



Faculty of Mathematic and Natural Sciences

Summer semester 2025

Bachelor's Thesis

in the field of

Theoretical Chemistry

on the topic

**"Excited-state Characterisation of Cytosine De-methylation: A
Benchmark Theoretical Study"**

Submitted to

First assessor: Prof. Dr. Shirin Faraji
Second assessor: Jun.-Prof. Dr. Markus Suta
Co-supervisor: Dr. David Picconi
Institute of Theoretical Chemistry
Faculty of Mathematic and Natural Sciences
Heinrich-Heine-University Düsseldorf

by

Carla Jordans
E-Mail: carla.jordans@hhu.de
Matrikel-Nr.: 3161616

Submission Date: 28.07.2025

Statutory Declaration

I hereby declare that I wrote this thesis on my own and did not use any help that is not clearly mentioned. All sources and materials I used are listed properly. Anything that came from other people is clearly marked. I understand that if this is not true, it will be considered a case of plagiarism.

Carla Jordans
28.07.2025

Contents

1 Abstract	4
2 Introduction	4
2.1 Demethylation of Cytosine	4
2.2 UV-Vis Spectroscopy and Objective on This Study	6
2.3 Theoretical Background	7
2.3.1 TD-DFT	7
2.3.2 EOM-CCSD	9
2.3.3 ADC(2)	10
3 Computational Details	12
3.1 Studied Molecules	12
3.2 Methods	13
4 Results	14
4.1 Cytosine	14
4.1.1 Effect of the basis set and the equilibrium structure	19
4.2 5-Methyl Cytosine	21
4.3 Protonated Cytosine	25
4.3.1 Protonated Enol Cytosine	25
4.3.2 Protonated Keto Cytosine	27
4.4 Protonated 5-Methyl Cytosine	30
4.4.1 Protonated Enol 5-Methyl Cytosine	30
4.4.2 Protonated Keto 5-Methyl Cytosine.	32
5 Conclusion and outlook	36

1 Abstract

This study is a benchmark of the electronic excitation energies of cytosine, 5-methyl cytosine and their protonated forms in the ultraviolet-visible domain. The equation-of-motion coupled-cluster method (EOM-CCSD), the algebraic diagrammatic construction method (ADC(2)) and the time-dependent density functional theory (TD-DFT) were employed to analyze excitation energies, oscillator strengths, and electronic transition characters. Natural transition orbitals (NTOs) were used to visualize and interpret the transitions. With EOM-CCSD as a benchmark, the results compare the methods, showing their strengths and weaknesses.

2 Introduction

2.1 Demethylation of Cytosine

5-methylcytosine (5mC) is the most common DNA modification in mammals and plays a major role in the regulation of transcription, the mediation of genomic imprinting and the X-chromosome-inactivation and the repression of transposable elements [1]. The methylation of cytosine is reversible in several ways. The first one is known as passive DNA demethylation. In this case an impaired methylation maintenance leads to demethylation in the replication of the DNA strands. Another way is the active demethylation through TET (ten-eleven translocation) proteins. In this process 5mC is first oxidized to 5-hydroxymethylcytosine (5hmC) by TET1. It is suggested that different TET proteins also catalyze the oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). When the oxidized forms are removed through replication-dependent dilution, this process is called active modification-passive dilution (AM-PD). The processes in which the oxidation is followed by the excision of 5fC or 5caC by thymine DNA glycosylase (TDG) and base excision repair (BER) is called active modification-active removal (AM-AR). The different ways of active demethylation are also shown in Figure 1.

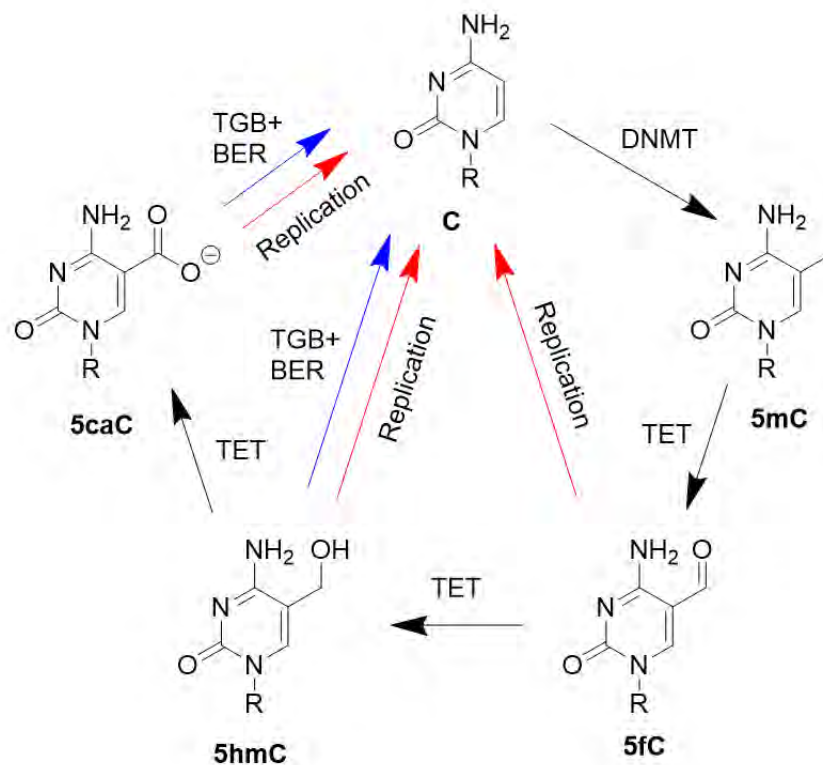


Figure 1: Overview of the active DNA demethylation pathways. DNA methyltransferase (DNMTs) modifies cytosine to 5-methylcytosine. TET enzymes oxidize 5-methylcytosine (5mC) stepwise to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). All three are diluted in replication, 5fC and 5caC can also be removed via TDG and BER. Redrawn based on Ref [1].

Mutations of TET proteins lead to aberrant active DNA demethylation. The malfunction of these proteins results in the increased self renewal of haematopoietic stem or progenitor cells and thereby promotes malignancies. One possible source of DNA damage is the exposure to UV-Vis radiation, that might lead to photochemical damage. One of the most common DNA mutations caused by UV-Vis radiation involves the formation of cyclobutane pyrimidine dimers (CPDs), which occur when two pyrimidine bases (cytosine and/or thymine) become covalently linked. This usually involves UVB or UVC radiation. At CPD sites the DNA pairing is disrupted which over time may lead to C \rightarrow T mutations [2, 3]. 5-methyl cytosine in particular is prone to undergo spontaneous deamination to thymine. Furthermore the occurrence of 5mC is correlated to the formation of CPDs [4]. Especially in tumor suppressor genes those mutations can induce skin cancer by inactivating the gene [2, 3]. Figure 2 shown the formation of a CPD involving 5-methyl cytosine and thymine.

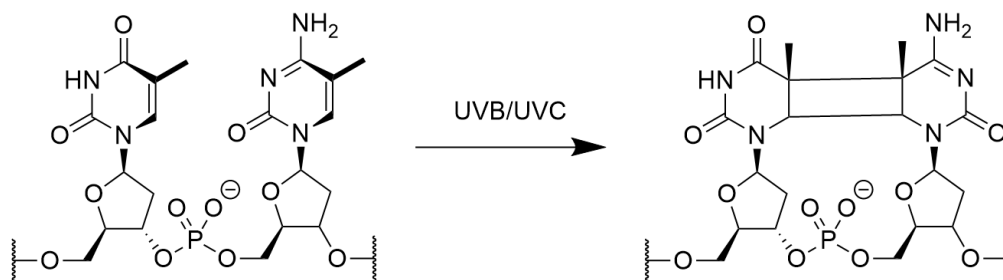


Figure 2: UV radiation induced formation of a cyclobutane pyrimidine dimer involving thymine and 5-methyl cytosine. Redrawn based on Ref [2].

To understand the mechanism of photodamage it is important to characterize the excited states of DNA fragments, which can be done via UV-Vis.

2.2 UV-Vis Spectroscopy and Objective on This Study

Ultraviolet-visible (UV-Vis) spectroscopy is a widely used technique for probing the electronic structure of molecules. It involves the absorption of ultraviolet (200–400 nm) and visible (400–800 nm) light, which leads to the excitation of electrons from occupied molecular orbitals to unoccupied ones. The transitions involved are valence and Rydberg transitions. For valence excitations these are typically $\pi\pi^*$ or $n\pi^*$ transitions where an electron is excited from a bonding π or non-bonding to an anti-bonding $\pi\pi^*$ orbital. Rydberg transitions are promotions of electrons from valence to diffuse orbitals, where the electron is only loosely associated to the electron.

The excitations as well as the absorption spectra can be simulated through different computational methods, which gives insight into the excited state structure and and simplify the interpretation of experimental data.

There is a wide range of different methods to calculate excited states. Different methods use different approaches and predict the transitions with varying accuracy. Their accuracy is linked to their computational costs. The more precise methods are therefore not suited to calculate large molecules. So when choosing a method, the accuracy has to be weighted against the computational cost in order to get reasonable results.

The aim of this study is to characterize the excited states of cytosine and 5-methylcytosine as well as their protonated variants. Furthermore different approaches and basis sets are compared, including EOM-CCSD, ADC(2) and TD-DFT. A particular focus is on the assessment of TD-DFT in terms of accuracy and reliability.

The study on the protonated derivatives might be relevant for the photophysics in a protic solvent. Furthermore, this investigation is instrumental to the interpretation of gas-phase experiments where the molecule is prepared by electrospray ionization[5]. Differing from the solid state or aqueous solutions, not only the canonical form of cytosine is present in the gas phase[6]. Therefore this study also includes a brief analysis of the most common tautomers and their composite absorption spectra.

2.3 Theoretical Background

The following sections include brief summaries on time-dependent density functional theory (TD-DFT)[7], the equation of motion coupled cluster theory (EOM-CCSD)[8] and the algebraic diagrammatic construction scheme (ADC(2))[9].

2.3.1 TD-DFT

Time dependent density functional theory (TD-DFT) is one of the most commonly used methods for the computation of excited states of medium to large molecules. A main reason for that is its low computational cost, while the method itself is still reasonably accurate[10]. TD-DFT is closely related to the ground state, time-independent density functional theory (DFT). Similar to the first Hohenberg-Kohn theorem (HKI)[11], the Runge-Gross theorem[7] states that every time-dependent electron density $\rho(r, t)$ uniquely determines a time-dependent external potential $V(r, t)$ up to a time-dependent function $C(t)$. Thus making the wave function a functional of $\rho(r, t)$ that can be determined up to a time-dependent phase factor $\alpha(t)$,

$$\Psi(r, t) = \Psi[\rho(t)](t)e^{-i\alpha(t)} , \quad (1)$$

where $\frac{d}{dt}\alpha(t) = C(t)$.

The second Hohenberg-Kohn (HKII) theorem postulates the need of a variational principle[11]. In TD-DFT this is given by the action integral (in atomic units)

$$A[\rho] = \int_{t_0}^{t_1} dt \langle \Psi[\rho](r, t) | i \frac{\partial}{\partial t} - \hat{H}(r, t) | \Psi[\rho](r, t) \rangle . \quad (2)$$

The time-dependent Hamiltonian $\hat{H}(r, t)$ is defined as

$$\hat{H}(r, t) = \hat{T}(r) + \hat{V}_{el-nuc} + \hat{V}_{el-el} + \hat{V}(t) , \quad (3)$$

where $\hat{T}(r)$, \hat{V}_{el-nuc} and \hat{V}_{el-el} are the kinetic energy operator, the electron-nuclei attraction and the electron-electron repulsion and $\hat{V}(t)$ describes a time-dependent external potential. Thus the action integral can be rewritten as

$$A[\rho] = B[\rho] - \int_{t_0}^{t_1} dt \int d^3r \rho(r, t) v(r, t), \quad (4)$$

where B is independent of $v(r, t)$.

$$B[\rho] = \int_{t_0}^{t_1} dt \langle \Psi[\rho](r, t) | i \frac{\partial}{\partial t} - \hat{T}(r) - \hat{V}_{el-el}(r) | \Psi[\rho](r, t) \rangle \quad (5)$$

The exact electron density can be obtained by the variation of the action integral with respect to the density according to the Eulers equation

$$\frac{\delta A[\rho]}{\delta \rho(r, t)} = 0 . \quad (6)$$

In both the time-dependent and time-independent Kohn-Sham equations, it is assumed that there is a non-interacting reference system with a external one particle potential v_s , such that the density ρ_s of the non-interacting system is equal to the density ρ of the real interacting system. In TD-DFT this density can be written as

$$\rho(r, t) = \rho_S(r, t) = \sum_i^N |\phi_i(r, t)|^2, \quad (7)$$

with ϕ_i being single-electron orbitals.

The time dependent single particle potential is given in Equation 8.

$$v_s(r, t) = v(r, t) + \int d^3r' \frac{\rho(r', t)}{|r - r'|} + \frac{\delta A_{XC}[\rho]}{\delta \rho(r, t)} \quad (8)$$

The applied potential $v(r, t)$ is the sum of the external potential and the electron nuclei attraction. The second term describes the Coulomb potential, while the last term is the exchange-correlation potential V_{XC} . $A_{XC}[\rho]$ is the exchange-correlation part of the action integral defined as

$$A_{XC}[\rho] = B_S[\rho] - \int_{t_0}^{t_1} dt \int d^3r \rho(r, t) v(r, t) - \frac{1}{2} \int_{t_0}^{t_1} dt \int d^3r \int d^3r' \frac{\rho(r, t) \rho(r', t)}{|r - r'|} - B[\rho]. \quad (9)$$

Ideally the exchange correlation accounts for everything missing in the non-interacting system.

The insertion of Equation 8 in the time-dependent Schrödinger equation yields the time-dependent Kohn-Sham equation.

$$i \frac{\partial}{\partial t} \phi_i(r, t) = \left(-\frac{1}{2} \nabla_i^2 + v_s(r, t) \right) \phi_i(r, t) \quad (10)$$

$$i \frac{\partial}{\partial t} \phi_i(r, t) = \left(-\frac{1}{2} \nabla_i^2 + v(r, t) + \int d^3r' \frac{\rho(r', t)}{|r - r'|} + \frac{\delta A_{XC}[\rho]}{\delta \rho(r, t)} \right) \phi_i(r, t) \quad (11)$$

Like DFT, TD-DFT is a formally exact method, but since the exact exchange correlation functional is not known approximate functionals are used to compute excited states. These functionals are the same for DFT and TD-DFT and as the name already suggests an approximation and do not hold the same level of accuracy for every molecular structure. This way there is not one ideal functional and one might has to try multiple different ones in order to find the functional that best describes the given system. The functionals can be divided into different into different categories. The first and simplest one being local density approximation (LDA) functionals such as Vosko-Wilk-Nussair (VWN)[12] or Slater-Vosko-Wilk-Nussair (SVWN). The functionals only depend on the local electron density, which is constant with a uniform electron gas.

$$E_{XC}^{LDA} = \int \epsilon_{XC}[\rho(r)] \rho(r) dr \quad (12)$$

LDA does not show correct asymptotic behavior, since real changes in density influence the system. The group of functionals that take this into account are gradient-corrected ones like Becke-Lee-Yang-Parr (BLYP)[13],[14]. These functionals correct the LDA by introducing a gradient of density into E_{XC} . Since it depends both on the density and its gradient, it can reproduce correct asymptotic behavior of

$\rho(r)$.

$$E_{XC}^{GGA} = E_{XC}^{LDA}[\rho(r)] + \Delta E_{XC} \left[\frac{|\nabla\rho|}{\rho^{4/3}} \right] \quad (13)$$

The third group consists of hybrid functionals. One example is Becke3-Lee-Young-Parr (B3LYP)[15]. These functionals mix GGA with Hartree-Fock exchange.

$$E_{XC} = E_C^{GGA} + E_X^{GGA} + C_{HF} E_X^{HF} \quad (14)$$

Range separated hybrid functionals like CAM-B3LYP[16] divide the HF-exchange into a long and a short range term, which enables it to handle long range excitations like Rydberg transitions better. The excitation energies and oscillator strengths can be obtained by using the linear response theory on top of DFT. Linear response introduces small Used for this is the Cassida equation whose matrix representation is shown below.

