# Table of contents

List of tables iv

List of figures vii

Abstract x

Chapter 1: Introduction and overview 1
  1.1 Potential energy surfaces and their intersections 1
  1.1.1 Methods for the ground electronic state 4
  1.1.2 Methods for excited electronic states 10
  1.1.3 Density functional theory methods 18
  1.2 Previous studies of the green fluorescent protein chromophore 19
  1.3 Previous studies of the excited states of uracil 26
  1.4 Chapter 1 references 29

Chapter 2: Electronic properties of the green fluorescent protein chromophore 42
  2.1 Chapter 2 introduction 42
  2.2 Computational details 43
    2.2.1 Continuum solvation models 44
    2.2.2 Effective fragment potential-based QM/MM method 45
  2.3 Anionic form of HBDI 46
    2.3.1 Molecular orbital framework 46
    2.3.2 Vertical electron detachment energy of the HBDI anion 48
    2.3.3 Vertical excitation energy and electronic properties of the singlet and triplet $\pi\pi^*$ transitions of the HBDI anion 50
    2.3.4 Cis-trans isomerization pathway of anionic HBDI 61
  2.4 Chapter 2 conclusions 73
  2.5 Chapter 2 references 77

Chapter 3: The effect of oxidation on the electronic structure of the green fluorescent protein chromophore 81
  3.1 Chapter 3 introduction 81
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>Computational details</td>
<td>86</td>
</tr>
<tr>
<td>3.3</td>
<td>Results and discussion</td>
<td>89</td>
</tr>
<tr>
<td>3.4</td>
<td>Chapter 3 conclusions</td>
<td>99</td>
</tr>
<tr>
<td>3.5</td>
<td>Chapter 3 references</td>
<td>102</td>
</tr>
</tbody>
</table>

**Chapter 4: Electronically excited states of uracil**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Chapter 4 introduction</td>
<td>107</td>
</tr>
<tr>
<td>4.2</td>
<td>Computational details</td>
<td>108</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and discussion</td>
<td>110</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Structure and electronic states of uracil</td>
<td>110</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Effect of the one-electron basis set</td>
<td>114</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Effects due to including higher excitations</td>
<td>118</td>
</tr>
<tr>
<td>4.3.4</td>
<td>MRCI results</td>
<td>121</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Solvent effects</td>
<td>123</td>
</tr>
<tr>
<td>4.3.6</td>
<td>Best theoretical estimate of vertical excitation energies in uracil</td>
<td>125</td>
</tr>
<tr>
<td>4.4</td>
<td>Chapter 4 conclusions</td>
<td>128</td>
</tr>
<tr>
<td>4.5</td>
<td>Chapter 4 references</td>
<td>129</td>
</tr>
</tbody>
</table>

**Chapter 5: Direct location of the minimum point on intersection seams of potential energy surfaces with equation-of-motion coupled-cluster methods**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Chapter 5 introduction</td>
<td>134</td>
</tr>
<tr>
<td>5.2</td>
<td>Theory</td>
<td>137</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Adiabatic approximation</td>
<td>137</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Crossing location and minimization</td>
<td>140</td>
</tr>
<tr>
<td>5.3</td>
<td>Computational details</td>
<td>143</td>
</tr>
<tr>
<td>5.4</td>
<td>Benchmarks and examples</td>
<td>143</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Cyclic trinitrogen cation</td>
<td>143</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Nitrogen dioxide</td>
<td>146</td>
</tr>
<tr>
<td>5.4.3</td>
<td>para-Benzene diradical</td>
<td>149</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion: Performance of the algorithm and details of implementation</td>
<td>153</td>
</tr>
<tr>
<td>5.6</td>
<td>Chapter 5 conclusions</td>
<td>155</td>
</tr>
<tr>
<td>5.7</td>
<td>Chapter 5 references</td>
<td>156</td>
</tr>
</tbody>
</table>

**Chapter 6: Future work**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Chapter 6 references</td>
<td>168</td>
</tr>
</tbody>
</table>

**Bibliography**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>170</td>
</tr>
</tbody>
</table>

**Appendix: Tensor algebra library**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>187</td>
</tr>
</tbody>
</table>
List of tables

