribute to the absorption band, whereas SS-CASPT2 predicts them with equal oscillator strength (**Figure 3a**) while MS-CASPT2 favors the $\pi_{S2} \to \pi^{*}_{1}$ transition (**Figure 3b**). The presence of two bright transitions under the spectral envelope rationalizes the experimentally observed shoulder at 290 nm, which appears when the sulfur lone pair is included in the AS (red solid line in **Figure 3a**). The reason for the less satisfactory agreement of the MS-CASPT2 spectrum, over-emphasizing the $\pi_{S2} \to \pi^{*}_{1}$ contribution, can be rationalized by examining the CASSCF wave functions, SS-CASPT2 energies and MS-CASPT2 couplings as a function of the AS size. With CAS(10,8) the maxima of the $\pi_{S2} \to \pi^{*}_{1}$ and the $\pi_{S2} \to \pi^{*}_{2}$ bands are 0.55 eV apart at the CASSCF level (22 nm, **Figure S1**, left). Correspondingly, the wave functions of the two transitions mix only weakly and the $\pi_{S2} \to \pi^{*}_{2}$ transition appears less intense in both the SS-CASPT2 and the MS-CASPT2 spectra. Increase of the AS red-shifts all transitions already at CASSCF level due to the portion of correlation energy described at this level. Notably, the stabilizing effect is more pronounced for the $\pi_{S2} \to \pi^{*}_{2}$ state, thereby reducing the gap between the two lowest bright states to ca. 0.35 eV (17 nm, **Figure S1**). This causes a more pronounced wave function mixing, evident both by the increasing oscillator strength of the $\pi_{S2} \to \pi^{*}_{2}$ transition at the CASSCF level (**Figure S1**, left) and by the increasing values of the MS-CASPT2 couplings (from 0.001 on average
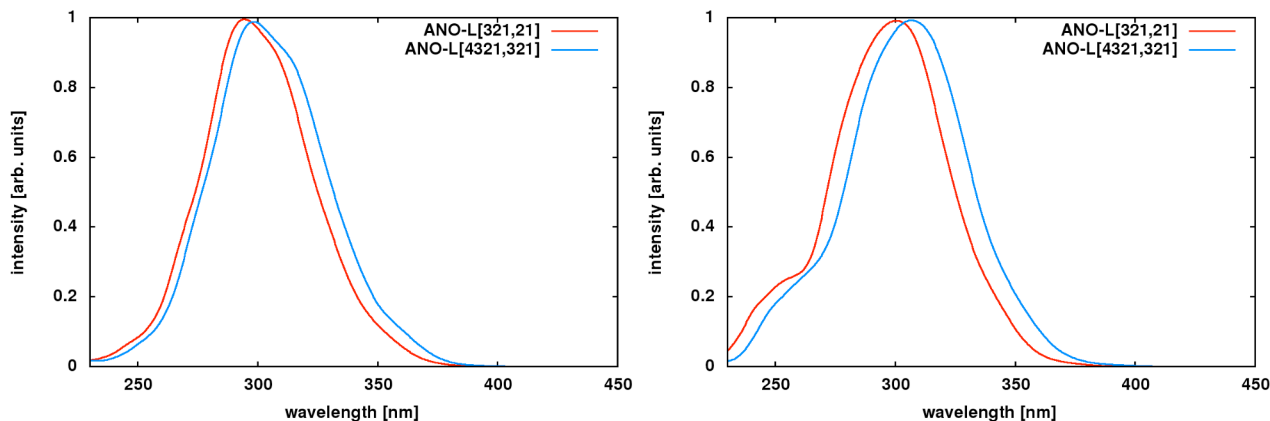
Figure 4. *Dependence of the LA spectrum of 2-thiouracil on the basis set size in the SS (left) and MS (right) flavors*

with CAS(10,8) to 0.004 with RAS(10,8|2,4), **Figure S2**, right). Due to the energetic proximity of $\pi_{S2}\rightarrow\pi^*_1$ and the $\pi_{S2}\rightarrow\pi^*_2$ after SS-CASPT2 correction (ca. 0.30 eV) the larger off-diagonal elements become the source of overestimation of the level of wave function un-mixing and energy splitting. We note that in both SS- and MS-CASPT2 spectra the $\pi_{S2}\rightarrow\pi^*_1$ band is predicted to peak at ~300 nm. It is the oscillator strength carried predominantly by the $\pi_{S2}\rightarrow\pi^*_1$ transition at the MS-CASPT2 level which is responsible for the overall red-shift of the absorption band.

The lowest $n\pi^*$ transition $n_{S2}\rightarrow\pi^*_1$ in the system is found to overlap energetically with the $\pi_{S2}\rightarrow\pi^*_1$ transition (more clearly seen in the MS-CASPT2 spectrum, **Figure 3b**), in agreement with the static computations predicting a gap of solely 0.10 eV (**Table 1**). This observation is also in agreement with recently published molecular dynamics simulations carried on 2-thiouracil in gas-phase, reporting an ultrafast, sub-100 fs decay to the $n\pi^*$ after excitation of the lowest bright state [64].

Finally, **Figure 4** demonstrates the effect of the basis set size on the total spectrum. Despite the twofold increase of the basis functions (300 with ANO-L$_L$ vs. 148 with ANO-L$_S$) we observe only a ca. 4 nm (SS-CASPT2) to 7 nm (MS-CASPT2) red-shift of the entire band which does not affect the line shape.

4-thiouracil

Table 2 summarizes the electronic structure of 4-thiouracil in gas phase and in aqueous solution. The lowest $\pi\pi^*$ transitions are the $\pi_{S4}\rightarrow\pi^*_1$, $\pi_4\rightarrow\pi^*_1$ and the $\pi_{S4}\rightarrow\pi^*_2$ absorbing at 3.83 eV (323 nm), 4.50 eV (275 nm) and 5.15 eV (241 nm) at MS-CASPT2 level, respectively. Two essentially dark $n\pi^*$ transitions involving the sulfur lone pair appear in the window above 230 nm, the $n_{S4}\rightarrow\pi^*_1$ at 2.99 eV (415 nm) and the $n_{S4}\rightarrow\pi^*_2$ at 5.03 eV (247 nm). We note the interchanged energetic order of $\pi_4\rightarrow\pi^*_1$ and $\pi_{S4}\rightarrow\pi^*_2$ with respect to 2-thiouracil, as well as the larger $\pi_{S4}\rightarrow\pi^*_1/n_{S4}\rightarrow\pi^*_1$ gap of 0.80 eV. SS- and MS-CASPT2 give similar results, $\pi_{S4}\rightarrow\pi^*_2$ the sole outlier which exhibits a 0.28 eV blue-shift at MS-CASPT2 level. Gas phase calculations show a strong resemblance to the computations in water, the $n_{S4}\rightarrow\pi^*_1/\pi_{S4}\rightarrow\pi^*_1$ gap growing to 1.00 eV. The solvatochromic shifts vary in the range of +0.10 eV (e.g. $n_{S4}\rightarrow\pi^*_1$) to -0.15 eV (e.g. $n_{S4}\rightarrow\pi^*_2$).