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \quad (15)$$

The Matrices \mathbf{A} and \mathbf{B} are defined as

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + (ia|jb) + (ia|f_{XC}|ij) \quad (16)$$

$$B_{ia,jb} = (ia|bj) + (ia|f_{XC}|bj) \quad (17)$$

The first term in the \mathbf{A} matrix describes the energy difference between the orbitals involved in the transition. The other terms, which also make up the \mathbf{B} matrix corresponds to the chosen xc potential.

2.3.2 EOM-CCSD

Another method is the equation of motion coupled cluster (EOM-CC)[8] which is generally more accurate than TD-DFT, but this comes at higher computational costs. This makes the method unsuitable for the calculation of large molecules.

In EOM-CC the excited state wavefunction is given by

$$|\Psi\rangle = R |\Psi_0\rangle, \quad (18)$$

where $|\Psi_0\rangle$ refers to the ground state wave function and R is a linear excitation operator. In general EOM-CC this operator is written as

$$R = R_0 + R_1 + R_2 + R_3 + \dots \quad (19)$$

$$R_n = \frac{1}{n!} \sum r_{ijk\dots}^{abc\dots} a^\dagger i b^\dagger j c^\dagger k \dots \quad (20)$$

In EOM-CCSD the wavefunction is truncated after the doublet excitations, which leads to the following.

$$R = R_0 + R_1 + R_2 \quad (21)$$

$$= r_0 + \sum r_i^a a^\dagger i + \frac{1}{4} \sum r_{ij}^{ab} a^\dagger i b^\dagger j \quad (22)$$

The ground state $|\Psi_0\rangle$ is described by the coupled-cluster wave function.

$$|\Psi_0\rangle = e^T |\Phi_0\rangle \quad (23)$$

Here $|\Phi_0\rangle$ is a single slater determinant usually the HF wave function. T is an excitation operator similar to R .

Inserting Equation 18 and Equation 23 into the Schrödinger equation yields Equation 24. This equation can be transformed into Equation 26 because the excitation operators commute.

$$H |\Psi_0\rangle = E |\Psi_0\rangle \quad (24)$$

$$H R e^T |\Phi_0\rangle = E R e^T |\Phi_0\rangle \quad (25)$$

$$e^{-T} H R e^T |\Phi_0\rangle = E R |\Phi_0\rangle \quad (26)$$

By introducing the similarity transformed Hamiltonian $\bar{H} \equiv e^{-T} H e^T$ this equation can be reduced to

$$\bar{H} R |\Phi_0\rangle = E R |\Phi_0\rangle . \quad (27)$$

The spectrum of \bar{H} is equal to the spectrum of H regardless of T . This way in a complete basis EOM-CC would be identical to FCI (full configuration interaction).

2.3.3 ADC(2)

The algebraic diagrammatic construction (ADC) originates from Green's function theory performing a diagrammatic perturbation expansion of the polarization propagator[9]. A derivation is also possible through the intermediate state representation (ISR)[17]. In ISR the exact electronic states are given by

$$|\Psi_n\rangle = \sum_J X_{Jn} |\tilde{\Psi}_J\rangle , \quad (28)$$

where $|\tilde{\Psi}_J\rangle$ is an intermediate state that derives from a correlated excited state

$$|\bar{\Psi}_J^N\rangle = \hat{C}_J |\Psi_0^N\rangle . \quad (29)$$

\hat{C}_J is a physical excitation operator to the exact ground state $|\Psi_0\rangle$.

$$\{\hat{C}_J\} = \{c_a^\dagger c_i; c_a^\dagger c_b^\dagger c_i c_j i < j, a < b; \dots\} \quad (30)$$

The construction of the intermediate states is done by the orthogonalizations of the correlated excited states. This procedure includes the Gram-Schmidt orthogonalization of the different excitation classes including the ground state. The Hermitian ADC matrix formed by the resulting intermediate states is a representation of subtracted Hamiltonian $\hat{H} - E_0$

$$M_{IJ} = \langle \tilde{\Psi}_I | \hat{H} - E_0 | \tilde{\Psi}_J \rangle , \quad (31)$$

which leads to the Hermitian eigenvalue problem

$$\mathbf{MX} = \mathbf{X}\Omega, \quad \mathbf{X}^\dagger \mathbf{X} = \mathbf{1} , \quad (32)$$

with \mathbf{X} being a matrix of eigenvectors and Ω denoting the eigenvalues, which can be identified as the excitation energies

$$\Omega_n = E_n - E_0 . \quad (33)$$

3 Computational Details

3.1 Studied Molecules

The molecules studied in this thesis are cytosine, 5-methyl cytosine and their protonated keto and enol form, all of which are depicted in Figure 3. Furthermore the three lowest-energy tautomers were analyzed to predict the absorption spectrum of cytosine in the gas phase.

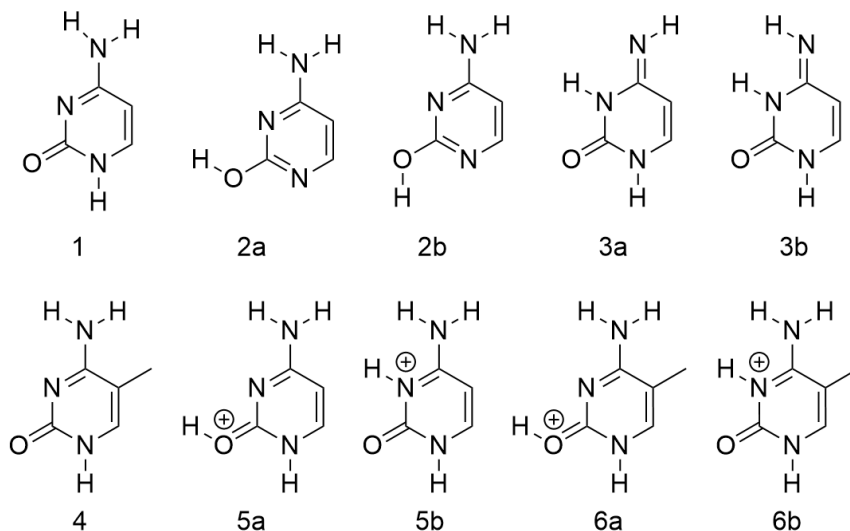


Figure 3: This figure shows the three most stable tautomers of cytosine: oxo(-amino) (1), hydroxy(-amino) (2) and (oxo)-imino (3) as well as 5-methyl cytosine (4) and both their protonated forms (5) and (6) in their enol (a) and keto (b) form. Both rotamers of 2 and 3 are shown.

The protonation is more likely to occur at the nitrogen, but since there is still the possibility for the molecule to be protonated at the oxygen both tautomers are taken into account.

The main focus of this paper are the singlet excitation of the different molecules for different methods. The excitations of cytosine, 5-methyl cytosine as well as their protonated forms are computed using EOM-CCSD, ADC(2) and TD-DFT with the functionals CAM-B3LYP and B3LYP. Oscillator strengths were also obtained in order to predict the absorption spectra. NTOs have been used for the assignment of the characters. For cytosine and methyl cytosine the first ten excitations were analyzed. The protonated forms show less excitations within the detectable range leading to the analysis of only five excitations for each structure.

The different spectra were calculated using the following formula

$$\sigma(\omega) = \sum_n f_{0n} g(\lambda - \lambda_n), \quad (34)$$

where f_{0n} is the oscillator strength and g is a lineshape function. Here the lineshape function is a Lorentzian,

$$g(\lambda - \lambda_n) = \frac{1}{\pi} \cdot \frac{\gamma}{(\lambda - \lambda_n)^2 + \gamma^2}, \quad (35)$$

where γ is the half-width at half-maximum. The value used for all spectra in this study is $\gamma = 0.25$. The tautomer populations and resulting composite spectrum were calculated using Boltzmann-weighted averaging

$$P_j = \frac{e^{-G_j/k_B T}}{\sum_j e^{-G_j/k_B T}}, \quad (36)$$

with the Boltzmann constant k_B and a temperature of $T = 298.15\text{K}$.

3.2 Methods

All calculations were performed using the electronic structure package QChem[18]. To maintain comparability, the geometries of the structures were obtained at the same level of theory. The method of choice for these calculations was MP2, which generally provides reliable predictions for the equilibrium structure, and is also the starting point for ADC(2) calculations. Three different methods were used to predict the excited states. These involved EOM-CCSD, ADC(2) and TD-DFT. The functionals for TD-DFT were B3LYP as well as CAM-B3LYP leading to a total of four calculations for each structure. In addition to the optimization and excited state calculations, frequency calculations were performed on the tautomers, in order to obtain the Gibbs free energy. 6-31+G** was chosen as the basis set for all calculations. Excited state calculations were performed on the optimized cytosine using TD-DFT/B3LYP to confirm this choice. The results of this comparison are listed in Table 4. In order to visualize the transition, the natural transition orbitals (NTOs) were computed and plotted using jmol.

4 Results

4.1 Cytosine

The calculation of excited states of cytosine has been done many times using different levels of theory. A study conducted in 2012 uses EOM-CCSD(T) for benchmarking [19], since it is comparable to CC3-LR, which is considered one of the most accurate methods for the excitations in cytosine. As already shown in ref [19], EOM-CCSD calculations for cytosine tend to overestimate the excitation energies by 0.2-0.3 eV. Similar results for EOM-CCSD are calculated in this thesis. The results of all methods are listed in Table 1. The character of the electronic transition is assigned by inspecting the natural transition orbitals (NTOs) described below.

Table 1: Comparison of the methods for the excitation energy (eV) and oscillator strength of cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	4.948	0.063	4.554	0.046	4.940	0.062	4.605	0.036
1A''	$n\pi^*$	5.457	0.003	4.749	0.000	5.271	0.002	5.100	0.001
2A''	πRy	5.710	0.004	5.510	0.005	5.780	0.004	5.323	0.005
2A'	$\pi\pi^*$	5.892	0.179	5.523	0.178	5.852	0.126	5.390	0.082
3A''	$n\pi^*$	5.939	0.000	5.217	0.002	5.818	0.001	5.603	0.000
4A''	πRy	6.203	0.007	6.034	0.008	6.285	0.004	5.888	0.002
5A''	$n\pi^*$	6.294	0.000	5.535	0.000	5.992	0.000	5.953	0.016
6A''	πRy	6.412	0.007	6.217	0.003	6.436	0.008	5.733	0.002
3A'	$\pi\pi^*$	6.574	0.625	6.210	0.662	6.559	0.437	6.247	0.195
7A''	πRy	6.672	0.004	6.477	0.002	6.719	0.005		

A first inspection of Table 1 shows that the same $\pi\pi^*$ and πRy transitions appear in almost all methods, with the $\pi\pi^*$ transitions being the only transitions with an oscillator strength above 0.01. Only the TD-DFT calculation using the B3LYP functional fails to compute the fourth Rydberg state. The strongest transition is always $3(\pi\pi^*)$, followed by $2(\pi\pi^*)$. EOM-CCSD predicts three $\pi\pi^*$ and three $n\pi^*$ transitions as well as four πRy as the lowest ten singlet excitations.

ADC(2) considerably underestimates the excitation energies compared to EOM-CCSD. However these shifts are relatively consistent for each character of the transitions. The shift is about 0.2 eV for πRy and between 0.3 and 0.4 eV for $\pi\pi^*$ transitions. A comparison of the $n\pi^*$ states is more difficult due to differences in the transitions. This will be discussed further in the following.

The difference in the excitation energies between EOM-CCSD and CAM-B3LYP lies between 0.0-0.3 eV. This difference is especially low for the $\pi\pi^*$ and πRy states with less than 0.1 eV. Only the $n\pi^*$ states show a larger shift in between the methods. It is also notable that the shift in those states lead to a change in the order of excitation states. Since the $n\pi^*$ states have little oscillator strength the difference is not visible in the calculated spectra. Compared to EOM-CCSD CAM-B3LYP seems to show a slight red shift and an increasing underestimation of the oscillator strength of the A' excitations. While also red shifted the TD-DFT calculations using the B3LYP functional not only significantly

underestimate the oscillator strength, it is also less consistent when it comes to the differences in the energies. The energy differences are between 0.3 and 0.7 eV. Like ADC(2) the $n\pi^*$ states differ from the ones predicted by EOM-CCSD. Furthermore it fails to compute the last Rydberg transition. The different shifts also lead to a swap of the two states.

The spectra resulting from the calculations are compared to an experiment conducted in water [20] (Figure 4).