1.1 Selected theoretical estimates of the $S_0$-$S_1$ vertical excitation energy ($\Delta E$) and the corresponding wavelengths for the gas phase GFP chromophore. The experimental value is also given. ........................................ 22

2.1 The interaction of the bright $\pi\pi^*$ state with the electron-detached continuum as reflected by the diminishing oscillator strength ($f_L$) in CIS and TD-DFT/BNL calculations. Also see Fig. 2.4. ......................... 54

2.2 Vertical $\pi\pi^*$ electronic excitation energies ($\Delta E$, in eV), corresponding wavelengths (nm) and oscillator strength ($f_L$) for the gas phase GFP chromophore (Fig. 1). The reference experimental value is 2.59 eV (479 nm). ............................................................... 57

2.3 Vertical triplet $\pi\pi^*$ electronic excitation energies ($\Delta E$, in eV) of the HBDI anion in the gas phase. ............................................................... 60

2.4 Geometry parameters (distances in Å, angles in degrees) at the stationary points for the gas-phase cis-trans chromophore isomerization. .............. 64

2.5 Cumulative natural charges on the fragments of the HBDI molecule: the phenyl and dimethylimidazolin rings and the CH bridge calculated with CASSCF(12/11) and DFT (PBE0 functional). ................................. 67

2.6 Cis-trans energy difference $\Delta E_{tc}$ and the energy barrier $E_a$ for the cis-trans isomerization of the HBDI anion calculated at various levels of theory for the chromophore molecule and solvent. ......................... 72

3.1 Vertical excitation energies (eV) and oscillator strengths of the $\pi\pi^*$ states of the doublet HBDI radical. The values of $\langle S^2 \rangle$ are also shown. .... 89

3.2 Vertical excitation energies (eV) and oscillator strengths of the excited states of the HBDI cation at its equilibrium geometry. ......................... 92
3.3 Standard oxidation potentials corresponding to different one-electron oxidation transitions computed using gas-phase detachment and ionization energies (see text). 95

3.4 Selected geometric parameters (bond lengths in Å, angles in degrees) of the HBDI anion, and its singly and doubly oxidized forms. 98

4.1 Equilibrium geometry of uracil. The calculated structure is optimized with B3LYP/6-311G(2df,2pd). Experimental values are obtained by averaging dimensions found in crystal structures. Atomic labels are defined in Fig. 4.1, distances in Å, angles in degrees. 113

4.2 Leading electronic configurations in the EOM-CCSD/aug-cc-pVDZ wave function. \( r \) is the weight of a configuration, \( R_1^2 \) is the norm of the EOM singles amplitudes. 114

4.3 Vertical EOM-CCSD excitation energies of uracil (eV). Ground state CCSD energies are shown in hartree, oscillator strengths are given in parentheses. The singlet CCSD reference wave function was used for both singlet and triplet EOM calculations. 116

4.4 Effect of triple excitations on vertical excitation energies. Shift in excitation energies due to triple excitations is shown in parentheses. All energies are in eV. 119

4.5 Norms of the amplitudes \( R_{m,k} \) from Eq. (1.15) in the EOM-CC(2,3)/6-31G(d) wave function of uracil. The (10,10) active space is used. Total norm is not unity in virtue of biorthogonal properties of EOM-CC. 120

4.6 MRCI excitation energies (eV) of the four lowest singlet excited states of uracil. Oscillator strengths are given in parentheses. 122

4.7 Vertical excitation energies (eV) of uracil in the gas phase and water solution calculated with QM/MM. Energy shifts due to the solvent are shown in parentheses. 124

4.8 Best estimates of the vertical excitation energies (eV) of the lowest singlet and triplet states compared with previously published results. 126

5.1 Geometries and total energies of the intersection seam minima and points of triple degeneracy in \( N_3^+ \) calculated by EOM-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees, all energies are given in a.u. 145
5.2 Results of energy minimization of the $2^1A_2/1^1B_1$ intersection seam in $N_3^+$ calculated by EOM-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees.

5.3 Points of the $C_{2v}$ constrained seam of the $X^2A_1/A^2B_2$ intersection in nitrogen dioxide. All bond lengths are given in angstroms, all angles are given in degrees, all energies are given in a.u.