Figure 5 shows the LA spectra of 4-thiouracil obtained by semi-classical sampling (see Computational Methods) at room temperature at the SS- and at the MS-CASPT2 levels of theory. A minimal full valence-π CAS(10,8) (red solid line), as well as two active spaces augmented by four and eight extra-valence virtual orbitals, denoted as RAS(10,8|2,4) (green solid line) and RAS(10,8|2,8) (blue solid line) are compared to the experimental spectrum (black solid line).

	In solution		Gas phase		orbitals
					4-THIOURACIL
$n_{S4} \rightarrow \pi^*_1$	2.99(2.89)	0.00(0.00)	2.91(2.81)	0.00(0.00)	
$\pi_{S4} \rightarrow \pi^*_1$	3.83(3.73)	0.62(0.52)	3.89(3.91)	0.65(0.45)	
$\pi_4 \rightarrow \pi^*_1$	4.50(4.54)	0.04(0.03)	4.49(4.47)	0.03(0.11)	
$n_{S4} \rightarrow \pi^*_2$	5.03(4.89)	0.02(0.00)	5.18(5.08)	0.00(0.00)	
$\pi_{S4} \rightarrow \pi^*_2$	5.15(4.87)	0.04(0.15)	5.21(4.82)	0.07(0.15)	

Table 2. Electronic spectra and active space orbitals of 4-thiouracil ground state minimum (see Computational Details section) at the MS- and SS-RASPT2/SA-6-RASSCF(12,9|2,4)/ANO-L$_S$ level in gas phase and in water. SS-RASPT2 energies and oscillator strengths are given in parenthesis.

The LA spectrum of 4-thiouracil above 230 nm is characterized by an intense band peaking at 330 nm and a weak band around 250 nm. As in the case of 2-thiouracil, at the SS-CASPT2 level we notice that the minimal CAS(10,8) red-shifts the absorption band by ca. 20 nm (~0.20 eV at this wavelength, **Figure 5a**). The addition of four and eight virtual orbitals corrects the red-shift to give an excellent agreement with the experimental absorption maximum for RAS(10,8|2,8) (**Figure 5a**). We note that the simulated spectrum decays smoother at long wavelengths compared to the experiment. On the other hand, at MS-CASPT2 level already the minimal CAS(10,8) appears red-shifted by only 7 nm. Increase of the AS size leads to a near-quantitative agreement with the experiment. Notably, the MS-CASPT2 spectra tend to underestimate the intensity of the 250 nm absorption band, while SS-CASPT2 reproduces accurately the high-energy region.

Next we focus on the individual contributions to the spectrum. As in the case of 2-thiouracil we expand the RAS(10,8|2,4), shown above to produce spectra in excellent agreement with the experimental data, by including the sulfur lone pair thus obtaining a RAS(12,9|2,4) capable of resolving the relative positions of the bright and dark bands. **Figure 6** demonstrates that the inclusion of the lone pair introduces a 5 nm red-shift in the SS-CASPT2 spectrum (compare solid (RAS(12,9|2,4)) and dotted (RAS(10,8|2,4)) red lines), not present in the MS-CASPT2 spectra. The lowest absorption band emerges exclusively due to the $\pi_{S4} \rightarrow \pi^*_1$ transition (purple). The weak band around 250 nm has contributions from the $\pi_4 \rightarrow \pi^*_1$ (brown) and $\pi_{S4} \rightarrow \pi^*_2$ (cyan) transitions, whereas MS-CASPT2 predicts higher oscillator strength for the former, while SS-CASPT2 favors the latter. Unlike 2-thiouracil, the $\pi_{S4} \rightarrow \pi^*_1$ state in 4-thiouracil is well separated from the higher lying transitions, both at the CASSCF (0.80 eV, **Figure S2**) and at the SS-CASPT2 (~0.75 eV, **Figure 6a**) levels. Consequently, SS- and MS-
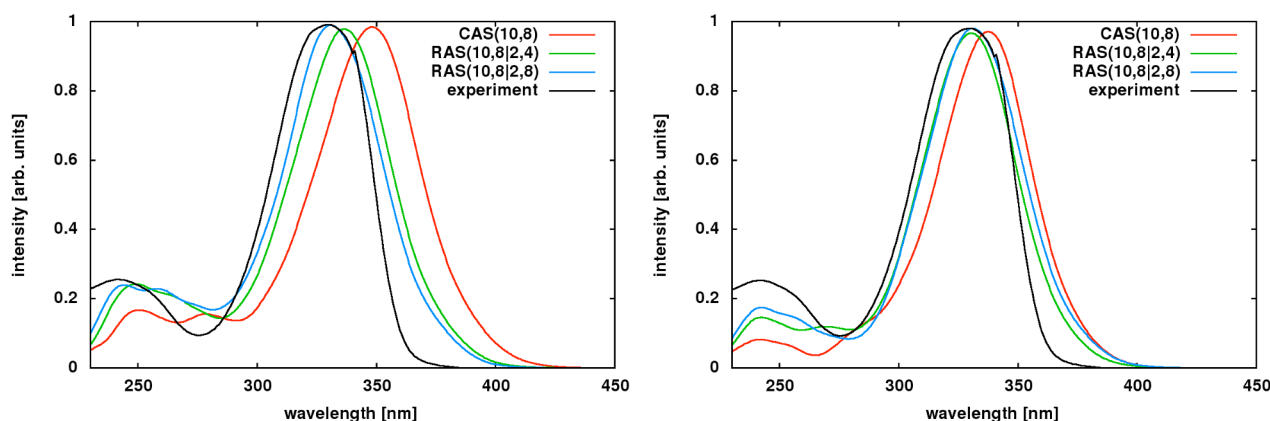