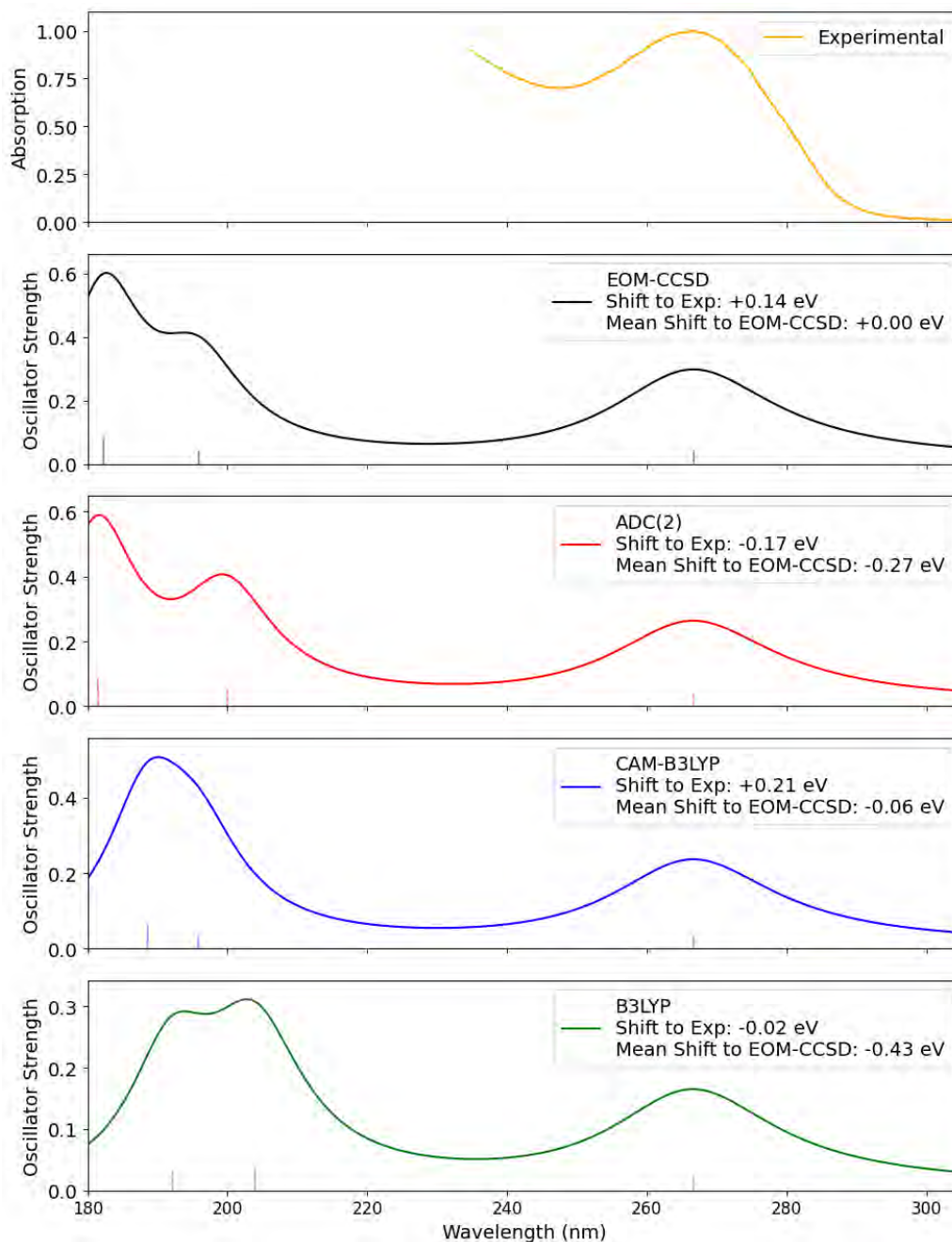


Figure 4: Experimental and calculated absorption spectra of cytosine at different level of theories.

Since the experiment was conducted with a minimal wavelength of 235 nm only the peak indicating the first $\pi\pi^*$ transition can be identified at about 267 nm. The existence of a second transition is however indicated by the gradient of the graph at 235 nm. Compared to the experiment, the calculations show an overestimation of the EOM-CCSD and TD-DFT calculation using the CAM-B3LYP functional, as well as an underestimation of ADC(2) and TD-DFT using B3LYP. The shift of the first peak for B3LYP is the lowest however it has the greatest mean shift compared to the benchmark EOM-CCSD. Closest to the benchmark is CAM-B3LYP with a mean shift of -0.06 eV. ADC(2) is red shifted by -0.27 eV. All spectra show similarities, but differ in the distance between the peaks. Only B3LYP stands out with a higher strength of the second peak.

Finally, the spectra of different tautomers is considered. Table 2 reports the relative energies and Gibbs free energies in regard to hydroxy(-amino) rotamer 2b and their percentage in the tautomeric composition.

Table 2: Relative electronic energies (ΔE), Gibbs free energies (ΔG) in kcal mol^{-1} and percentage of the tautomer in the tautomeric composition.

Tautomer	ΔE	ΔG	%
1	1.334	0.712	18.8
2a	0.809	0.283	17.7
2b	-	-	62.5
3a	2.666	2.471	1.0
3b	4.389	4.079	0.1

With 62.5% tautomer 2b has by far the highest population. Its rotamer differs by $0.809 \text{ kcal mol}^{-1}$ in the electronic energy and $0.283 \text{ kcal mol}^{-1}$ in Gibbs free energy leading to 17.7% in the tautomeric composition. The relative energies of the oxo(-amino) tautomer 1 are $\Delta E = 1.334 \text{ kcal mol}^{-1}$ and $\Delta G = 0.712 \text{ kcal mol}^{-1}$. Therefore, the canonical form is the second most common tautomer with a population of 18.8%. The rotamers 3a and 3b of the (oxo-)imino differ by more than 2 and 4 kcal mol^{-1} in both the electronic energy and the Gibbs free energy. They have the lowest percentages with only 1.0% and 0.1%.

For further calculations only the $\pi\pi^*$ transitions are taken into account, due to the weak oscillator strength of the other transitions. The excitation energy and oscillator strength of the $\pi\pi^*$ transitions are given in Table 3.

Table 3: Excitation energy (eV) and oscillator strength of the first three $\pi\pi^*$ transitions of the tautomers.

	1		2a		2b		3a		3b	
	Energy	Stren.	Energy	Stren.	Energy	Stren.	Energy	Stren.	Energy	Stren.
1A'	4.948	0.063	5.090	0.112	5.133	0.104	5.433	0.262	5.568	0.276
2A'	5.892	0.179	6.136	0.156	6.095	0.169	6.352	0.017	6.583	0.128
3A'	6.574	0.625	6.881	0.190	6.836	0.213	6.637	0.255	6.662	0.117

Compared to tautomer 1 the excitation energies of the transitions are higher in all other tautomers. The resulting composite spectrum is shown in Figure 5. Due to the high percentage of 2b, the the composite spectra is dominated by this tautomer. Since the rotamer has similar energies it also contributes to the peaks. The contribution of the (oxo)-imino tautomers is not relevant due to their low percentage. Traces of the canonical cytosine tautomer can only be seen as a shoulder in the highest peak. Apart from that, tautomer 1 does not show any relevant contribution to the spectra.

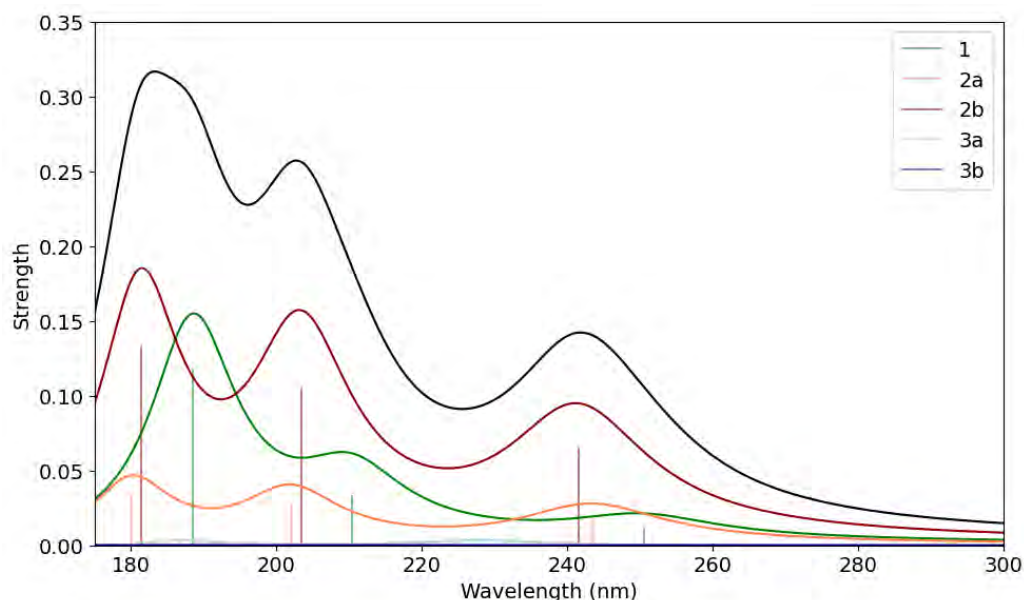
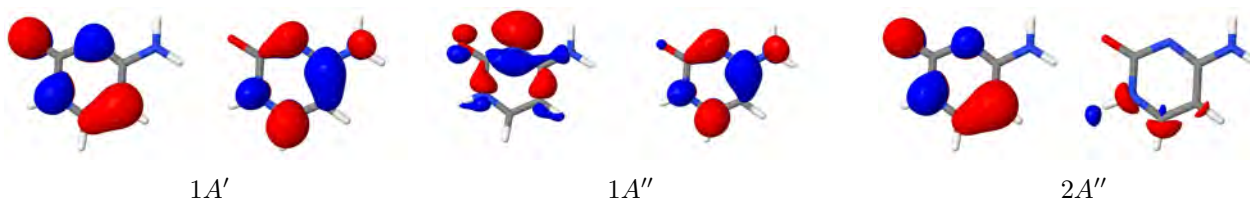


Figure 5: Computed spectra of cytosine tautomers in regard to the tautomeric composition. The black line shows the composite spectrum, obtained by Boltzmann averaging the spectra of the individual tautomers.

According to the calculations using the EOM-CCSD/6-31+G** level of theory, the lowest 10 singlet can be described by the following NTOs. These orbitals resemble those in [21].



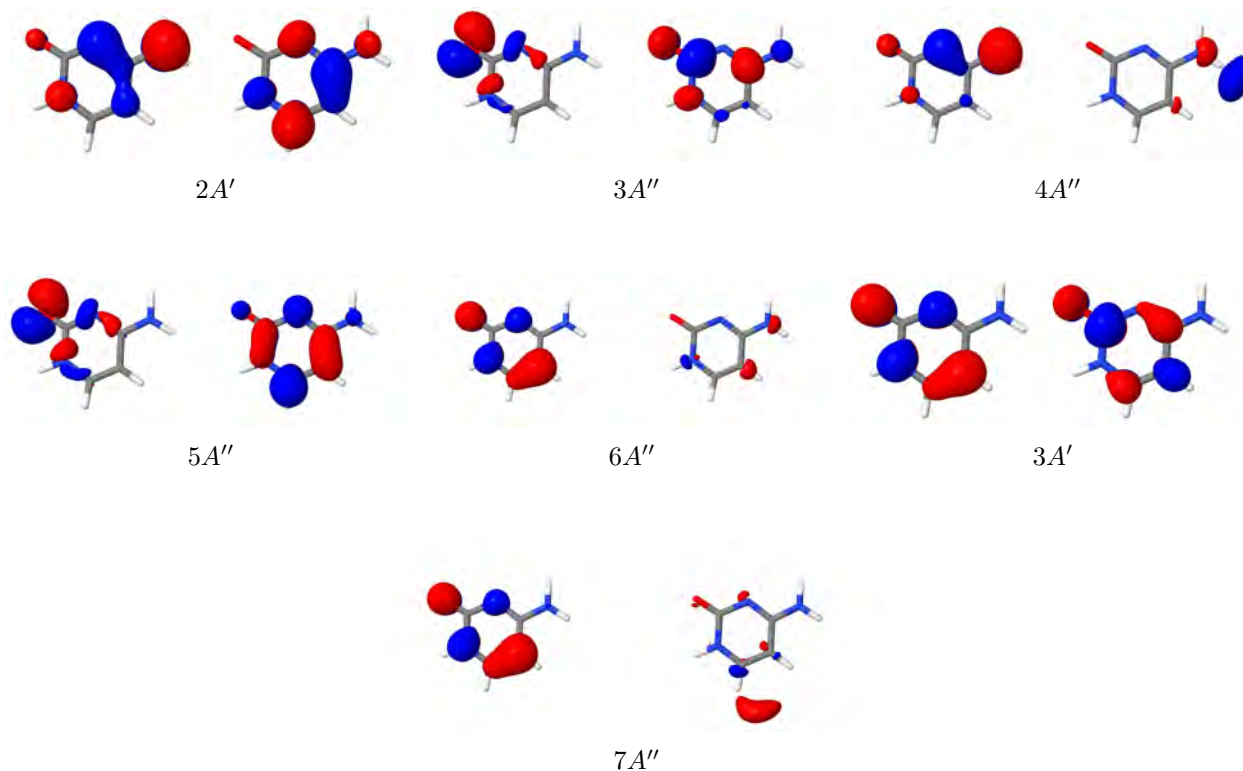
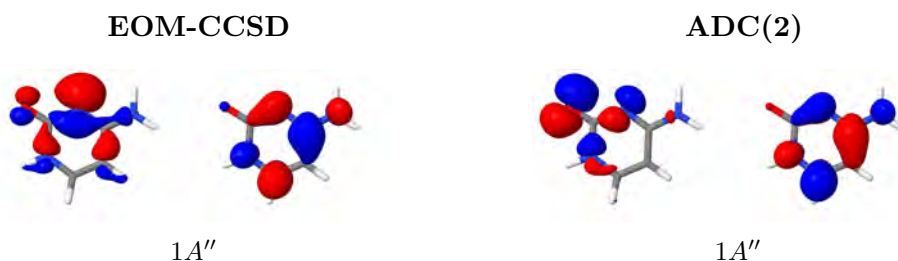


Figure 9: Natural Transition orbitals of the lowest 10 singlet excitations of cytosine computed with EOM-CCSD/6-31+G^{**}. For each transition the occupied and virtual NTOs are shown at the left and right side, respectively.

The NTOs for the other methods compare to those results when it comes to the orbitals of $\pi\pi^*$ and πRy transitions. The method does however have an influence on the $n\pi^*$ transitions. While the orbitals are quite similar for EOM-CCSD and TD-DFT/CAM-B3LYP, ADC(2) and TD-DFT/B3LYP show different results. A comparison of the NTOs can be seen in Figure 12. Ref [22] shows a similar difference between EOM-CC2 and EOM-CCSD. Both EOM-CCSD and ADC(2) predict two transitions with an excitation from the oxygen lone pair and one from the nitrogen. The transitions involving nitrogen are $1A''$ for EOM-CCSD and $2A''$ for ADC(2). They are hardly distinguishable, while the excitations from the oxygen differ significantly. A further explanation for this behavior is not given in this paper, but it explains the differences in the calculated excitation energies.



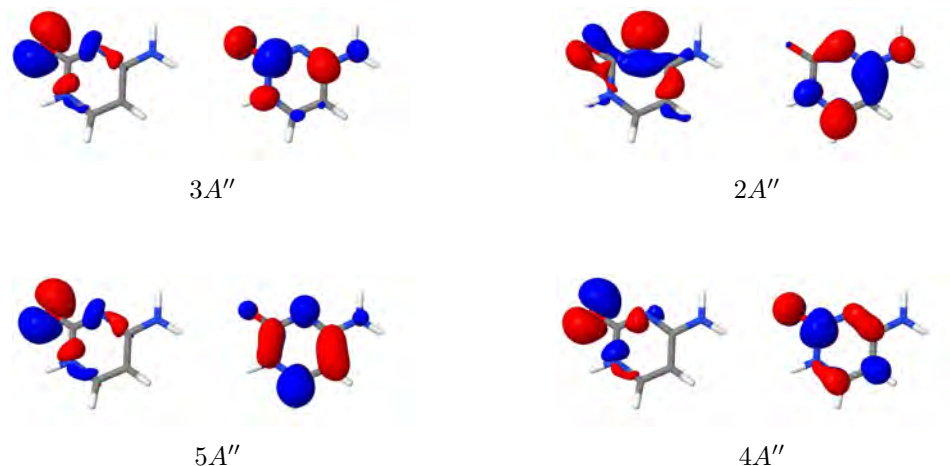


Figure 12: Natural Transition orbitals of the $n\pi^*$ transitions in cytosine computed with EOM-CCSD and ADC(2).