5.4 Results of $C_{2v}$-constrained energy minimization of the $X^2A_1/A^2B_2$ intersection seam in NO$_2$ calculated by EOMIP-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees.

5.5 Equilibrium geometries and the MECP of the $^1A_g$ and $^3B_{1u}$ electronic states calculated by EOMSF-CCSD/6-31G*. Geometrical parameters are defined in Fig. 5.5. All bond lengths are given in angstroms, all angles are given in degrees.
List of figures

1.1 Calculated adiabatic potential energy curves of the O₂ molecule. At each internuclear distance, the lowest-energy solution of the electronic Schrödinger equation yields the ground state energy. Higher-energy solutions describe electronically excited states. .............................................. 3

1.2 Action of different flavors of the EOM excitation operator forms different sets of target states giving rise to a suite of EOM-CC methods. ........... 12

1.3 The chemical structure and atomic labels of the anionic form of the model GFP chromophore, 4'-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in the cis-conformation. ......................................................... 20

2.1 The chemical structure and atomic labels of the anionic form of the model GFP chromophore, 4'-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in the cis- (top) and trans- (bottom) conformations. ............. 43

2.2 Two resonance structures of the HBDI anion. ........................................ 46

2.3 Two molecular orbitals of the HBDI anion giving rise to the ππ* state. The character of the orbitals can be explained by considering linear combination of two localized π-bonding orbitals. ........................................ 47

2.4 Effect of increasing the number of diffuse functions in the basis set on the density of states and the convergence of the lowest excited and the bright ππ* state. The calculations were performed with CIS (top) and TD-DFT/BNL (bottom). The basis set was varied from 6-311G(2pd,2df) to 6-311(4+,2+)G(2pd,2df). ......................................................... 53

2.5 Computed energy profiles of cis-trans ground-state isomerization of the HBDI anion in the gas phase: (1) IRC calculated with DFT(PBE0)/6-31+G(d,p); (2) MEP calculated with CASSCF(12/11)/cc-pVDZ; (3) MEP calculated with SOS-MP2/cc-pVDZ. ........................................ 62

2.6 Frontier valence molecular orbitals of the HBDI anion in the cis-form (left) and at the transition state of the cis-trans isomerization path (right). 66
2.7 Cis-trans isomerization of the chromophore inside a cluster of water molecules. Relative energies of the stationary points are computed with QM(CASSCF(12/11)/cc-pVDZ)/EFP/MM(TIP3P). 71

3.1 Structures of deprotonated HBDI (HBDI anion, top) and the cation (bottom). The structure of the radical is similar to that of the cation, with the unpaired electron being on the carbon that hosts positive charge in the cation. 85

3.2 Relevant molecular orbitals of deprotonated HBDI (HF/6-311G*). In the ground state of the anion, both $\pi_1$ and $\pi_2$ are doubly occupied, and the bright state is derived by the $\pi_1 \rightarrow \pi^*$ excitation. Oxidized forms are derived by removing the electrons from $\pi_1$. 90

3.3 Energy levels and electronic states of the HBDI anion and radical. The bright absorbing $\pi_1\pi^*$ state at 2.62 eV is a resonance state embedded in a photodetachment continuum beginning at 2.54 eV. The respective MOs are shown in Fig. 3.2. 91

3.4 Electronic states of the doublet HBDI radical. The respective MOs are shown in Fig. 3.2. 91

3.5 Electronic states of the doubly ionized HBDI anion. The respective MOs are shown in Fig. 3.2. 92

3.6 Energy diagram of the relevant electronic states of deprotonated HBDI. 93

3.7 Infrared spectra of the anionic and cationic forms of deprotonated HBDI computed with RI-MP2/cc-pVTZ. 99

4.1 Structure of uracil defining atom labels referred to in Table 4.1. 111

4.2 Frontier molecular orbitals of uracil at the ground state equilibrium geometry. Orbital energies are calculated with RHF/aug-cc-pVDZ. 112

4.3 Excitation energies (upper panel) and oscillator strengths (lower panel) of the lowest excited states of uracil calculated with EOM-CCSD and various one-electron basis sets. 115

4.4 Absorption spectrum of uracil with theoretical estimations of electronic transitions calculated with EOM-CCSD/aug-ANO-DZ and CR-EOM-CCSD(T)/aug-ANO-DZ. 127
5.1 The potential energy surface slices for the ground and the first excited states of a substituted ethylene. There is an energy barrier that prevents the molecule from rotating freely around the double bond. The top of the barrier on the ground state PES corresponds to the minimum on the excited state PES.