Figure 5. *Dependence of the LA spectrum of 4-thiouracil on the size of the active space size in the SS (left) and MS (right) flavors. Experimental spectrum in aqueous solution given as reference.*
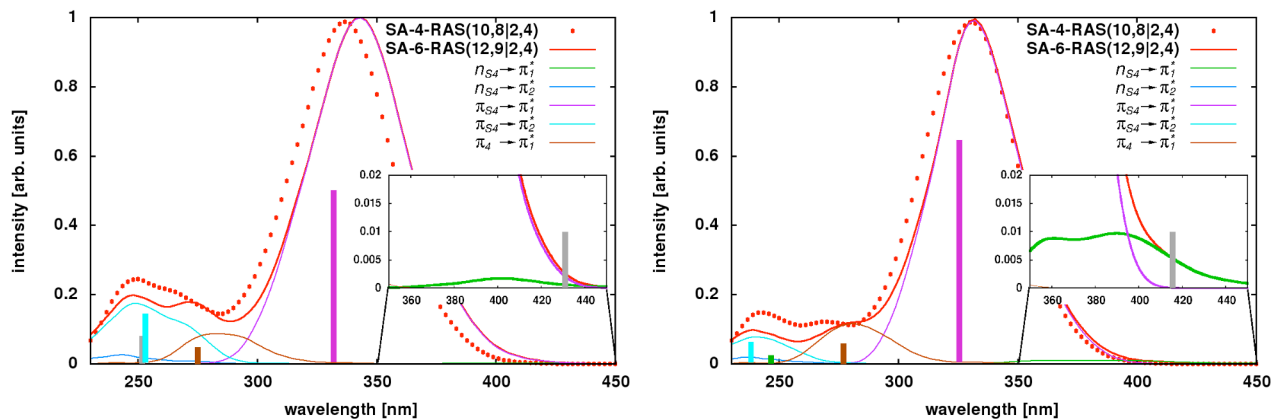
Figure 6. *State assignment in 4-thiouracil at the RASPT2/SA-6-RASSCF(12,9|2,4) level in the SS (left) and MS (right) flavors. Values at the equilibrium geometry (Table 1) are shown as stick spectra.*

CASPT2 are equally competent in describing the 330 nm band. Regarding the high energy region we observe a gradual decrease of the magnitude of the off-diagonal elements of the MS-Hamiltonian from 0.005 a.u. to 0.002 a.u. (**Figure S3**) with the increase of the AS.

The lowest $n\pi^*$ transition $n_{S4}{\rightarrow}\pi^*_1$ in the system is found to overlap with red part of the $\pi_{S4}{\rightarrow}\pi^*_1$ band (**Figure 6**), despite the 0.80 eV energy gap between both transitions at the equilibrium geometry (**Table 2**). The second $n\pi^*$ transition $n_{S4}{\rightarrow}\pi^*_2$ falls under the envelope of the second absorption band around 240 nm.

Finally, **Figure 7** demonstrates the effect of the basis set size on the total spectrum. While the SS-CASPT2 spectrum is unaffected by the basis set increase, the MS-CASPT2 spectrum exhibits a 7 nm red-shift similar to what was observed in the case of 2-thiouracil.
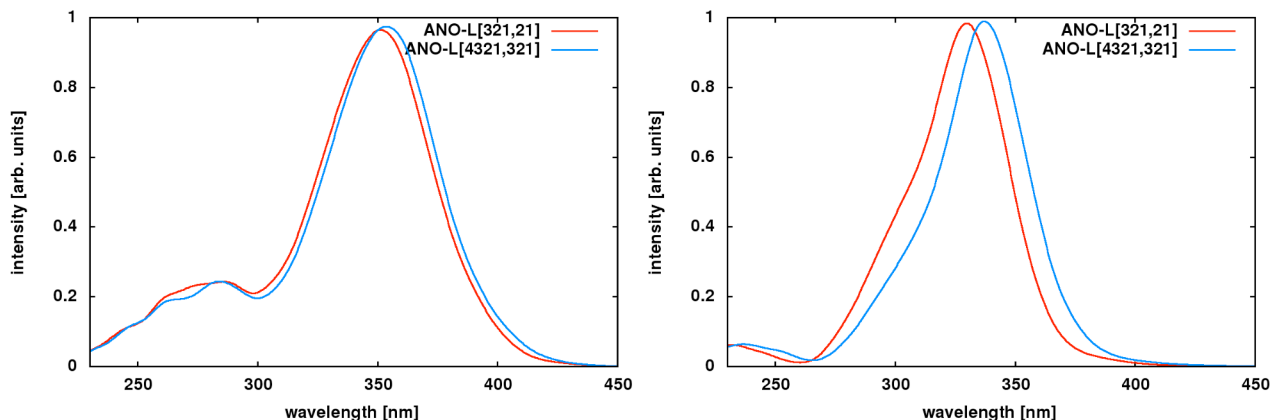

Figure 7. *Dependence of the LA spectrum of 4-thiouracil on the basis set size in the SS (left) and MS (right) flavors*

2,4-thiouracil

Table 3 summarizes the electronic structure of 2,4-thiouracil in gas phase and in aqueous solution. The lowest $\pi\pi^*$ transitions are the $\pi_{S42}\rightarrow\pi^*_1$, followed by the $\pi_{S24}\rightarrow\pi^*_1$, the $\pi_{S42}\rightarrow\pi^*_2$ and the $\pi_{S24}\rightarrow\pi^*_2$ absorbing at 3.76 eV (330 nm), 4.22 eV (294 nm), 4.56 eV (272 nm) and 4.85 eV (256 nm) at MS-CASPT2 level, respectively. Four dark $n\pi^*$ transitions involving the sulfur lone pairs appear in the window above 230 nm, the $n_{S42}\rightarrow\pi^*_1$ at 3.51 eV (253 nm), the $n_{S42}\rightarrow\pi^*_2$ at 3.90 eV (318 nm), the $n_{S24}\rightarrow\pi^*_1$ at 4.32 eV (287 nm) and the $n_{S24}\rightarrow\pi^*_2$ at 5.08 eV (244 nm). We note that SS- and MS-CASPT2 give similar results, whereas MS-CASPT2 tends to blue-shift all transitions by up to 0.20 eV. Gas phase calculations show a strong resemblance to the computations in water, $\pi\pi^*$ transitions in solution roughly matching with the gas phase values. In the case of $n\pi^*$ transitions we observe a significant 0.5 eV solvatochromic blue-shift for the lowest ($n_{S42}\rightarrow\pi^*_1$, shift from 3.00 eV to 3.50 eV) and highest ($n_{S24}\rightarrow\pi^*_2$, shift from 4.60 eV to 5.10 eV) bands. This leads to pronounced reduction of the gap between the lowest $\pi\pi^*$ and $n\pi^*$ states upon solvation from 0.90 eV to 0.25 eV. Notably, the triad of states $\pi_{S42}\rightarrow\pi^*_1$, $n_{S42}\rightarrow\pi^*_2$ and $\pi_{S24}\rightarrow\pi^*_1$ appear close in energy in gas-phase at SS-CASPT2 level (between 3.84 eV and 3.98 eV). Same holds for the triad of states $\pi_{S42}\rightarrow\pi^*_2$, $n_{S24}\rightarrow\pi^*_2$ and $n_{S24}\rightarrow\pi^*_2$ which absorb between 4.34 eV and 4.47 eV. We attribute the pronounced deviations in the magnitudes of the oscillator strengths between SS- and MS-CASPT2 in gas-phase (as well as between gas-phase and solution) to the strong wave function mixing, originating from the higher density and energetic proximity of the electronic states.