4.1.1 Effect of the basis set and the equilibrium structure

The choice of the basis set is important for the outcome of a calculation. However, as one can see in Table 4, the differences between basis sets can be small.

Table 4: Comparison of the excitation energy (eV) and oscillator strength of cytosine using different basis functions.

	Char.	6-31G*		6-31+G*		6-31+G**		6-31++G**	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
S_1	$\pi\pi^*$	5.048	0.055	4.947	0.062	4.940	0.062	4.941	0.062
S_2	$n\pi^*$	5.297	0.001	5.291	0.002	5.271	0.002	5.271	0.002
S_3	πRy			5.768	0.004	5.780	0.004	5.648	0.005
S_4	$n\pi^*$	5.870	0.001	5.828	0.001	5.818	0.001	5.815	0.000
S_5	$\pi\pi^*$	6.023	0.131	5.852	0.127	5.852	0.126	5.851	0.125
S_6	$n\pi^*$	6.056	0.000	5.999	0.000	5.992	0.000	5.991	0.000
S_7	πRy			6.259	0.005	6.285	0.004	6.122	0.005

While 6-31G* is not able to predict Rydberg transitions and shows a significant blue shift compared to the other basis sets (Figure 13), the other methods hold similar results, especially for the $\pi\pi^*$ and $n\pi^*$ transitions. The introduction of diffuse orbitals for H-atoms leads to slightly different results for the πRy transitions, whereas the excitation energies of the $n\pi^*$ and $\pi\pi^*$ states are essentially not affected. Nevertheless, the difference does not influence the predicted spectra since all πRy transitions would be considered dark. The use of the 6-31+G** basis set therefore is an acceptable choice and is adopted for the other molecules studied in this work.

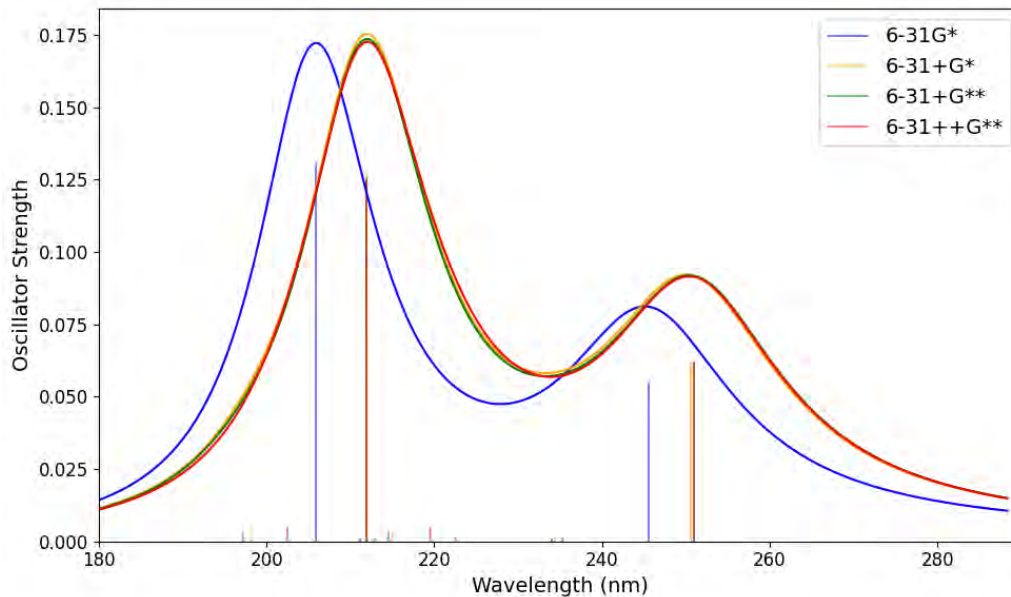


Figure 13: Comparison of the spectra resulting from the use of different basis sets for the calculation of the excited states in cytosine.

The TD-DFT/CAM-B3LYP calculations for cytosine are also used to compare the influence of different optimizations on the outcome of the excited state calculations. For this the results are compared to those using TD-DFT/CAM-B3LYP and 6-31+G** for both calculations ([21]).

Table 5: Comparison of the excitation energy (eV) and the oscillator strength of cytosine different optimization

	Char.	MP2		DFT/B3LYP	
		Energy	Strength	Energy	Strength
S_1	$\pi\pi^*$	4.940	0.062	5.00	0.06
S_2	$n\pi^*$	5.271	0.002	5.31	0.00
S_3	πRy	5.780	0.004	5.77	0.00
S_4	$n\pi^*$	5.818	0.001	5.88	0.00
S_5	$\pi\pi^*$	5.852	0.126	5.94	0.13
S_6	$n\pi^*$	5.992	0.000	6.10	0.00
S_7	πRy	6.285	0.004	6.29	0.00

As one can see in Table 5 the use DFT for the optimization leads to a slight blue shift compared to the MP2 calculations used for this thesis. The differences are slight with only about 0.1 eV at most leading to comparable results for calculations based on both optimizations.

4.2 5-Methyl Cytosine

The results of the calculations for 5-methyl cytosine are listed in Table 6.

Table 6: Comparison of the methods for the excitation energy (eV) and oscillator strength of 5-methyl cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	4.849	0.064	4.448	0.048	4.800	0.065	4.483	0.062
1A''	$n\pi^*$	5.411	0.003	4.690	0.000	5.235	0.002	4.709	0.001
2A''	πRy	5.558	0.000	5.359	0.001	5.640	0.000	5.166	0.000
2A'	$\pi\pi^*$	5.862	0.174	5.491	0.174	5.833	0.127	5.380	0.085
3A''	$n\pi^*$	5.934	0.000	5.168	0.003	5.778	0.000	5.062	0.001
4A''	πRy	6.147	0.000	5.899	0.001	6.097	0.001	5.389	0.001
5A''	πRy	6.224	0.013	6.084	0.010	6.301	0.011	5.828	0.010
6A''	$n\pi^*$	6.326	0.000	5.616	0.000	6.062	0.000	5.690	0.000
7A''	πRy	6.539	0.001	6.332	0.001	6.574	0.001	6.011	0.001
3A'	$\pi\pi^*$	6.555	0.505	6.207	0.561	6.533	0.479		

Except for B3LYP all methods seem to predict similar transitions. EOM-CCSD predicts three $\pi\pi^*$, three $n\pi^*$ and four πRy transitions that are similar to the ones in cytosine. which can be seen by looking at the NTOs (Figure 17). Compared to cytosine the order, however, changes slightly. The oscillator strengths of the $\pi\pi^*$ states are increasing with the third being significantly higher than the other two.

Compared to this, there are significant changes in the order of the excitations using ADC(2), including a change in the order of $\pi\pi^*$ states. The third and fourth transitions are swapped leading to an additional peak in the spectrum (Figure 14). The excitation energy of this transition is 5.786 eV. It is comparably weak with a strength of only 0.040. The shifts are ranging from about 0.2 eV for some of the Rydberg transitions up to 0.4 eV for the first $\pi\pi^*$ transitions. A comparison of the $n\pi^*$ transitions is once again difficult due to the different mixing of the orbitals (Figure 20). All transitions are estimated to have lower excitation energies contributing to a red shift of the whole spectrum.

The TD-DFT calculation also predict different orders for the transitions. Using the CAM-B3LYP functional leads to a switch of the second $n\pi^*$ and the second $\pi\pi^*$ as well as the third $n\pi^*$ and the third Rydberg transition. The shift of the $\pi\pi^*$ states is less than 0.05 eV, while the $n\pi^*$ states shift up to 0.3 eV. The Rydberg transitions are once again quite similar with shifts of less than 0.1 eV in each direction.

The B3LYP calculation show the same $n\pi^*$ transitions as ADC(2) making a comparison to EOM-CCSD difficult. The shifts of the other transitions are between 0.4 and 0.9 eV. It also fails to calculate the third $\pi\pi^*$ transition.

The different spectra are shown in Figure 14.

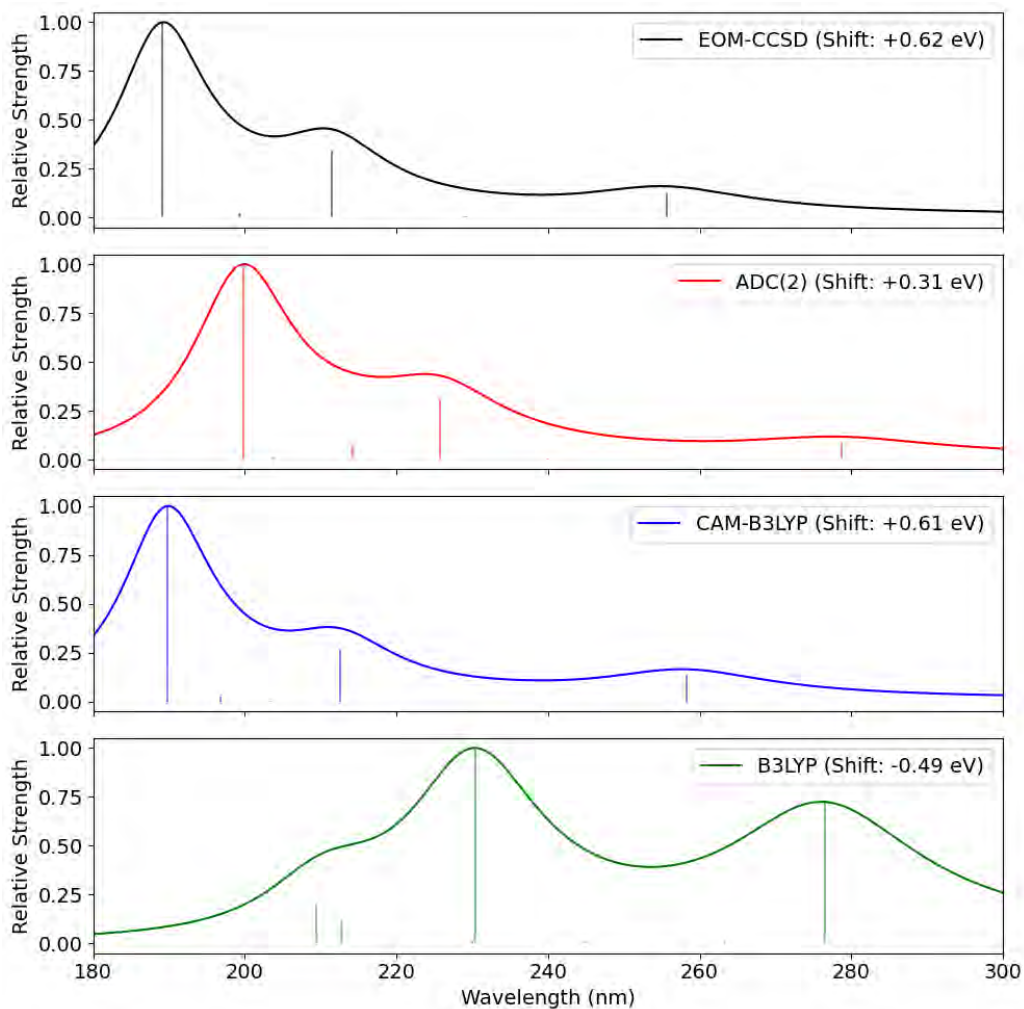


Figure 14: Absorption spectra of 5-methyl cytosine calculated by the different methods.

Similar to cytosine the spectra show strong resemblance to each other. The only exception is the one resulting from the TD-DFT calculation using B3LYP, since it is missing the third $\pi\pi^*$ transition. CAM-B3LYP is especially close to EOM-CCSD- with a mean shift of only -0.1 eV. With -0.31 eV the shift of the ADC(2) spectrum is considerably larger. Despite that the spectrum itself is slightly different in its form.

The NTOs calculated by EOM-CCSD are visualized in Figure 17. As mentioned before ADC(2) and TD-DFT using the B3LYP functional show a different mixing of the orbitals of the $n\pi^*$ transitions. Those are shown in Figure 20.

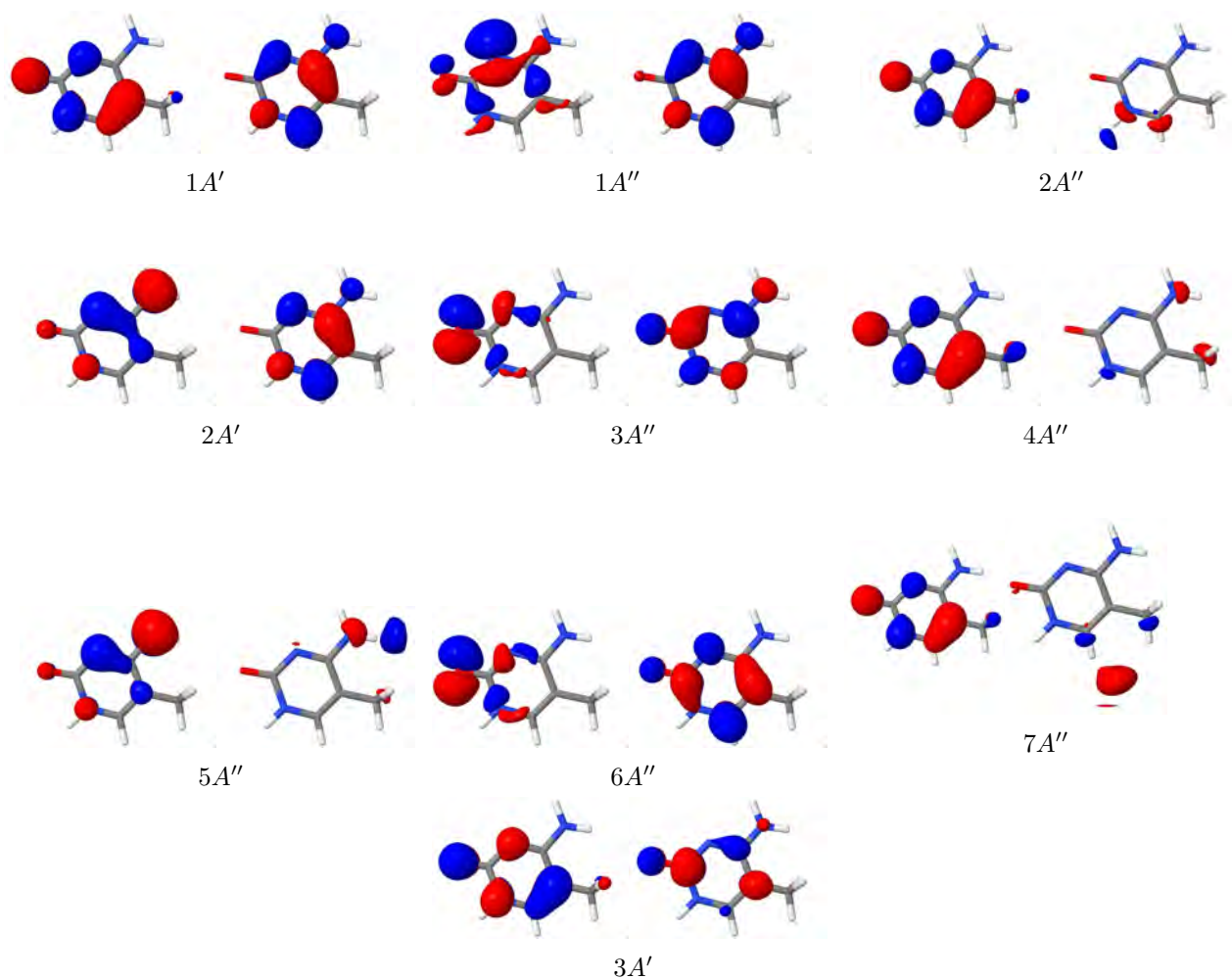
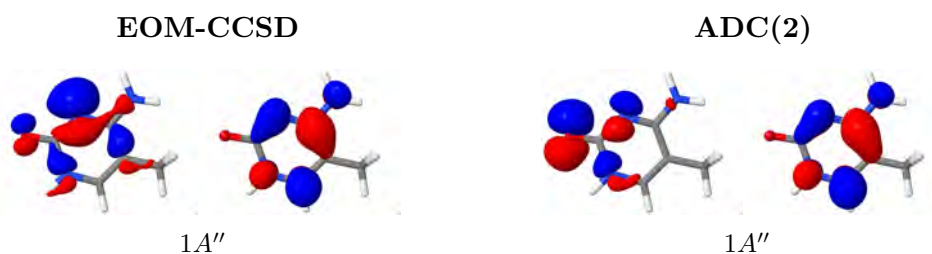


Figure 17: Natural Transition orbitals of the lowest 10 singlet excitations in 5-methyl cytosine computed with EOM-CCSD/6-31+G**. For each transition the occupied and virtual NTOs are shown a the left and right side, respectively.