5.2 $D_{3h}$-constrained scans of adiabatic PESs of $N_3^+$ around the points of intersections calculated by EOM-CCSD/6-31G.

5.3 The seam of the $X^2A_1/A^2B_2$ intersection in nitrogen dioxide. At every N–O distance, the PESs of both states were scanned to find the O–N–O angle at which the states are degenerate.

5.4 The singlet and the $M_z=0$ triplet target states of the $para$-benzyne diradical derived from the high-spin triplet reference by spin-flip.

5.5 Geometrical parameters used to describe the $para$-benzyne diradical in Table 5.5.

5.6 Convergence of the $^1A_g/ ^3B_{1u}$ intersection optimization in the $para$-benzyne diradical by EOMSF-CCSD/6-31G*. In the upper part, the relative energies of both states are plotted for each iteration. The corresponding energy gaps are shown in the lower part. Convergence is achieved in about 50 iterations.

5.7 Relative energies of the $^1A_g$ and $^3B_{1u}$ states of $para$-benzyne diradical at their equilibrium structures and MECP. Both at the singlet and triplet equilibrium geometry, the singlet is below the triplet. The two states intersect 0.65 eV above the ground state minimum.

6.1 Large biological systems such as this active center of GFP require treatment of the QM core and the MM environment using hybrid QM/MM methods.

A.1 Components of the tensor algebra library.
Abstract

This work demonstrates electronic structure techniques that enable predictive modeling of the properties of biologically relevant species. Chapters 2 and 3 present studies of the electronically excited and detached states of the chromophore of the green fluorescent protein, the mechanism of its cis-trans isomerization, and the effect of oxidation. The bright excited ππ* state of the chromophore in the gas phase located at 2.6 eV is found to have an autoionizing resonance nature as it lies above the electron detachment level at 2.4 eV. The calculation of the barrier for the ground-state cis-trans isomerization of the chromophore yields 14.8 kcal/mol, which agrees with an experimental value of 15.4 kcal/mol; the electronic correlation and solvent stabilization are shown to have an important effect. In Chapter 3, a one-photon two-electron mechanism is proposed to explain the experimentally observed oxidative reddening of the chromophore. Chapter 4 considers the excited states of uracil. It demonstrates the role of the one-electron basis set and triples excitations in obtaining the converged values of the excitation energies of the nπ* and ππ* states. The effects of the solvent and protein environment are included in some of the models.

Chapter 5 describes an implementation of the algorithm for locating and exploring intersection seams between potential energy surfaces. The theory is illustrated with examples from atmospheric and combustion chemistry.
Chapter 1: Introduction and overview

1.1 Potential energy surfaces and their intersections

In electronic structure theory, the full Hamiltonian in Schrödinger’s equation is separated into the nuclear and electronic parts

$$H = H_{\text{nuc}} + H_{\text{el}} = -\frac{1}{2} \nabla^2_R + H_{\text{el}}$$  \hspace{1cm} (1.1)

This decomposition known as Born–Oppenheimer, or adiabatic, approximation\(^1\) is made possible by the separation of the nuclear and electronic degrees of freedom with the assumption that the electrons in the system adjust instantaneously to any change of the nuclei’s positions. Because the nuclei move infinitely slower than the electrons, the electronic part \(H_{\text{el}}\) can be solved at fixed nuclear configurations. The solutions of the electronic Schrödinger equation

$$H_{\text{el}} \Psi_i = E_i \Psi_i$$  \hspace{1cm} (1.2)

are the eigenfunctions \(\Psi_i\) of the electronic Hamiltonian \(H_{\text{el}}\) and are called electronic wavefunctions. Their corresponding eigenvalues \(E_i\) are state energies. The set of all the eigenpairs \(\{(E_i, \Psi_i)\}\), where each pair is identified by a nonnegative integer \(i\), establishes the universe of all possible electronic states of the system at the given nuclear
configuration. Electronic states with the smallest value of energy $E_i$ are called ground electronic states. States with the same value of energy are called degenerate states. It is possible to have degenerate ground electronic states, however in this work we are mainly concerned with the class of systems that have only one ground electronic state.