	In solution		Gas phase		orbitals
			2,4-THIOURACIL		
$n_{S42}\rightarrow\pi^*_1$	3.51(3.42)	0.02(0.00)	3.01(2.91)	0.00(0.00)	
$\pi_{S42}\rightarrow\pi^*_1$	3.76(3.63)	0.18(0.18)	3.89(3.95)	0.15(0.40)	
$n_{S42}\rightarrow\pi^*_2$	3.90(3.80)	0.00(0.00)	3.94(3.84)	0.00(0.00)	
$\pi_{S24}\rightarrow\pi^*_1$	4.22(4.15)	0.36(0.37)	4.26(3.98)	0.21(0.02)	
$n_{S24}\rightarrow\pi^*_1$	4.32(4.20)	0.00(0.00)	4.45(4.34)	0.00(0.00)	
$\pi_{S42}\rightarrow\pi^*_2$	4.56(4.46)	0.36(0.40)	4.61(4.42)	0.10(0.23)	
$\pi_{S24}\rightarrow\pi^*_2$	4.85(4.65)	0.23(0.16)	4.74(4.59)	0.64(0.45)	
$n_{S24}\rightarrow\pi^*_2$	5.08(4.94)	0.01(0.03)	4.59(4.47)	0.08(0.00)	

Table 3. Electronic spectra and active space orbitals of 2,4-thiouracil ground state minimum (see Computational Details section) at the MS- and SS-RASPT2/SA-9-RASSCF(2,2|14,10|2,2)/ANO-L$_S$ level in gas phase and in water. SS-RASPT2 energies and oscillator strengths are given in parenthesis.

Figure 8 shows that LA spectra of 2,4-thiouracil obtained by semi-classical sampling (see Computational Methods) at room temperature at the SS- and at the MS-CASPT2 levels of theory. A minimal full valence-π CAS(10,8) (red solid line), as well as two active spaces augmented with four and eight extra-valence virtual orbitals, denoted as RAS(10,8|2,4) (green solid line) and RAS(10,8|2,8) (blue solid line) are compared to the experimental spectrum (black solid line).

The LA spectrum of 2,4-thiouracil above 230 nm is characterized by a complex structure with two bands, a less intense one around 360 nm and a more intense one around 275 nm with a clearly distinguishable shoulder at 260 nm. Both SS- and MS-CASPT2 have difficulties in reproducing the two-peak spectral line shape. In particular, the SS-CASPT2 simulations red-shift strongly the 270 nm band, the use of a minimal CAS(10,8) resulting thereby in a 50 nm shift (0.70 eV at this wavelength). At the same time the 360 nm band appears weaker in intensity, yet it is clearly visible as a shoulder in the red part of the spectrum. The addition of virtual orbitals partially recovers the red-shift of the 270 nm band. Still, we observe a more pronounced mismatch with the experimental data compared to 2- and 4-thiouracil. Notably, despite the red-shift, SS-CASPT2 accurately reproduces the line shape of the
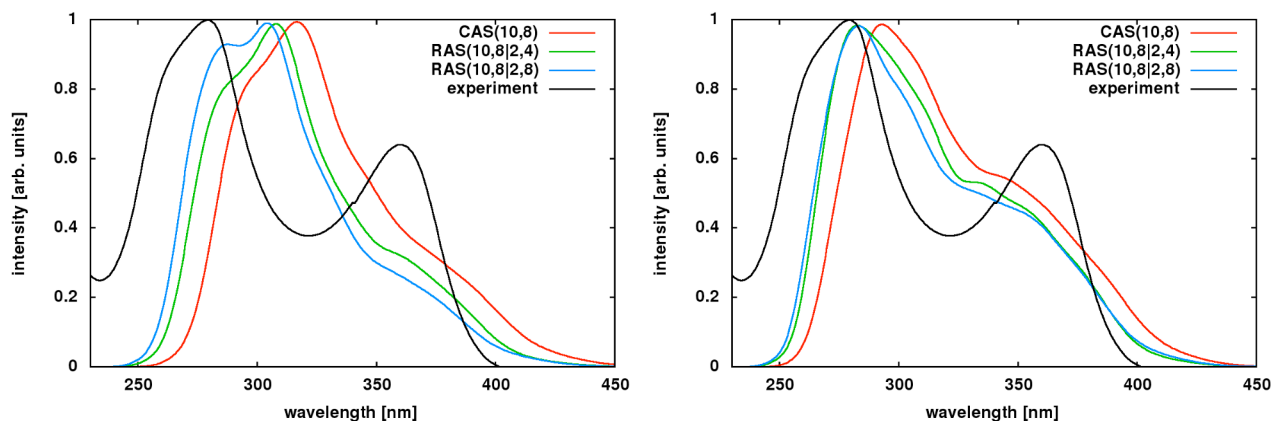
Figure 8. *Dependence of the LA spectrum of 2,4-thiouracil on the size of the active space size in the SS (left) and MS (right) flavors. Experimental spectrum in aqueous solution given as reference.*

270 nm band even with the minimal CAS(10,8), clearly resolving the shoulder in the blue. The MS-CASPT2 spectra reproduce better the position of the problematic 270 nm peak, its maximum being only 10 nm red-shifted with RAS(10,8|2,4) and RAS(10,8|2,8). In addition, the 360 nm band appears more intense in MS-CASPT2 computations, thus improving the agreement with the experiment.

Next we focus on the individual contributions to the LA spectrum. As in the case of 2- and 4-thiouracil we expand the RAS(10,8|2,4), this time including two sulfur lone pairs in order to obtain a RAS(14,10|2,4) capable of resolving the relative positions of the bright and dark states. **Figure 9** demonstrates that the inclusion of the lone pair introduces a 10 nm red-shift in the SS-CASPT2 spectrum (compare solid (RAS(12,9|2,4)) and dotted (RAS(10,8|2,4)) red lines) accompanied by a weak increase of the intensity of the shoulder at 360 nm. At the MS-CASPT2 level a red-shift is observed only for the high energy region. The lowest absorption band emerges due to the $\pi_{S42} \rightarrow \pi^*_1$ transition (purple). The intense band at 270 nm instead is built by contributions from $\pi_{S24} \rightarrow \pi^*_2$ (ocher, highest oscillator strength among all $\pi\pi^*$ states), $\pi_{S24} \rightarrow \pi^*_1$ (brown) and $\pi_{S42} \rightarrow \pi^*_2$ (cyan). It is, thus, clear that the precise spectral line shape is a function of the interaction of these three near-degenerate states. We encounter MS-CASPT2 off-diagonal couplings of up to 0.005 a.u. between these states, i.e. of the same magnitude as the energy gaps (**Figure S4**).
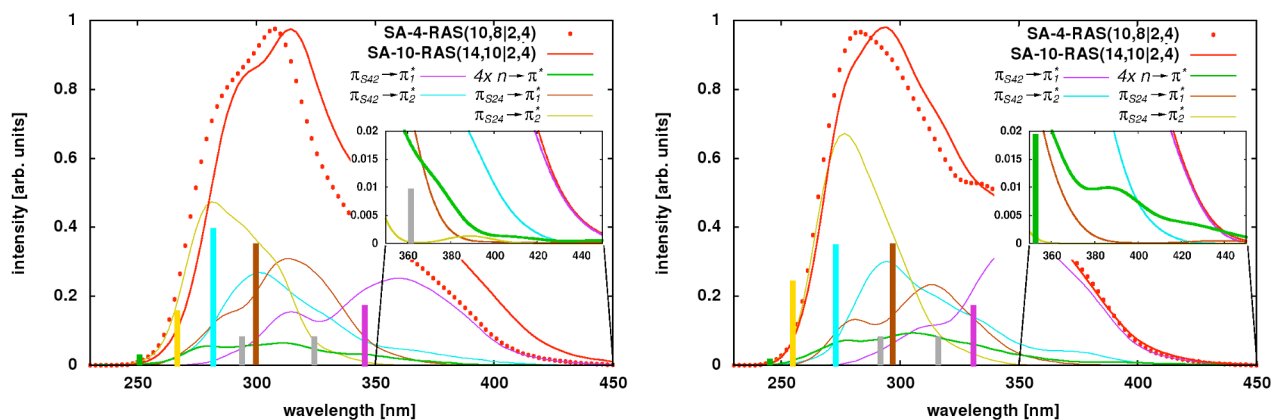