Similar to cytosine there are two excitations from the nitrogen and on from the oxo group. In EOM-CCSD the second and third transitions are from oxygen, in ADC(2) it is the first and third one. The excitations involving nitrogen are relatively similar, while the others



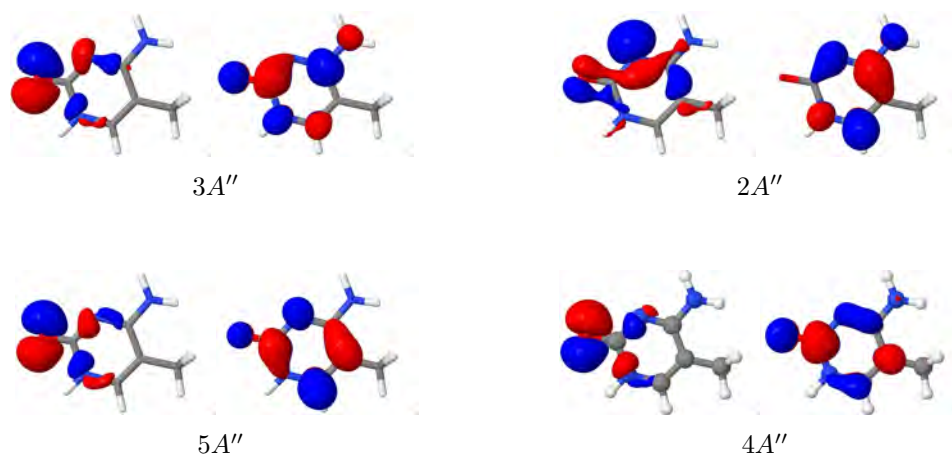


Figure 20: Natural Transition orbitals of the $n\pi^*$ transitions of 5mCyt computed with EOM-CCSD and ADC(2).

Figure 21 compares the spectra calculated with EOM-CCSD of cytosine and 5-methyl cytosine. The spectra only show a slight shift of -0.07 eV.

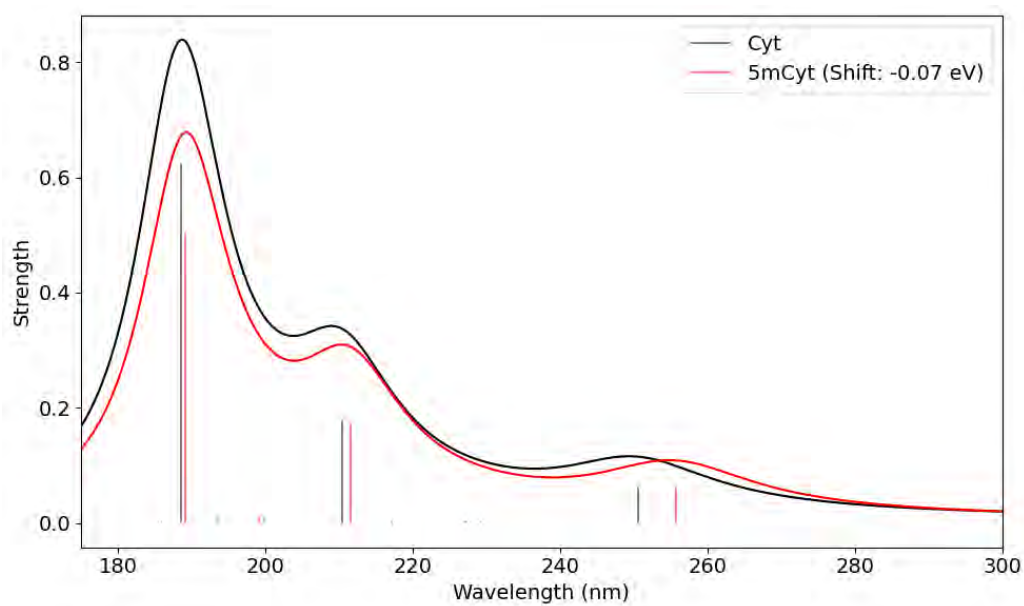


Figure 21: Comparison of the absorption spectra of cytosine and 5-methyl cytosine calculated with EOM-CCSD.

4.3 Protonated Cytosine

4.3.1 Protonated Enol Cytosine

The protonated cytosine in its enol form only has five singlet excitations with an excitation energy below 7 eV in EOM-CCSD. Those include three $\pi\pi^*$ as well as two $n\pi^*$ excitations, which are listed in Table 7.

Table 7: Comparison of the methods for the excitation energy (eV) and oscillator strength of protonated enol cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	5.471	0.220	4.939	0.305	5.506	0.152	5.287	0.098
2A'	$\pi\pi^*$	6.009	0.199	5.362	0.108	6.111	0.160	5.853	0.161
1A''	$n\pi^*$	6.275	0.001	5.407	0.002	6.030	0.001	5.685	0.001
3A'	$\pi\pi^*$	6.732	0.251	6.006	0.185	6.522	0.179	5.996	0.096
2A''	$n\pi^*$	6.980	0.005	6.090	0.003	6.589	0.004	6.268	0.003

The energies and oscillator strengths are not consistent throughout the methods. EOM-CCSD predicts the first two transitions to be $\pi\pi^*$ followed by a $n\pi^*$, a $\pi\pi^*$ and another $n\pi^*$. The oscillator strengths of the $\pi\pi^*$ states are relatively similar, with the second one being the weakest with 0.199 and the third one being the strongest with a strength of 0.251.

Compared to this, there are no changes in the order of the excitations using ADC(2). However there is a significant difference in the oscillator strength. With 0.308 the first transition is by far the brightest and the second states is far weaker with 0.199. The shifts are ranging from about 0.5 to 0.9 eV and increase with the energy for each character. The differences are once again more significant for the $n\pi^*$ transitions. All transitions are estimated to have lower excitation energies contributing to a red shift of the whole spectrum.

The TD-DFT calculation predicts different orders for the transitions and further differences when it comes to the oscillator strength. Using the CAM-B3LYP functional leads to a switch of the first $n\pi^*$ and the second $\pi\pi^*$. The shift of the of the $\pi\pi^*$ states is within 0.3 eV in each direction, while the $n\pi^*$ states shift up to 0.4 eV. All shifts increase with the excitation energy and so do the oscillator strengths of the $\pi\pi^*$ transitions. The range of the strengths is comparable to the one in EOM-CCSD, though they are weaker on average.

The B3LYP calculation predicts the second transitions to be the strongest with an oscillator strength of 0.161. The other two $\pi\pi^*$ are both slightly below 0.100. The shifts are in a similar range as they are for the other functional but only in one direction leading to a considerable redshift.

The difference in the excitations is reflected in the spectra (Figure 22). Once again seeing the great similarities between the EOM-CCSD and the CAM-B3LYP spectra.

The ADC(2) spectrum is red shifted by -0.79 eV and while the peaks are spaced similarly to EOM-CCSD the intensities are to different to match.

In B3LYP the peaks originating from the second and third $\pi\pi^*$ excitation merge into one leading to a spectra that significantly differs from the others.

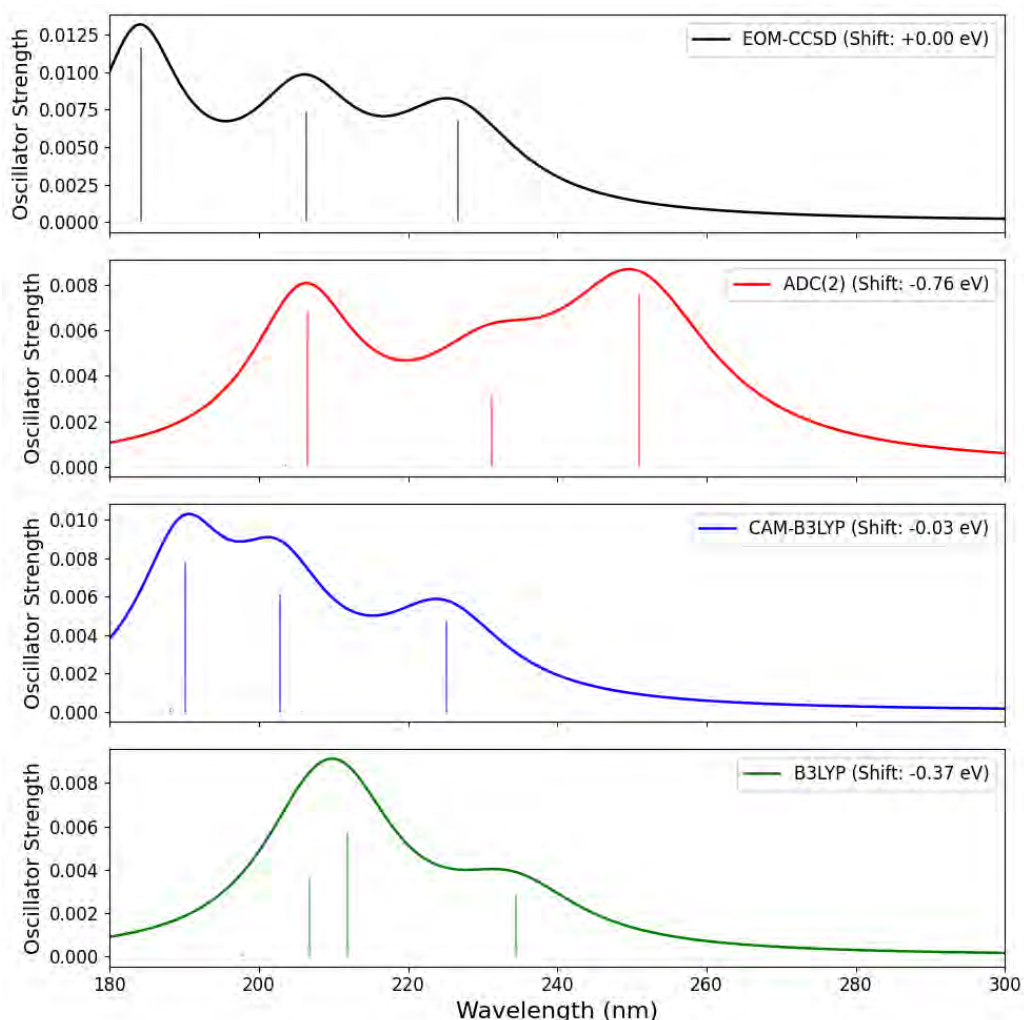
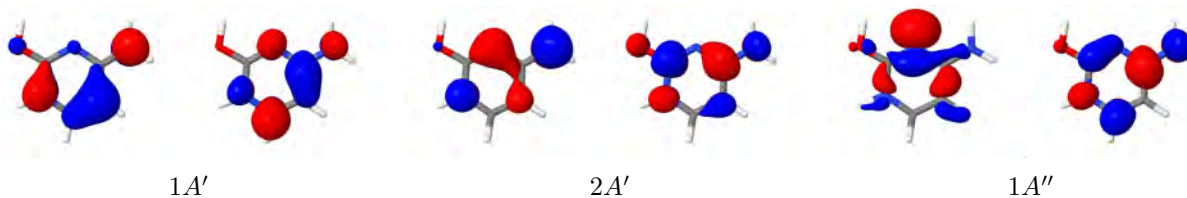


Figure 22: Absorption spectra of protonated enol cytosine calculated by the different methods.

Even though all the methods have little in common when it comes to the predicted spectra, they, differing from the previous calculations, predict the same transitions regardless of character, which are visualized in Figure 24.



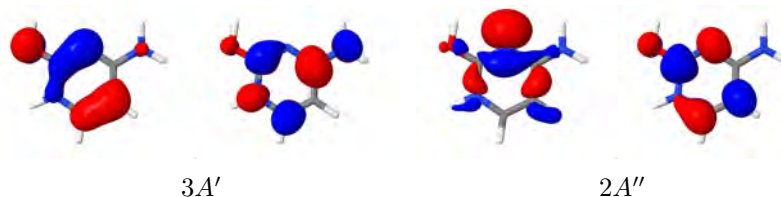


Figure 24: Natural Transition orbitals of the lowest 5 singlet excitations computed with EOM-CCSD/6-31+G**. For each transition the occupied and virtual NTOs are shown on the left and right side, respectively.

4.3.2 Protonated Keto Cytosine

The protonated keto cytosine also shows three $\pi\pi^*$ and two $n\pi^*$ under 7 eV for EOM-CCSD, which are summarized in Table 8. Looking at Figure 25 the transitions seem to be more consistent throughout the methods.

Table 8: Comparison of the methods for the excitation energy (eV) and oscillator strength of protonated keto cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	4.785	0.224	4.473	0.195	4.862	0.177	4.630	0.121
1A''	$n\pi^*$	6.230	0.000	5.214	0.000	5.965	0.000	5.028	0.000
2A'	$\pi\pi^*$	6.465	0.219	6.027	0.255	6.543	0.181	6.060	0.188
2A''	$n\pi^*$	6.542	0.000	5.961	0.000	6.260	0.000	5.970	0.000
3A'	$\pi\pi^*$	6.936	0.421	6.658	0.425	6.787	0.294	6.438	0.164

In EOM-CCSD the five transition alter between $\pi\pi^*$ and $n\pi^*$ transitions. While The first two $\pi\pi^*$ transitions show similar strengths around 0.220, the third is considerably stronger with 0.421.