Each electronic wave function $\Psi_i$ belongs to one of the irreducible representations of the molecule’s point group symmetry. The corresponding state is labeled with that irrep, which is called the “state symmetry.” In addition, each wave function $\Psi_i$ is an eigenfunction of the spin-squared operator $S^2$, and the respective eigenvalue is the “state spin” or “state multiplicity.” States of the same symmetry and spin form manifolds that are uncoupled from each other. This makes it possible to solve Eq. 1.2 with appropriate constraints for each irrep and multiplicity independently.

Let us introduce an order of the electronic states by their energies such that

$$E_i \leq E_j \forall i, j : i < j$$

(1.3)

Then $(E_0, \Psi_0)$ is the ground state.

Solving the electronic Schrödinger equation at every nuclear configuration $R \in \mathbb{R}^{3N}$ ($N$ is the number of atoms, and $3N$ is the total number of three-dimensional Cartesian coordinates) and then ordering the solutions as prescribed by Eq. 1.3 allows us to define functions $V_i : \mathbb{R}^{3N} \rightarrow \mathbb{R}$ that yield the energy of $i$-th state at the given positions of the nuclei $R$

$$E_i|_R = V_i(R)$$

(1.4)

The function $V_i(R)$ is called the adiabatic potential energy surface (PES) of $i$-th electronic state. Examples are shown in Fig. 1.1 (Reproduced from Ref. 2).
Figure 1.1: Calculated adiabatic potential energy curves of the O₂ molecule. At each internuclear distance, the lowest-energy solution of the electronic Schrödinger equation yields the ground state energy. Higher-energy solutions describe electronically excited states.

Consider a situation in which two or more states are degenerate over a set of nuclear configurations

\[ V_i(R) = V_j(R) \quad \forall R \in S \subseteq \mathbb{R}^{3N} \quad i \neq j \]  

(1.5)

Set \( S \) can be all possible nuclear positions \( S = \mathbb{R}^{3N} \), but that case is not of particular interest. Let us look at the case where \( S \subset \mathbb{R}^{3N} \). The set \( S \) defines the subspace of
degeneracy of the states, and we can define a function $U_i : S \rightarrow \mathbb{R}$, which we call the intersection seam.

The sets $S$ and $\mathbb{R}^{3N}$ are not completely independent. The Neumann–Wigner theorem\(^3\) (the noncrossing rule) states that two conditions have to be imposed on the Hamiltonian in order for two states of the same symmetry to be degenerate. In diatomic molecules with only one internal degree of freedom, PES of the same symmetry may not intersect (hence “noncrossing rule”). In larger molecules the rule limits the dimensionality of the intersection hyperline to $M - 2$ where $M$ is the number of internal degrees of freedom. For states of different symmetry or spin, there is only one condition (degeneracy), and therefore such crossings are possible even in diatomic molecules. Section 5.2 presents more details on the theory of PES intersections.

PES crossings occur often in real systems\(^4\). Because they allow the system to start on one state’s PES and continue on another’s, they facilitate radiationless transitions between electronic states. In Chapter 5 we present an implementation of algorithms for locating and optimizing intersection seams and demonstrate their work with examples.

### 1.1.1 Methods for the ground electronic state

The ground state is of particular interest in electronic structure theory because chemical reactions mostly occur on the lowest-energy PES. One of the conventional approaches to approximately solve for the ground state energy is to use Hartree–Fock method, which employs variational principle and is based on the mean-field description of electron–electron interactions. Using a set of one-electron basis set functions, one solves a system of Roothaan equations self-consistently to obtain molecular orbitals and their energies.

Hartree–Fock method can be roughly described as taking the following steps. The reader is encouraged to consult electronic structure texts\(^5,6\) for more details.
1. A one-electron basis set is selected for each atom of the molecule. In this work we use Pople’s split-valence\textsuperscript{7–9} and Dunning’s\textsuperscript{10} collections of basis sets, which are based on the Cartesian Gaussian representation of Slater-type orbitals. Using the basis set, orbitals of each type ($s$, $p$, $d$, etc.) are created. They are called atomic orbitals $\phi_\mu$.