Figure 9. *State assignment in 2,4-thiouracil at the RASPT2/SA-6-RASSCF(12,9|2,4) level in the SS (left) and MS (right) flavors. Values at the equilibrium geometry (Table 1) are shown as stick spectra.*

The inclusion of the sulfur lone pairs in the AS introduces four $n\pi^*$ states in the range above 250 nm, arising through linear combination of transitions from each of the two lone pairs in either π^*_1 or π^*_2. The $n\pi^*$ transitions are found to cover the entire spectrum from 250 nm to 400 nm (green solid line in **Figure 9**). It comes to our attention that the $n\pi^*$ states bear a weak oscillator strength in the case of 2,4-thiouracil.

Discussion and conclusion

In the previous section we assessed the performance of the RASPT2/RASSCF protocol in simulating LA spectra for three thio-functionalized uracil derivatives as a function of several parameters: active space (AS) size and composition, CASPT2 flavor and basis set. Through a QM/MM scheme coupled to a room-temperature Wigner sampling we are in the position to simulate condensed phase spectra, permitting us to compare our results against experimental data recorded in aqueous solution. Vertical transitions at equilibrium do not correspond to the maximum of the LA band and appear blue-shifted[82]. Sampling is important to obtain the correct band maxima. This is particularly well exemplified in the spectrum of 2,4-thiouracil where the vertical transitions (shown as stick spectra in Figs. 9) exhibit a 0.2-0.3 eV blue-shift for all $\pi\pi^*$ states. The benchmarking demonstrates that full-π valence AS (referred to as minimal AS) tend to red-shift the absorption band maxima by 20 nm at the SS-CASPT2 level (**Figures 2a, 5a, 8a**). Limited active spaces leave extensive part of the correlation computation over to the CASPT2 routine. This could lead to overestimation of the contribution when the energy gap between the reference (zeroth-order) CASSCF wave function and the correction terms of the first-order wave function is too small [33-35]. Increasing the AS by including extra-valence virtual orbitals rectifies the problem (**Figures 2a, 5a, 8a**) by increasing the gap (and thus the value of the denominator in the correction terms), thereby effectively reducing the perturbative contribution to the correlation. At the same time increasing portion of the correlation is treated variationally at the CASSCF level. Upon convergence, the transition energies (and all other state properties) should become independent on whether further extra-valence orbitals are treated variationally or perturbationally as seen when comparing the spectra obtained with four and eight extra-valence orbitals. At the MS-CASPT2 level the red-shift seems less pronounced as exemplified by 4- and 2,4-thiouracil (**Figures 5b and 8b**). Yet, MS-CASPT2 on top of minimal AS suffers from over-estimation of the degree of wavefunction un-mixing due to large off-diagonal elements in the MS-Hamiltonian between electronic states which end up energetically close at the CASSCF level and, as a consequence, mix strongly. This is evident in the case of 2-thiouracil (**Figure 2b**) in which the MS-CASPT2 performs poorly even with extra-valence orbitals in the AS due to the mixing of the $\pi_{S2}\rightarrow\pi^*_1$ and $\pi_{S2}\rightarrow\pi^*_2$ states, their MS-CASPT2 coupling reaching values of 0.006 a.u. (**Figure S1b**). In fact, we remind that in the seminal paper on MS-CASPT2 a value of 0.002 a.u. has been suggested as an upper limit for the off-diagonal elements of the MS-Hamiltonian [83]. Generally, increasing the AS reduces the size of the off-diagonal elements as demonstrated for 4-thiouracil (**Figure 5b**, **Figure S3**) but in certain cases one can observe also the opposite situation as seen for 2-thiouracil (**Figure 2b**, **Figure S1b**). Thus, the multi-state formulation of CASPT2 is not always superior to the single-state formulation and one should critically examine the magnitude of the MS-CASPT2 couplings.

Overall, SS-CASPT2 with four extra-valence orbitals in the AS is found to reproduce better the spectral lines shapes in all three systems. In particular, it successfully reproduces the position and shoulder of the 270 nm band in 2-thiouracil, the intensity of the 250 nm band in 4-thiouracil and the shoulder of the 275 nm band in 2,4-thiouracil. MS-CASPT2 underperforms in the case of 2-thiouracil and is equally adequate for the description of the intense 330 nm band of 4-thiouracil, yet underestimating the intensity of the high energy 250 nm band. It reaches a somewhat better agreement with the experiment in the most complex case of 2,4-thiouracil where it reproduces the intensity of the 360 nm peak.

2,4-thiouracil exhibits the most complex electronic structure of all three systems with four bright $\pi\pi^*$ states and four nπ^* states in close energetic proximity. Both SS- and MS-CASPT2 do not succeed in reproducing the line shape of its LA spectrum with satisfactory agreement. 2,4-thiouracil could become an intriguing test case for assessing the potential of further QM methods and QM/MM protocols to improve the agreement. Recent CASPT2 extensions such as the extended multi-state (XMS) variant, QM/MM improvements such as inclusion of environment polarization or explicit consideration of the first layer of solvent molecules in the QM region could be benchmarked.

Finally, it was demonstrated on the example of 2- and 4-thiouracil that going beyond double-ζ basis sets with one set of polarization functions introduces a minor stabilization effect without influencing the spectral line shape.