In ADC(2) the $\pi\pi^*$ transitions shift by 0.3-0.5 eV. The $n\pi^*$ are less consistent with the first one differing by less than 0.1 eV and the second one by more than 0.5. The large difference leads to a swap of the $n\pi$ and the second $\pi\pi^*$ transition. Alike all previous calculation the whole spectrum is red shifted. TD-DFT/CAM-B3LYP again shows similarities to EOM-CCSD. Their excitation energies for $\pi\pi^*$ differ by about 0.1 eV. The direction of the shift differs leading to a main shift of -0.06 eV. The $n\pi^*$ transitions show no strength but are underestimated by 0.2-0.3 eV.

Once again the relative strength of the $\pi\pi^*$ B3LYP differs from the other methods. It is the only one predicting the second $\pi\pi^*$ transition as the most intense with a strength of 0.188. The differences in energy to the $\pi\pi^*$ transitions in EOM-CCSD is increasing from 0.15 to 0.50 eV, while the $n\pi^*$ transitions are under estimated by over 1 eV. The whole spectrum is shifted by -0.43 eV.

As it has been mentioned before, the spectra predicted by the different methods resemble each other again. Especially ADC(2) and CAM-B3LYP show a great resemblance to the EOM-CCSD spectrum. The greatest difference being the spacing between the peaks indicating the second and third $\pi\pi^*$ excitation. In ADC(2) it is enlarged, while CAM-B3LYP only leaves a small shoulder for the second

peak.

The shift of the B3LYP spectrum is significant, but except for the intensity of the second peak it still resembles the other spectra.

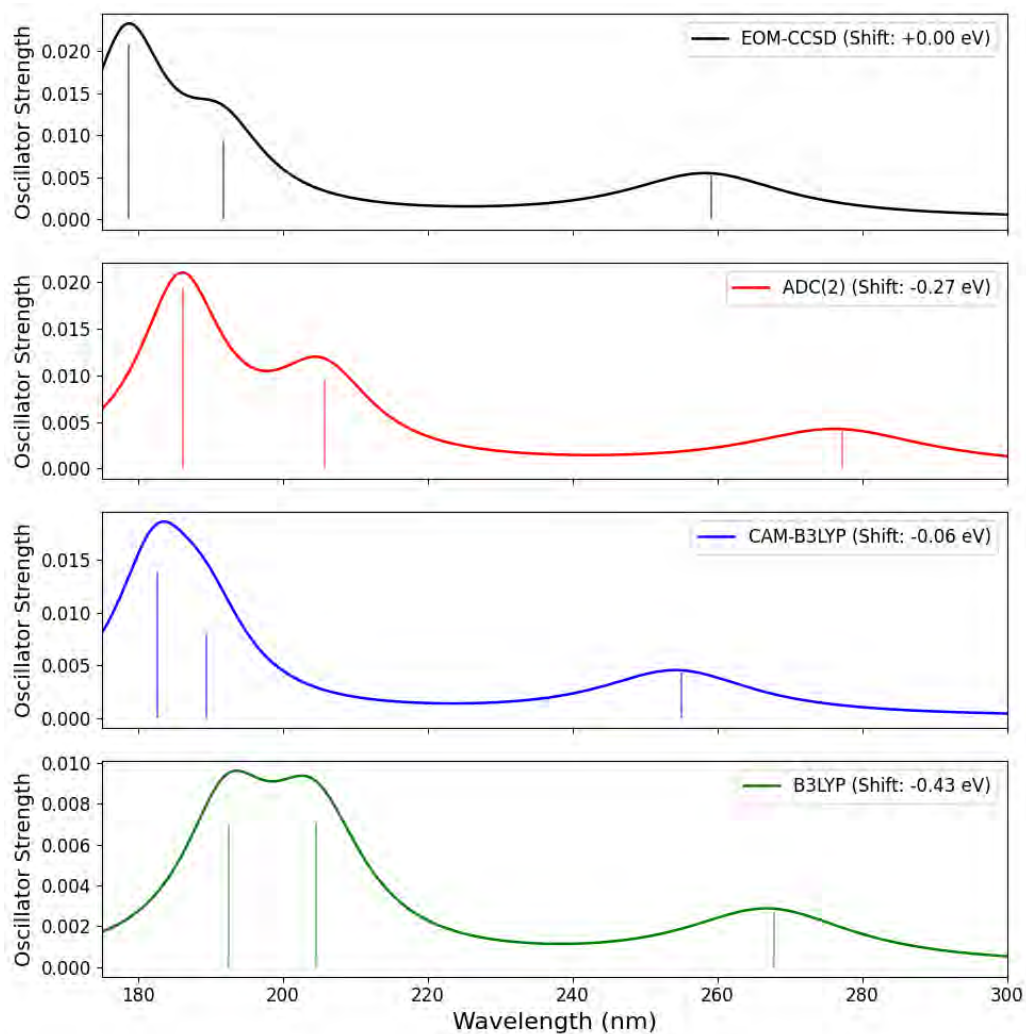


Figure 25: Absorption spectra of protonated keto cytosine calculated by the different methods.

The extreme shifts in ADC(2) and TD-DFT/B3LYP can again be explained by a difference in the $n\pi^*$. A comparison is shown in Figure 29. All NTOs calculated by EOM-CCSD are visualized in Figure 27.

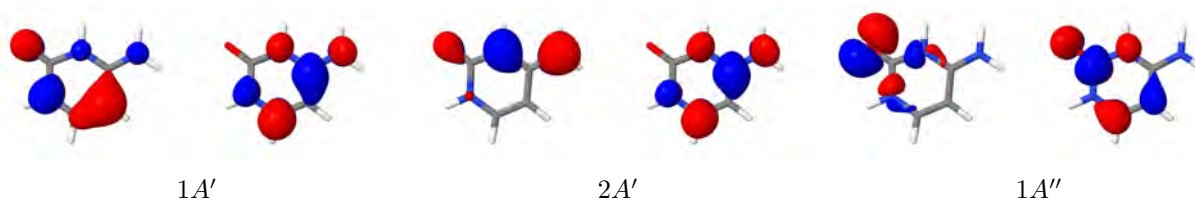




Figure 27: Natural Transition orbitals of the lowest 5 singlet excitations computed with EOM-CCSD/6-31+G**. For each transition the occupied and virtual NTOs are shown on the left and right side, respectively.

The two $n\pi^*$ transitions are both excitations from the lone pair in oxygen. While the n orbitals are indistinguishable, the π^* orbitals all differ in their shape, as well as the mixing with the carbonyl and amino group.

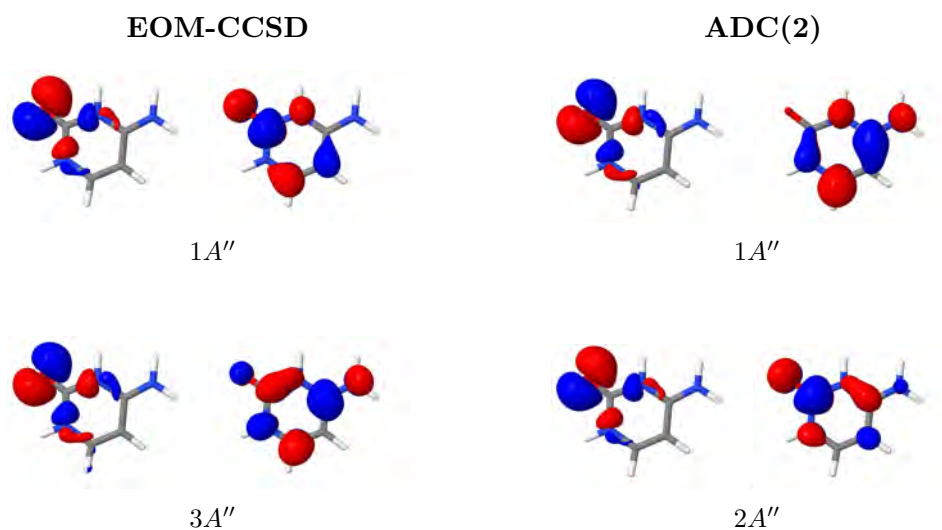


Figure 29: Natural Transition orbitals of the $n\pi^*$ transitions of the protonated keto cytosine computed with EOM-CCSD and ADC(2).

As one can see in Figure 30 the spectra of the different tautomers differ not only from each other, but also significantly from the neutral cytosine. While the first transition of the keto tautomer has an absorption at a lower energy than cytosine, the larger gap between the first and the second peak overall leads to a slight blue shift. In the enol form, however, all peaks are higher compared to their equivalents in cytosine leading to a more significant mean shift of -0.20 eV.

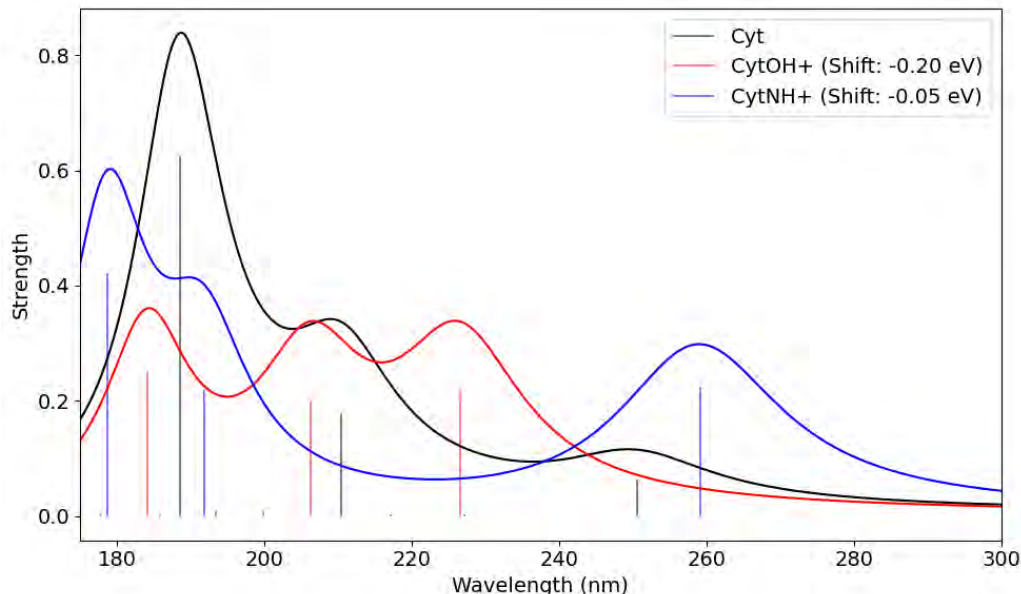


Figure 30: Comparison of the absorption spectra of cytosine and the protonated enol and keto forms calculated with EOM-CCSD.

4.4 Protonated 5-Methyl Cytosine

4.4.1 Protonated Enol 5-Methyl Cytosine

The results for the protonated 5-methyl cytosine in its keto form are listed in Table 9.

Table 9: Comparison of the methods for the excitation energy (eV) and oscillator strength of protonated enol 5-methyl cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	5.377	0.220	5.159	0.246	5.344	0.166	5.159	0.126
2A'	$\pi\pi^*$	6.016	0.131	5.709	0.121	5.978	0.100	5.760	0.110
1A''	$n\pi^*$	6.222	0.001	5.890	0.001	6.049	0.001	5.634	0.001
3A'	$\pi\pi^*$	6.687	0.337	6.255	0.320	6.475	0.258	5.973	0.124
2A''	$n\pi^*$	6.973	0.005	6.664	0.005	6.673	0.004	6.349	0.002

Similar to the tautomers of the protonated cytosine, EOM-CCSD predicts three $\pi\pi^*$ and two $n\pi^*$ transitions as the first five singlets. Their order is also the same to the demethylated equivalent. The only method showing a differing from EOM-CCSD is TD-DFT/B3LYP, where the first $n\pi^*$ is switched with the second $\pi\pi^*$ transition. In all other methods the third $\pi\pi^*$ transition as the strongest, in TD-DFT/B3LYP it is the first one. For ADC(2) the shifts of all transitions increase again with increasing excitation energy. Their shifts ranging from 0.2 to 0.4 eV.

The first $\pi\pi^*$ transitions for TD-DFT using the CAM-B3LYP functional shift by less than 0.1 eV. The greatest difference is between the second $n\pi^*$ transition with 0.3 eV. Their intensities, while smaller, have similar ratios.

As already mentioned before B3LYP predicts different relations in the strengths of the transitions as well as in the order of them. The transitions are shifted by up to 0.7 eV leading to a significant mean shift of 0.52 eV.

Despite the small shift of the CAM-B3LYP spectra compared to EOM-CCSD, ACD(2) shows a greater resemblance in the itself. The increasing shift in the excitation energies is compromised by the inverse proportionality of the energy and wavelength. B3LYP once again shows the least resemblance to the other spectra.

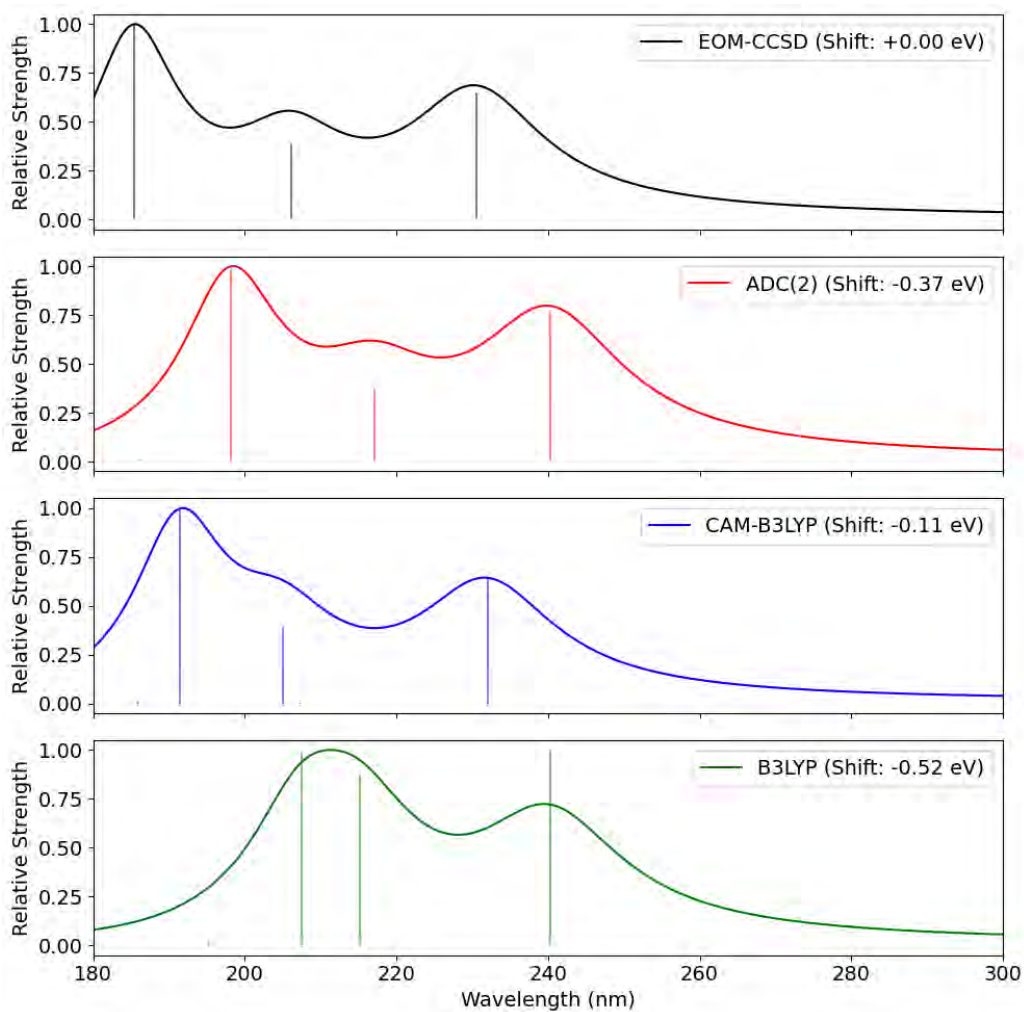


Figure 31: Absorption spectra of protonated enol 5-methyl cytosine calculated by the different methods.