2. Initial molecular orbital matrix $C$ and Fock matrix $F$ are formed using core Hamiltonian, superposition of atomic densities, Hückel theory or another similar strategy. The orbital overlap matrix $S$ is computed.

3. Roothaan equation

$$FC = SCE$$

is solved for the matrix $C$ and the diagonal matrix $E$. Because Fock matrix $F$ itself depends on $C$, the equation has to be solved self-consistently. Iterations continue until the convergence criterion (usually $|C^{(i)} - C^{(i-1)}| < \varepsilon$) is met.

Resulting from this procedure, the $C$ matrix contains in its rows the expansion coefficients for obtaining molecular orbitals $\psi_i$ in terms of atomic orbitals

$$\psi_i = \sum_\mu c_{i\mu} \phi_\mu$$

(1.6)

and the diagonal matrix $E$ contains corresponding molecular orbital energies.

Hartree–Fock theory does not include electronic correlation, but provides a starting point for correlated methods.
**Coupled-clusters theory**

Coupled-clusters (CC) theory offers a size-extensive approach to include electronic correlation into the description of the ground state. Similar to the configuration interaction (CI) method, CC theory represents the wave function of the ground state as an expansion in terms of the reference determinant, then all determinants with single substitutions, followed by double substitutions, and so on. Unlike CI, which uses a linear excitation operator to generate the expansion, CC theory introduces an exponential ansatz:

\[
|\Psi_{CC}\rangle = \exp(T)|\Phi_0\rangle
\]

where \(|\Phi_0\rangle\) is called the reference determinant (usually Hartree–Fock wavefunction).

The cluster operator \(T\) is a series of \(n\)-tuple excitation operators:

\[
T = T_1 + T_2 + T_3 + \ldots
\]

Each term in the series can be written in the second quantization form:

\[
T_1 = \sum_{i a} t^a_i a^\dagger_a a_i \\
T_2 = \frac{1}{4} \sum_{i j a b} t^{a b}_{i j} a^\dagger_a a^\dagger_b a_j a_i \\
T_3 = \frac{1}{6} \sum_{i j k a b c} T^{a b c}_{i j k} a^\dagger_a a^\dagger_b a^\dagger_c a_k a_j a_i
\]

and so on. These equations follow the conventional notation: \(i, j, \ldots\) stand for occupied in the reference determinant \(|\Phi_0\rangle\) spin-orbitals, \(a, b, \ldots\) designate virtual (unoccupied) spin-orbitals. The creation and annihilation operators \(a^\dagger_p\) and \(a_p\) create or remove an electron from the spin-orbital \(|p\rangle\).
The wave function and energy are obtained by solving the system of CC equations for the $T$ amplitudes:

$$\langle \Phi_0 | \exp(-T) H \exp(T) | \Phi_0 \rangle = E_{CC}$$

$$\langle \Phi_0 | \exp(-T) H \exp(T) - E_{CC} | \Phi_i \rangle = 0$$

When all the terms of the expanded $T$ operator are included in the CC model, it becomes completely equivalent to the full CI series as all possible excited determinants are present in the expansion. The CC energy in that case contains 100% of the electronic correlation energy within the given basis set.

Using the full expansion of the cluster operator $T$, however, is impractical because the computational complexity of the method grows exponentially with the size of the system. Full CI results have only been reported for di- and triatomic molecules in rather small basis sets.

To reduce the scaling of the CC method, the $T$ operator is truncated. Limiting the expansion of $T$ to the first two terms, for example, gives rise to coupled-clusters with single and double excitations (CCSD) method\textsuperscript{11}

$$T = T_1 + T_2$$

Because of the exponential form of the wave function ansatz, CC theory retains size-extensivity even with the truncated cluster operator. This is unlike CI, in which only the full expansion is size-extensive. In addition, as evident from the Taylor expansion of
exp(T), the CCSD wavefunction effectively includes contributions from higher than single and double excitations:

\[
\exp(T_1 + T_2) = 1 + T_1 + T_2 + \frac{1}{2}(T_1 + T_2)^2 + \ldots = 1 + T_1 + T_2 + \frac{T_1^2}{2} + \frac{T_2^2}{2} + \ldots
\]

This argument is applicable to any level of truncation of the \( T \) operator.