In conclusion, we demonstrate that, while the computational accuracy of the RASSCF/RASPT2 approach can be rigorously improved by increasing the active space, the approach comes with a number of fine-tuning knobs which require expert knowledge. The choice of proper parameters has to be made according to the problem of interest through rigorous benchmarking. Mismatch between experiment and simulation, often attributed to the weaknesses of QM/MM models with electrostatic embedding (i.e. missing feedback from the solute on the solvent), should be also discussed in the perspective of insufficient accuracy of the applied QM methodology.

Acknowledgments

References

1. H. Lin, D. G. Truhlar, QM/MM: what have we learned, where are we, and where do we go from here?, Theo. Chem. Acc., 117 (2007) 185, doi: 10.1007/s00214-006-0143-z.
2. E. Brunk, U. Rothlisberger, Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States, Chem. Rev., 115 (2015), 6217-6263, doi: 10.1021/cr500628b.
3. O. Weingart, Combined Quantum and Molecular Mechanics (QM/MM) Approaches to Simulate Ultrafast Photodynamics in Biological Systems, Cur. Org. Chem., 21 (2017) 586-601.
4. M. De Vett, M. F. S. J. Menger, J. J. Nogueira, L. González, Solvent Effects on Electronically Excited States: QM/Continuum Versus QM/Explicit Models, J. Phys. Chem. B, 122 (2018) 2975-2984, doi: 10.1021/acs.jpcb.7b12560
5. U. N. Morzan, D. J. Alonso de Armiño, N. O. Foglia, F. Ramírez, M. C. González Lebrero, D. A. Scherlis, D. A. Estrin, Spectroscopy in Complex Environments from QM-MM Simulations, Chem. Rev., Article ASAP, doi: 10.1021/acs.chemrev.8b00026
6. F. Maseras, K.Morokuma, IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states, J. Comp. Chem., 16 (1995) 1170-1179.
7. D. Bakowies, W. Thiel, Semiempirical treatment of electrostatic potentials and partial charges in combined quantum mechanical and molecular mechanical approaches, J. Comp. Chem., 17 (1996) 87-108.
8. R. D. J. Froese, D. G. Musaev, K. Morokuma, Theoretical Study of Substituent Effects in the Diimine−M(II) Catalyzed Ethylene Polymerization Reaction Using the IMOMM Method, J. Am. Chem. Soc., 120 (1998) 1581-1587, doi: 10.1021/ja9728334

9. H. M. Senn, W. Thiel, QM/MM Methods for Biological Systems, Top. Curr. Chem., 268 (2007) 173.

10. P. Altoè, M. Stenta, A. Bottoni, M. Garavelli, A Tunable QM/MM Approach to Chemical Reactivity, Structure and Physico-Chemical Properties Prediction, Theor. Chem. Acc., 118 (2007) 219.

11. K. F. Hall, T. Vreven, M. J. Frisch, M. J. Bearpark, Three-Layer ONIOM Studies of the Dark State of Rhodopsin: The Protonation State of Glu181, J. Mol. Biol., 383 (2008) 106.

12. J. M. Olsen, K. Aidas, J. Kongsted, Excited States in Solution through Polarizable Embedding, J. Chem. Theory Comput., 6 (2010) 3721.

13. K. Sneskov, T. Schwabe, J. Kongsted, O. Christiansen, The Polarizable Embedding Coupled Cluster Method, J. Chem. Phys., 134 (2011) 104108.

14. D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal, B. Mennucci, A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations, J. Chem. Theory Comput., 12 (2016) 3654-3661, doi: 10.1021/acs.jctc.6b00385

15. D. Loco, S. Jurinovich, L. Cupellini, M. F. S. J. Menger, B. Mennucci, The modeling of the absorption lineshape for embedded molecules through a polarizable QM/MM approach, Photochem. & Photobiol. Sci., Article ASAP (2018) doi: 10.1039/C8PP00033F.

16. B. O. Roos, P. R. Taylor, P. E. M. Siegbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, Chem. Phys., 48 (1980) 157-173.

17. B. O. Roos, in Advances in Chemical Physics; Ab Initio Methods in Quantum Chemistry-II, edited by K. P. Lawley (Wiley, Chichester, England, 1987), Chap. 69, pp. 399–445.

18. P. A. Malmqvist, A. Rendell, B. O. Roos, The restricted active space self-consistent-field method, implemented with a split graph unitary group approach, J. Phys. Chem., 94 (1990) 5477-5482.

19. K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, K. Wolinski, Second-order perturbation theory with a CASSCF reference function, J. Phys. Chem., 94 (1990) 5483-5488.

20. P. A. Malmqvist, K. Pierloot, C. J. Cramer, L. Gagliardi, The restricted active space followed by second-order perturbation theory method: theory and application to the study of CuO_2 and Cu_2O_2 systems, J. Chem. Phys., 128 (2008) 204109.

21. V. Sauri, L. Serrano-Andrés,, A. R. M. Shahi, L. Gagliardi, S. Vancoillie, K. Pierloot, Multiconfigurational Second-Order Perturbation Theory Restricted Active Space (RASPT2) Method for Electronic Excited States: A Benchmark Study, J. Chem. Theory Comput., 7 (2011) 153.

22. J. Finley, P.-Å. Malmqvist, B. O. Roos, L. Serrano-Andrés, The multi-state CASPT2 method, Chem. Phys. Lett., 288 (1998), 299-306.

23. T. Shiozaki, W. Gyorffy, P. Celani, H.-J. Werner, The extended multi-state CASPT2 method: Energy and nuclear gradients, J. Chem. Phys., 135 (2011) 081106.

24. B. O. Roos, K. Andersson, Multiconfigurational perturbation theory with level shift - the Cr2 potential revisited, Chem. Phys. Lett., 245 (1995) 215-223.

25. B. O. Roos, K. Andersson, M. P. Fülscher, L. Serrano-Andrés, K. Pierloot, M. Merchán, V. Molina, Applications of level shift corrected perturbation theory in electronic spectroscopy, J. Mol. Struct. Theochem., 388 (1996) 257-276.

26. N. Forsberg, P.-Å. Malmqvist, Multiconfiguration perturbation theory with imaginary level shift, Chem. Phys. Lett., 274 (1997) 196-204.

27. G. Ghigo, B. O. Roos, P.-Å. Malmqvist, A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2), Chem. Phys. Lett., 396 (2004) 142-149.

28. J. Patrick Zobel, J. J. Nogueira, L. González, The IPEA dilemma in CASPT2, Chem Sci., 8

(2017) 1482-1499.

29. S. Arslancan, L. Martínez-Fernández, I. Corral, Photophysics and photochemistry of canonical nucleobases' thioanalogs: From quantum mechanical studies to time resolved experiments, Molecules, 22 (2017) 998, doi:10.3390/molecules22060998.

30. M. Pollum, L. Martínez-Fernández, C. E. Crespo-Hernández, Photochemistry of Nucleic Acid Bases and Their Thio- and Aza-Analogues in Solution, In Top. Curr. Chem., Springer, 2014, 245–327, ISBN 978-3-540-73346-1.