The NTOs calculated by EOM-CCSD are shown in Figure 32.

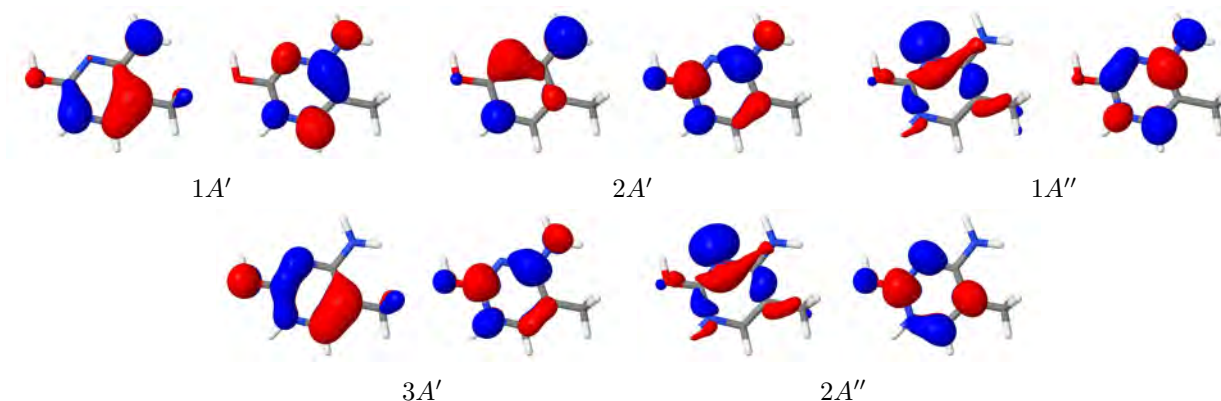


Figure 32: Natural Transition orbitals of the lowest 5 singlet excitations of the protonated enol 5-methyl cytosine computed with EOM-CCSD/6-31+G**. For each transition the occupied and virtual NTOs are shown on the left and right side, respectively.

4.4.2 Protonated Keto 5-Methyl Cytosine.

The last structure to be analyzed is the keto tautomer. Since the order of the fifth and sixth excitations are not the same across the methods, both transitions are taken into account. All transitions are listed in Table 10.

Table 10: Comparison of the methods for the excitation energy (eV) and oscillator strength of protonated keto 5-methyl cytosine.

	Char.	EOM-CCSD		ADC(2)		CAM-B3LYP		B3LYP	
		Energy	Strength	Energy	Strength	Energy	Strength	Energy	Strength
1A'	$\pi\pi^*$	4.641	0.204	4.309	0.178	4.650	0.164	4.425	0.117
1A''	$n\pi^*$	6.290	0.000	5.159	0.000	5.986	0.000	4.999	0.000
2A'	$\pi\pi^*$	6.446	0.183	6.005	0.224	6.514	0.121	6.059	0.170
2A''	$n\pi^*$	6.499	0.000	6.008	0.000	6.264	0.000	6.033	0.000
3A''	πRy	6.799	0.000	6.722	0.000	6.832	0.000	6.313	0.001
3A'	$\pi\pi^*$	6.885	0.539	6.586	0.541	6.690	0.412	6.317	0.255

All methods predict the first two transitions to be $\pi\pi^*$ followed by $n\pi^*$ as well as the third $\pi\pi^*$ transition to be the brightest. In EOM-CCSD the two first transitions are followed by another $\pi\pi^*$ and $n\pi^*$ transition as well as a πRy . The third $\pi\pi^*$ transition is the sixth with a strength of 0.539. The other two $\pi\pi^*$ transitions are significantly darker with 0.204 and 0.183. ADC(2) underestimates $\pi\pi^*$ the transitions by 0.30-0.45 eV. The first $n\pi^*$ transition has a shift of 1.1 eV and the second one shifts by 0.49 eV, while the Rydberg transition differs by less 0.1 eV. The different shifts lead to a change in the order of the transitions. In the ADC(2) calculation the third $\pi\pi^*$ occurs before the Rydberg transition.

Using the CAM-B3LYP functional leads to $\pi\pi^*$ transition that are quite similar to the ones in EOM-CCSD. They shift by less than 0.10 eV and while weaker have oscillator strengths that match in their

distribution. The $n\pi^*$ and the Rydberg transitions are underestimated by 0.2-0.3 eV which again results in a swap of the last two analyzed states.

B3LYP increasingly underestimates the transitions leading to a mean shift of -0.48 eV. The order of the excitations is however the same as it is in EOM-CCSD. The spectrum of EOM-CCSD only shows a shoulder as an indication for the second peak. This shoulder almost vanishes in the TD-DFT calculations, while ADC(2) shows a more distinct peak. The least shifted spectrum is CAM-B3LYP as it has been for all calculations.

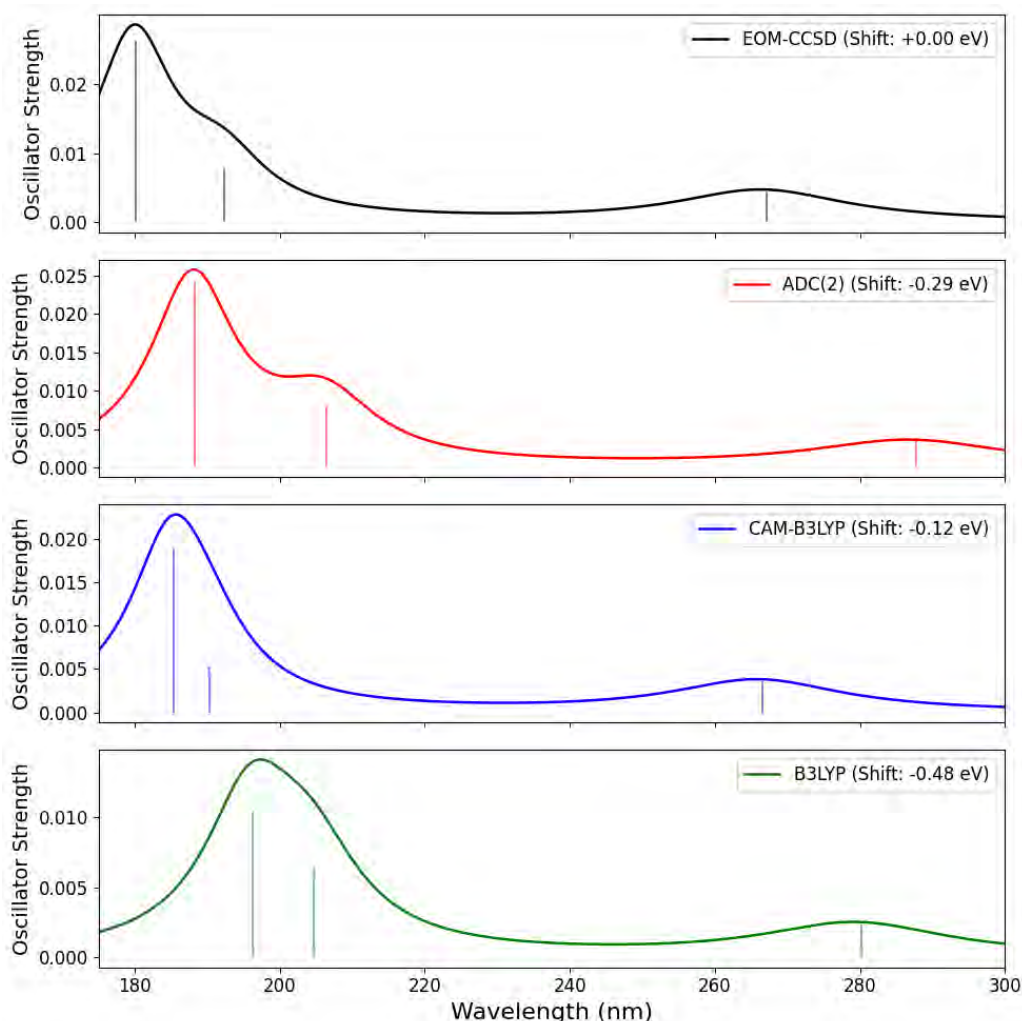
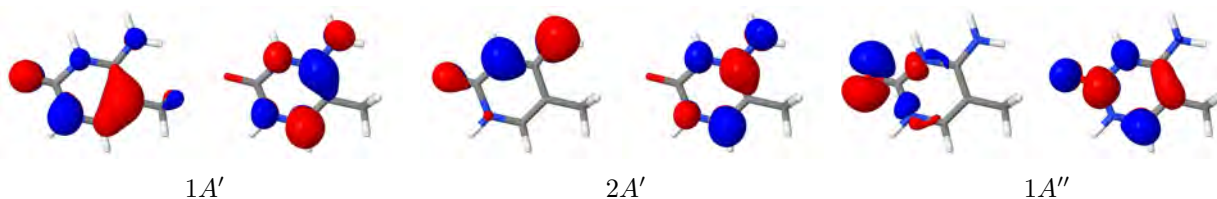


Figure 33: Absorption spectra of 5-methyl cytosine calculated by the different methods.

The NTOs of the tautomer are visualized in Figure 35. The $n\pi^*$ orbitals are once again different for the methods (Figure 37).



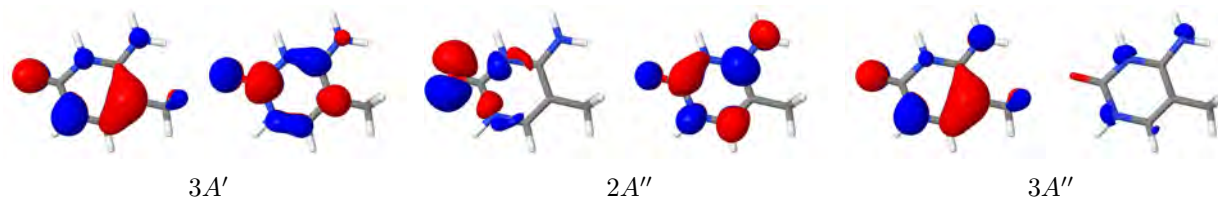


Figure 35: Natural Transition orbitals of the lowest 5 singlet excitations of the protonated keto 5-methyl cytosine computed with EOM-CCSD/6-31+G**. For each transition the occupied and virtual NTOs are shown on the left and right side, respectively.

The $n\pi^*$ orbitals of this tautomer are similar to the orbitals of the demethylated counterpart. All transitions originate from the oxygen and while the n orbitals are indistinguishable, the $\pi\pi^*$ orbitals differ significantly in their mixing.

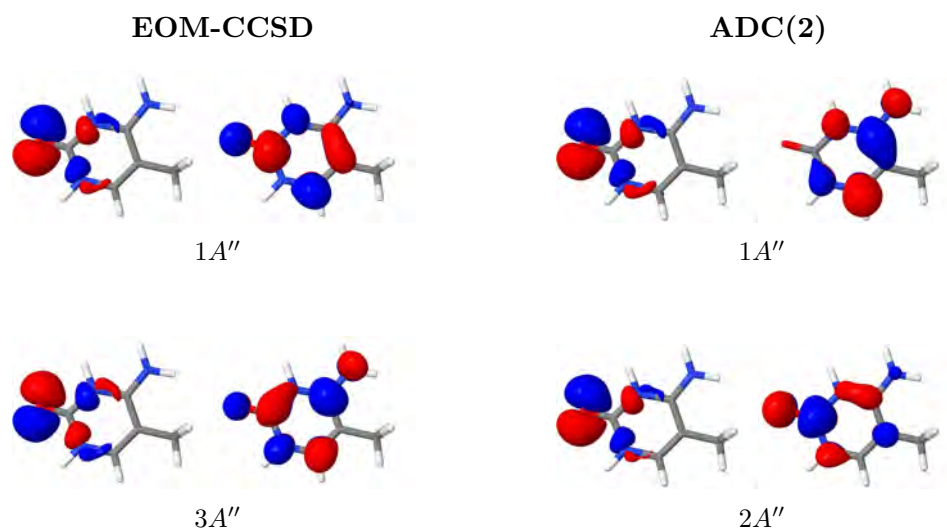


Figure 37: Natural Transition orbitals of the $n\pi^*$ transitions of the protonated keto 5-methyl cytosine computed with EOM-CCSD and ADC(2).

The comparison of the protonated forms with 5-methyl cytosine holds similar results to the ones of the demethylated cytosine. Again the first transition of the keto tautomer is the only peak lower in energy than the one in 5-methyl cytosine. In comparison the spectra are, however, a bit more red shifted. The keto form shows a mean shift of 0.06 eV and the enol form is shifted by 0.10 eV.

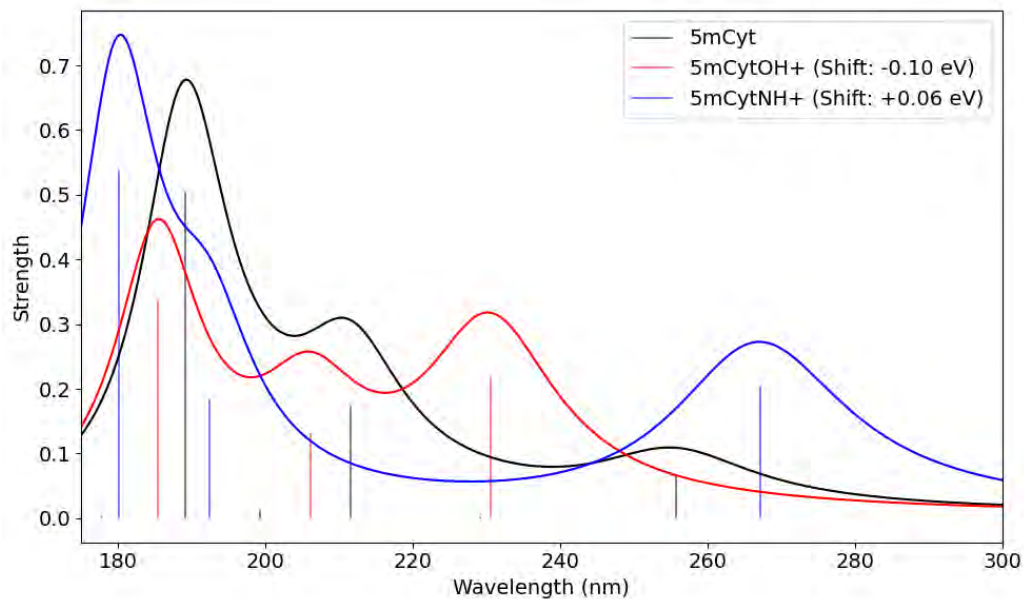


Figure 38: Comparison of the absorption spectra of 5-methyl cytosine and its protonated enol and keto tautomers calculated with EOM-CCSD.