These two qualities make the truncated CC models superior to their CI counterparts. Among all iterative CC methods, CCSD provides the most attractive balance between the computational cost and accuracy. When chemical accuracy is desired (orders of a few kcal/mol), the energy can be further improved using a non-iterative CCSD(T) correction via perturbation theory\(^{12}\).

In practice the CC equations are solved numerically using the Jacobi method for the system of linear equations in the matrix form. The vector solution is the CC amplitudes and the update procedure is formulated as a recurring expression for the \( T \) vector\(^{13}\). To improve the convergence of the scheme, the direct inverse of the iterative subspace (DIIS)\(^{14}\) method is applied.

**Multireference Møller–Plesset perturbation theory (MRMP2)**

The MRMP2 model\(^{15}\) is a special single-state case of multiconfigurational quasidegenerate second order perturbation theory (MCQDPT2)\(^{16}\) implemented in GAMESS(US)\(^{17}\) and PC GAMESS\(^{18}\). The zeroth-order (reference) wave functions are state-averaged CASSCF wave functions (\( \Psi_{\alpha}^{\text{CAS}} \)) for the target state \( \alpha \). The unperturbed Hamiltonian operator is a sum of one-electron Fock-like operators, in which the occupation numbers are replaced by the diagonal elements of the CASSCF density matrix. Beyond zeroth-order multiconfigurational functions, the complementary eigenfunctions of the complete
active space configuration interaction Hamiltonian, as well as the wave functions generated by one- and two-electron excitations from the reference functions (S-space), are considered via perturbation theory. The second-order corrections to the energy are given by:

$$E^{(2)}_{\alpha} = \sum_j \frac{\langle \Psi_{\text{CAS}}^{\alpha} | V | \Phi_j \rangle \langle \Phi_j V | \Psi_{\text{CAS}}^{\alpha} \rangle}{E_\alpha^{(0)} - E_j^{(0)}}$$  \hspace{1cm} (1.7)

where functions $\Phi_j$ belong to the S-space of uncontracted determinants. In a similar approach, CASPT2\textsuperscript{19}, the S-space consists of the contracted configurations. Both MRMP2 and CASPT2, diagonalize-then-perturb methods, are widely used to calculate the excitation energies of organic chromophores. A comprehensive benchmark study\textsuperscript{20} reports excellent agreement between zeroth-order-corrected CASPT2\textsuperscript{21} and an accurate approximation to the coupled-cluster with singles, doubles, and triples method (CC3)\textsuperscript{22,23}. Similar accuracy is observed in the recent computational studies employing the MRMP2 methodology\textsuperscript{16,24,25}.

**Scaled opposite-spin MP2**

Spin-component-scaled MP2 (SCS-MP2) is a semi-empirical approach based on scaling different spin contributions to the MP2 correction as proposed by Grimme\textsuperscript{26}:

$$E_c \approx E_{c,\text{SCS-MP2}} = p_{ss} E_{ss}^{(2)} + p_{os} E_{os}^{(2)}$$  \hspace{1cm} (1.8)

The parallel-spin component $E_{ss}^{(2)}$ and the antiparallel-spin component $E_{os}^{(2)}$ are scaled to correct for their unbalanced contributions to the MP2 correlation energy. Empirically found optimal values of the coefficients are $p_{ss} = \frac{1}{3}$ and $p_{os} = \frac{6}{5}$. SCS-MP2 provides an improved description for many systems, in which the ground state has a single-reference character.
As the opposite-spin contribution is the major one, Jung et al.\textsuperscript{27} further simplified the model by dropping the same-spin term altogether and scaling the opposite-spin contribution up. Along with the RI technique, SOS-MP2 offers a significant improvement in computational performance compared to the original MP2. The scaling coefficient — the only empirical parameters in the method — can be optimized for a wide variety of systems\textsuperscript{28}.