31. P. O. Widmark, P. A. Malmqvist, B. O. Roos, Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions, Theor. Chim. Acta, 77 (1990) 291-306.

32. J. Lorentzon, P. A: Malmqvist, M. Fulscher, B. O. Roos, A CASPT2 Study of the Valence and Lowest Rydberg Electronic States of Benzene and Phenol. Theor. Chim. Acta 91 (1995) 91-108.

33. A. Nenov, A. Giussani, J. Segarra-Martí, V. K. Jaiswal, I. Rivalta, G. Cerullo, S. Mukamel, M. Garavelli, Modeling the high-energy electronic state manifold of adenine: Calibration for nonlinear electronic spectroscopy, J. Chem. Phys., 142 (2015) 212443.

34. A. Nenov, I. Rivalta, S. Mukamel, M. Garavelli, Bidimensional electronic spectroscopy on indole in gas phase and in water from first principles, Comput. and Theor. Chem. 1040 (2014) 295-303.

35. A. Nenov, S. Mukamel, M. Garavelli, I. Rivalta, Two-dimensional electronic spectroscopy of benzene, phenol, and their dimer: an efficient first-principles simulation protocol, J. Chem. Theory Comput. 11 (2015) 3755-3771.

36. A. Nenov, A. Giussani, B. P. Fingerhut, I. Rivalta, E. Dumont, S. Mukamel, M. Garavelli, Spectral Lineshapes in Nonlinear Electronic Spectroscopy. Phys. Chem. Chem. Phys., 17 (2015) 30925-30936.

37. A. Nenov, R. Borrego-Varillas, A. Oriana, L. Ganzer, F. Segatta, I. Conti, J. Segarra-Marti, J. Omachi, M. Dapor, S. Taioli, C. Manzoni, S. Mukamel, G. Cerullo, M. Garavelli, UV-Light-Induced Vibrational Coherences: The Key to Understand Kasha Rule Violation in trans-Azobenzene, J. Phys. Chem. Lett. 9 (2018) 1534-1541

38. V. Barone, M. Cossi, Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model J. Phys. Chem. A, 102 (1998) 1995-2001.

39. M. Cossi, V. Barone, Solvent Effect on Vertical Electronic Transitions by the Polarizable Continuum Model J. Chem. Phys., 112 (2000) 2427-2435.

40. M. Cossi, N. Rega, G. Scalmani, V. Barone, Polarizable Dielectric Model of Solvation with Inclusion of Charge Penetration Effects J. Chem. Phys., 114 (2001) 5691-5701.

41. R. Cammi, B. Mennucci, J. Tomasi, Fast Evaluation of Geometries and Properties of Excited Molecules in Solution: A Tamm-Dancoff Model With Application to 4-dimethylaminobenzonitrile, J. Phys. Chem. A, 104 (2000) 5631-37.

42. M. Cossi, V. Barone, Time-Dependent Density Functional Theory for Molecules in Liquid Solutions, J. Chem. Phys., 115 (2001) 4708-17.

43. S. Mai, M. Pollum, L. Martínez-Fernández, N. Dunn, P. Marquetand, I. Corral, C. E. Crespo-Hernández, L. González, The origin of efficient triplet state population in sulfur-substituted nucleobases, Nat. Commun., 7 (2016) 13077, doi:10.1038/ncomms13077.

44. J. M. L. Pecourt, J. Peon, B. Kohler, Ultrafast internal conversion of electronically excited RNA and DNA nucleosides in water, J. Am. Chem. Soc., 122 (2000) 9348-9349, doi:10.1021/ja0021520.

45. C. T. Middleton, K. de La Harpe, C. Su, Y. K. Law, C. E. Crespo-Hernández, B. Kohler, DNA Excited-State Dynamics: From Single Bases to the Double Helix, Annu. Rev. Phys. Chem., 60 (2009) 217–239, doi:10.1146/annurev.physchem.59.032607.093719.

46. C. E. Crespo-Hernández, B. Cohen, P. M. Hare, B. Kohler, Ultrafast excited-state dynamics in

nucleic acids, Chem. Rev., 104 (2004) 1977-2019, doi:10.1021/cr0206770.

47. X. Zou, X. Dai, K. Liu, H. Zhao, D. Song, H. Su, Photophysical and Photochemical Properties of 4-Thiouracil: Time-Resolved IR Spectroscopy and DFT Studies, J. Phys. Chem. B, 118 (2014) 5864-5872, doi:10.1021/jp501658a.

48. L. Martínez-Fernández, G. Granucci, M. Pollum, C. E. Crespo-Hernández, M. Persico, I. Corral, Decoding the Molecular Basis for the Population Mechanism of the Triplet Phototoxic Precursors in UVA Light-Activated Pyrimidine Anticancer Drugs, Chem. Eur. J., 23 (2017) 2619-2627, doi:10.1002/chem.201604543.

49. C. Reichardt, C. E. Crespo-Hernaíndez, Room-temperature phosphorescence of the DNA monomer analogue 4-thiothymidine in aqueous solutions after UVA excitation, J. Phys. Chem. Lett., 1 (2010) 2239-2243, doi:10.1021/jz100729w.

50. C. Reichardt, C. Guo, C. E. Crespo-Hernández, Excited-state dynamics in 6-thioguanosine from the femtosecond to microsecond time scale, J. Phys. Chem. B, 115 (2011) 3263-3270, doi:10.1021/jp112018u.

51. M. Pollum, S. Jockusch, C. E. Crespo-Hernández, Increase in the photoreactivity of uracil derivatives by doubling thionation, Phys. Chem. Chem. Phys., 17 (2015) 27851-27861, doi:10.1039/C5CP04822B.

52. M. Pollum, S. Jockusch, C. E. Crespo-hernández, 2,4-Dithiothymine as a Potent UVA Chemotherapeutic Agent, J. Am. Chem. Soc., 136 (2014) 17930-17933.

53. M. Pollum, C. E. Crespo-Hernández, The dark singlet state as a doorway state in the ultrafast and efficient intersystem crossing dynamics in 2-thiothymine and 2-thiouracil, J. Chem. Phys., 140 (2014), 071101, doi:10.1063/1.4866447.

54. H. Yu, J. A. Sanchez-Rodriguez, M. Pollum, C. E. Crespo-Hernández, S. Mai, P. Marquetand, L. González, S. Ullrich, Internal conversion and intersystem crossing pathways in UV excited, isolated uracils and their implications in prebiotic chemistry, Phys. Chem. Chem. Phys., 18 (2016) 20168-20176, doi:10.1039/C6CP01790H.

55. J. Jiang, T. S. Zhang, J. D. Xue, X. Zheng, G. Cui, W. H. Fang, Short-time dynamics of 2-thiouracil in the light absorbing $S_2(\pi\pi^*)$ state, J. Chem. Phys., 143 (2015) 175103 doi:10.1063/1.4935047.