Differing from cytosine and 5-methylcytosine, the protonated 4-methyl cytosine tautomers both show a slight blue shift of 0.04 eV.

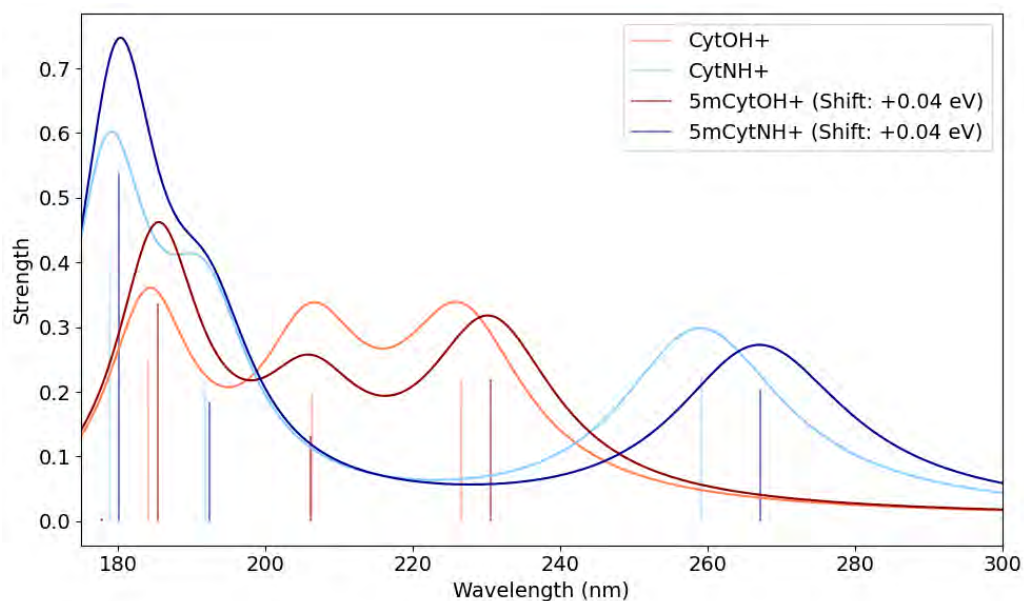


Figure 39: Comparison of the absorption spectra of the protonated enol and keto tautomers of cytosine and 5-methyl cytosine calculated with EOM-CCSD.

5 Conclusion and outlook

In this study the excited states of cytosine and methylated derivatives (in their neutral and protonated form) have been studied using a range of different theoretical methods including EOM-CCSD, ADC(2) and TD-DFT with both the CAM-B3LYP and the B3LYP functional.

The brightest state generally is S_0 or S_{10} for the neutral derivatives, S_4 or S_5 for their protonated forms and has $\pi\pi^*$ character. Most $n\pi^*$ states lie at lower excitation energies.

Since the brightest $\pi\pi^*$ states are relatively close to Rydberg states, the correct assessment of the nucleobase photophysics requires the use of diffuse basis functions. However, benchmark calculations for cytosine reveal that diffuse functions on non-H atoms are sufficient.

ADC(2) and EOM-CCSD give qualitatively similar results when it comes to the spectra. However ADC(2) shows a systematic redshift of 0.2-0.4 eV for $\pi\pi^*$ and πRy transitions and considerably larger deviations when it comes to $n\pi^*$ states. Therefore the ordering of the excited states differs from EOM-CCSD.

The TD-DFT results depend strongly on the choice of the functional. CAM-B3LYP predicts excitation energies that are in good agreement with EOM-CCSD. They typically stay within ± 0.3 eV and below 0.1 eV for most $\pi\pi^*$ transitions. In contrast B3LYP shows significant differences underestimating the energy by up to about 0.7 eV. However both functionals also frequently mispredict the state ordering especially for $n\pi^*$, but also for Rydberg transitions.

Besides helping interpreting future gas phase experiments on cytosine derivatives, these calculations are instrumental to nonadiabatic dynamics studies that are necessary to further clarify the mechanism and time scales of radiationless transitions [23]. To make accurate dynamical prediction the electronic structure is important.

The low error in the relative energy compared to EOM-CCSD make TD-DFT with the CAM-B3LYP functional an adequate alternative for gasphase experiments. However, because the ordering of the states is qualitatively wrong in TD-DFT compared to EOM-CCSD it is not suitable for accurate dynamical prediction. The same applies to ADC(2). This makes EOM-CCSD, despite the computational costs, the only eligible method for such applications presented in this study.

References

- [1] Xiaoji Wu and Yi Zhang. “TET-mediated active DNA demethylation: mechanism, function and beyond”. en. In: *Nature Reviews Genetics* 18.9 (Sept. 2017). Publisher: Nature Publishing Group, pp. 517–534. ISSN: 1471-0064. DOI: [10.1038/nrg.2017.33](https://doi.org/10.1038/nrg.2017.33). URL: <https://www.nature.com/articles/nrg.2017.33> (visited on 07/07/2025).
- [2] John Stephen Taylor. “Unraveling the Molecular Pathway from Sunlight to Skin Cancer”. In: *Accounts of Chemical Research* 27.3 (Mar. 1994). Publisher: American Chemical Society, pp. 76–82. ISSN: 0001-4842. DOI: [10.1021/ar00039a003](https://doi.org/10.1021/ar00039a003). URL: <https://doi.org/10.1021/ar00039a003> (visited on 07/24/2025).
- [3] Thierry Douki. “The variety of UV-induced pyrimidine dimeric photoproducts in DNA as shown by chromatographic quantification methods”. en. In: *Photochemical & Photobiological Sciences* 12.8 (Aug. 2013), pp. 1286–1302. ISSN: 1474-9092. DOI: [10.1039/c3pp25451h](https://doi.org/10.1039/c3pp25451h). URL: <https://doi.org/10.1039/c3pp25451h> (visited on 07/24/2025).
- [4] Lara Martinez-Fernandez et al. “UV-induced damage to DNA: effect of cytosine methylation on pyrimidine dimerization”. In: *Signal Transduction and Targeted Therapy* 2 (June 2017), p. 17021. ISSN: 2095-9907. DOI: [10.1038/sigtrans.2017.21](https://doi.org/10.1038/sigtrans.2017.21). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5661629/> (visited on 07/14/2025).
- [5] Thomas Schlathölder and Jean-Christophe Pouilly. “Radiation-Induced Molecular Processes in DNA: A Perspective on Gas-Phase Interaction Studies”. en. In: *Chemistry – A European Journal* 30.38 (2024). eprint: <https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/chem.202400633>, e202400633. ISSN: 1521-3765. DOI: [10.1002/chem.202400633](https://doi.org/10.1002/chem.202400633). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.202400633> (visited on 07/25/2025).
- [6] Gábor Bazsó et al. “Tautomers of cytosine and their excited electronic states: a matrix isolation spectroscopic and quantum chemical study”. en. In: *Physical Chemistry Chemical Physics* 13.15 (Mar. 2011). Publisher: The Royal Society of Chemistry, pp. 6799–6807. ISSN: 1463-9084. DOI: [10.1039/C0CP02354J](https://doi.org/10.1039/C0CP02354J). URL: <https://pubs.rsc.org/en/content/articlelanding/2011/cp/c0cp02354j> (visited on 07/08/2025).
- [7] Erich Runge and E. K. U. Gross. “Density-Functional Theory for Time-Dependent Systems”. en. In: *Physical Review Letters* 52.12 (Mar. 1984), pp. 997–1000. ISSN: 0031-9007. DOI: [10.1103/PhysRevLett.52.997](https://doi.org/10.1103/PhysRevLett.52.997). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.52.997> (visited on 06/30/2025).
- [8] John F. Stanton and Rodney J. Bartlett. “The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties”. In: *The Journal of Chemical Physics* 98.9 (May 1993), pp. 7029–7039. ISSN: 0021-9606. DOI: [10.1063/1.464746](https://doi.org/10.1063/1.464746). URL: <https://doi.org/10.1063/1.464746> (visited on 07/24/2025).

- [9] Jochen Schirmer. “Beyond the random-phase approximation: A new approximation scheme for the polarization propagator”. In: *Physical Review A* 26.5 (Nov. 1982). Publisher: American Physical Society, pp. 2395–2416. DOI: [10.1103/PhysRevA.26.2395](https://doi.org/10.1103/PhysRevA.26.2395). URL: <https://link.aps.org/doi/10.1103/PhysRevA.26.2395> (visited on 07/23/2025).
- [10] Andreas Dreuw and Martin Head-Gordon. “Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules”. In: *Chemical Reviews* 105.11 (Nov. 2005). Publisher: American Chemical Society, pp. 4009–4037. ISSN: 0009-2665. DOI: [10.1021/cr0505627](https://doi.org/10.1021/cr0505627). URL: <https://doi.org/10.1021/cr0505627> (visited on 06/30/2025).
- [11] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Physical Review* 136.3B (Nov. 1964). Publisher: American Physical Society, B864–B871. DOI: [10.1103/PhysRev.136.B864](https://doi.org/10.1103/PhysRev.136.B864). URL: <https://link.aps.org/doi/10.1103/PhysRev.136.B864> (visited on 07/23/2025).
- [12] S. H. Vosko, L. Wilk, and M. Nusair. “Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis”. en. In: *Canadian Journal of Physics* 58.8 (Aug. 1980), pp. 1200–1211. ISSN: 0008-4204 1208-6045. DOI: [10.1139/p80-159](https://doi.org/10.1139/p80-159). URL: <https://escholarship.org/uc/item/23j4q7zm> (visited on 07/23/2025).
- [13] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. en. In: *Physical Review A* 38.6 (Sept. 1988), pp. 3098–3100. ISSN: 0556-2791. DOI: [10.1103/PhysRevA.38.3098](https://doi.org/10.1103/PhysRevA.38.3098). URL: <https://link.aps.org/doi/10.1103/PhysRevA.38.3098> (visited on 06/30/2025).
- [14] Chengteh Lee, Weitao Yang, and Robert G. Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Physical Review B* 37.2 (Jan. 1988). Publisher: American Physical Society, pp. 785–789. DOI: [10.1103/PhysRevB.37.785](https://doi.org/10.1103/PhysRevB.37.785). URL: <https://link.aps.org/doi/10.1103/PhysRevB.37.785> (visited on 06/30/2025).
- [15] Axel D. Becke. “Density-functional thermochemistry. III. The role of exact exchange”. In: *The Journal of Chemical Physics* 98.7 (Apr. 1993), pp. 5648–5652. ISSN: 0021-9606. DOI: [10.1063/1.464913](https://doi.org/10.1063/1.464913). URL: <https://doi.org/10.1063/1.464913> (visited on 07/07/2025).
- [16] Takeshi Yanai, David P Tew, and Nicholas C Handy. “A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”. In: *Chemical Physics Letters* 393.1 (July 2004), pp. 51–57. ISSN: 0009-2614. DOI: [10.1016/j.cplett.2004.06.011](https://doi.org/10.1016/j.cplett.2004.06.011). URL: <https://www.sciencedirect.com/science/article/pii/S0009261404008620> (visited on 07/22/2025).
- [17] J. Schirmer and A. B. Trofimov. “Intermediate state representation approach to physical properties of electronically excited molecules”. In: *The Journal of Chemical Physics* 120.24 (June 2004), pp. 11449–11464. ISSN: 0021-9606. DOI: [10.1063/1.1752875](https://doi.org/10.1063/1.1752875). URL: <https://doi.org/10.1063/1.1752875> (visited on 07/22/2025).
- [18] *Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package — The Journal of Chemical Physics — AIP Publishing*. URL: <https://pubs.aip.org/aip/jcp/article/155/8/084801/1074802/Software-for-the-frontiers-of-quantum-chemistry-An> (visited on 07/24/2025).

- [19] Péter G. Szalay et al. “Benchmark Studies on the Building Blocks of DNA. 1. Superiority of Coupled Cluster Methods in Describing the Excited States of Nucleobases in the Franck–Condon Region”. In: *The Journal of Physical Chemistry A* 116.25 (June 2012). Publisher: American Chemical Society, pp. 6702–6710. ISSN: 1089-5639. DOI: [10.1021/jp300977a](https://doi.org/10.1021/jp300977a). URL: <https://doi.org/10.1021/jp300977a> (visited on 07/25/2025).
- [20] Masahiko Taniguchi and Jonathan S. Lindsey. “Database of Absorption and Fluorescence Spectra of 300 Common Compounds for use in PhotochemCAD”. en. In: *Photochemistry and Photobiology* 94.2 (2018). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/php.12860>, pp. 290–327. ISSN: 1751-1097. DOI: [10.1111/php.12860](https://doi.org/10.1111/php.12860). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/php.12860> (visited on 07/22/2025).
- [21] Martha Yaghoubi Jouybari et al. “Ultrafast Dynamics of the Two Lowest Bright Excited States of Cytosine and 1-Methylcytosine: A Quantum Dynamical Study”. In: *Journal of Chemical Theory and Computation* 16.9 (Sept. 2020). Publisher: American Chemical Society, pp. 5792–5808. ISSN: 1549-9618. DOI: [10.1021/acs.jctc.0c00455](https://doi.org/10.1021/acs.jctc.0c00455). URL: <https://doi.org/10.1021/acs.jctc.0c00455> (visited on 06/29/2025).
- [22] Dániel Kánnár and Péter G. Szalay. “Benchmarking coupled cluster methods on singlet excited states of nucleobases”. en. In: *Journal of Molecular Modeling* 20.11 (Nov. 2014), p. 2503. ISSN: 0948-5023. DOI: [10.1007/s00894-014-2503-2](https://doi.org/10.1007/s00894-014-2503-2). URL: <https://doi.org/10.1007/s00894-014-2503-2> (visited on 07/17/2025).
- [23] Shirin Faraji, David Picconi, and Elisa Palacino-González. “Advanced quantum and semiclassical methods for simulating photoinduced molecular dynamics and spectroscopy”. en. In: *WIREs Computational Molecular Science* 14.5 (2024). eprint: <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.e1731>. ISSN: 1759-0884. DOI: [10.1002/wcms.1731](https://doi.org/10.1002/wcms.1731). URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1731> (visited on 07/25/2025).