56. B. Ashwood, S. Jockusch, C. E. Crespo-Hernández, Excited-state dynamics of the thiopurine prodrug 6-thioguanine: Can N9-glycosylation affect its phototoxic activity?, Molecules, 22 (2017) pii: E379, doi:10.3390/molecules22030379.

57. D. Koyama, M. J. Milner, A. J. Orr-Ewing, Evidence for a Double Well in the First Triplet Excited State of 2-Thiouracil, J. Phys. Chem. B, 121 (2017) 9274-9280, doi:10.1021/acs.jpcb.7b06917.

58. Y. Harada, C. Okabe, T. Kobayashi, T. Suzuki, T. Ichimura, N. Nishi, Y. Z. Xu, Ultrafast intersystem crossing of 4-thiothymidine in aqueous solution, J. Phys. Chem. Lett., 1 (2010) 480-484, doi:10.1021/jz900276x.

59. J. A. Sánchez-Rodríguez, A. Mohamadzade, S. Mai, B. Ashwood, M. Pollum, P. Marquetand, L. González, C. E. Crespo-Hernández, S. Ullrich, 2-Thiouracil intersystem crossing photodynamics studied by wavelength-dependent photoelectron and transient absorption spectroscopies, Phys. Chem. Chem. Phys., 19 (2017) 19756-19766, doi:10.1039/C7CP02258A.

60. L. Martinez-Fernandez, T. Fahleson, P. Norman, F. Santoro, S. Coriani, R: Improta, Optical absorption and magnetic circular dichroism spectra of thiouracils: a quantum mechanical study in solution, Photochem. Photobiol. Sci., 16 (2017) 1415-1423, doi:10.1039/C7PP00105C.

61. L. Martínez-Fernández, I. Corral, G. Granucci, M. Persico, Competing ultrafast intersystem crossing and internal conversion: a time resolved picture for the deactivation of 6-thioguanine, Chem. Sci., 5 (2014) 1336, doi:10.1039/c3sc52856a.

62. M. Ruckenbauer, S. Mai, P. Marquetand, L. González, Photoelectron spectra of 2-thiouracil, 4-

thiouracil, and 2,4-dithiouracil, J. Chem. Phys., 144, (2016) 074303, doi:10.1063/1.4941948.

63. S. Mai, P. Marquetand, L. González, A Static Picture of the Relaxation and Intersystem Crossing Mechanisms of Photoexcited 2-Thiouracil, J. Phys. Chem. A, 119 (2015) 9524-9533, doi:10.1021/acs.jpca.5b06639.

64. S. Mai, P. Marquetand, L. González, Intersystem Crossing Pathways in the Noncanonical Nucleobase 2-Thiouracil: A Time-Dependent Picture, J. Phys. Chem. Lett., 7 (2016) 1978-1983, doi:10.1021/acs.jpclett.6b00616.

65. S. Bai, M. Barbatti, Why Replacing Different Oxygens of Thymine with Sulfur Causes Distinct Absorption and Intersystem Crossing, J. Phys. Chem. A, 120 (2016) 6342-6350, doi:10.1021/acs.jpca.6b05110.

66. G. Cui, W. H. Fang, State-specific heavy-atom effect on intersystem crossing processes in 2-thiothymine: A potential photodynamic therapy photosensitizer, J. Chem. Phys., 138 (2013) 044315, doi:10.1063/1.4776261.

67. B.-B. Xie, Q. Wang, W.-W. Guo, G. Cui, The excited-state decay mechanism of 2,4-dithiothymine in the gas phase, microsolvated surroundings, and aqueous solution, Phys. Chem. Chem. Phys., 19 (2017) 7689-7698, doi:10.1039/C7CP00478H.

68. A. Khvorostov, L. Lapinski, H. Rostkowska, M. J. Nowak, UV-Induced Generation of Rare Tautomers of 2-Thiouracils: A Matrix Isolation Study, J. Phys. Chem. A 109 (2005) 7700–7707.

69. R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package, WIREs Comput. Mol. Sci., 3, (2013) 198-210.

70. D. A. Case, T. A. Darden, T. E. C. III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. Kollman, AMBER 12, University of California, San Francisco, 2012.

71. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys., 79 (1983) 926.

72. J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys., 23 (1977) 327-341.

73. S. Miyamoto, P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem., 13 (1992) 952-962.

74. A. J. Pepino, J. Segarra-Martí, A. Nenov, R. Improta, M. Garavelli, Resolving ultrafast photoinduced deactivations in water-solvated pyrimidine nucleosides, J. Phys. Chem. Lett. 8 (2017) 1777-1783.

75. I. Conti, M. Garavelli, Evolution of the Excitonic State of DNA Stacked Thymines: Intrabase $\pi\pi* \rightarrow S_0$ Decay Paths Account for Ultrafast (Subpicosecond) and Longer (>100 ps) Deactivations, J. Phys. Chem. Lett. vol. 9, 9, (2018) 2373-2379.

76. I. Conti, L. Martínez-Fernández, L. Esposito, S. Hofinger, A. Nenov, M. Garavelli, R. Improta, Multiple Electronic and Structural Factors Control Cyclobutane Pyrimidine Dimer and 6–4 Thymine–Thymine Photodimerization in a DNA Duplex, Chemistry – A Eur. J. vol. 23, 60, (2017) 15177-15188.

77. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T.

Vreven, J. A. Montgometry Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Coss, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09 Rev. B01, Gaussian, Inc., Wallingford, CT, 2010.

78. F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. Fdez. Galván, N. Ferre, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, G. Hoyer, C. E. Li Manni, H. Lischka, D. Ma, P. A. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, B. Plasser, F. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata, R. Lindh, Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table, J. Comput. Chem., 37 (2016) 506.

79. P. A. Malmqvist, Calculation of transition density matrices by nonunitary orbital transformations, Int. J. Quantum Chem., 30 (1986) 479-494.

80. T. B. Pedersen, F. Aquilante, R. Lindh, Density Fitting with Auxiliary Basis Sets from Cholesky Decomposition, Theor. Chem. Acc., 124 (2009) 1-10.

81. L. Du, Z. Lan, An On-the-Fly Surface-Hopping Program JADE for Nonadiabatic Molecular Dynamics of Polyatomic Systems: Implementation and Application, J. Chem. Theory Comput., 11 (2015) 1360.

82. F. J. Avila Ferrer, J. Cerezo, E. Stendardo, R. Improta, F. Santoro, Insights for an Accurate Comparison of Computational Data to Experimental Absorption and Emission Spectra: Beyond the Vertical Transition Approximation, J. Chem. Theory Comput. 9 (2013), 2072-2082.

83. L. Serrano-Andrés, M. Merchán, R. Lindh, Computation of conical intersections by using perturbation techniques, J. Chem. Phys., 122 (2005) 104107.