1.1.2 Methods for excited electronic states

Equation-of-motion method

Applying the general linear excitation operator $R$ on the ground state CC wave function yields the excited state wave function

$$|\Psi^{(m)}\rangle = R^{(m)}|\Psi_{CC}\rangle = R^{(m)}\exp(T)|\Phi_0\rangle$$

(1.9)

(the integer $m$ identifies the excited state). The structure of the excitation operator $R$ is the same as in CI. It can be expanded as

$$R = R_1 + R_2 + R_3 + \ldots$$

(1.10)

where $R_1$ generates all possible singly excited with respect to the reference determinants, $R_2$ generates all double excited determinants, and so on.

In the limit of the zero cluster operator $T = 0$, this approach becomes regular CI as the excitation operator $R$ acts directly on the reference determinant. If the series for $R$ is not truncated, the operator generates all possible excitations. In this case the full CI result is recovered, no matter what the cluster operator $T$ is.
The idea can be reformulated and generalized by introducing a similarity transformation for the electronic Hamiltonian $H$ using the cluster operator:

$$\tilde{H} = \exp(-T)H\exp(T)$$  \hspace{1cm} (1.11)$$

The similarity transformed Hamiltonian $\tilde{H}$ has the same spectrum as the original Hamiltonian $H$. That allows us to “fold” the electronic correlation in the form of the cluster operator into the Hamiltonian and find the eigenvalues (target state energies) and eigenvectors (target state wave functions) in various subspaces of the Fock space. This approach is called equation-of-motion coupled-clusters (EOM-CC) method$^{29-37}$.

If the excitation operator $R$ promotes one or more electrons from occupied to virtual orbitals, it can be written in the second quantization form as

$$R_0^{(m)} = r_0 \hspace{1cm} R_1^{(m)} = \sum_{ia} r_i^a a_i^+ a_i \hspace{1cm} R_2^{(m)} = \frac{1}{4} \sum_{ijab} r_{ij}^{ab} a_i^+ a_b^+ a_j a_i$$  \hspace{1cm} (1.12)$$

This form of $R$ preserves the number and spin of electrons. It is used in EOM-CC for excitation energies (EOMEE-CC), which yields the energies and wave functions for electronically excited states.

Operator $R$ can be designed to remove or attach electrons, change their spin (Fig. 1.2). This gives rise to a suite of EOM-CC methods: for ionization energies (EOMIP-CC), for attachment energies (EOMEA-CC), spin-flip (EOMSF-CC). It is a powerful toolkit that allows one to start with a convenient well-behaved reference state, and form target states by applying the different flavors of the operator $R^{38}$. 

11
Figure 1.2: Action of different flavors of the EOM excitation operator forms different sets of target states giving rise to a suite of EOM-CC methods.
EOM operator $R$ is written in the tensor form in the basis of molecular orbitals. The elements of this tensor ($r_i^a$, $r_i^{ab}$ in Eq. 1.12) are called EOM amplitudes. They are found by solving the eigenproblem

$$
\langle \Phi_\mu | \bar{H} - E^{(m)} | R^{(m)} \Phi_0 \rangle = 0
$$

(1.13)

where $\langle \Phi_\mu |$ are all $\mu$-tuply excited with respect to the reference determinants.

Just like the cluster operator $T$ in CC, the operator $R$ can be truncated at different excitation levels. In EOM-CCSD$^{31,32,39,40}$, both $T$ and $R$ include terms up to $T_2$ and $R_2$. In the EOM-CC method with single, double, and full triple excitations (EOM-CCSDT)$^{41-44}$, the series for $T$ and $R$ are truncated after $T_3$ and $R_3$. It provides superb quality of wave functions, but remains prohibitively expensive for all but tiny systems. The EOM-CC(2,3)$^{45,46}$ approximation is obtained when $R$ is truncated after $R_3$, and $T$ is the same as in EOM-CCSD. The CC2 and CC3 methods are iterative approximations to full EOM-CCSD and EOM-CCSDT, respectively$^{47}$.

Explicit inclusion of triples, as in EOM-CCSDT or EOM-CC(2,3), results in the $O(N^8)$ scaling versus the $O(N^6)$ scaling of EOM-CCSD, where $N$ is the number of one-electron basis set functions. Computational cost can be reduced by using a restricted subset of triple excitations via an active space, where the three-body operators $T_3$ and $R_3$ are replaced by operators $t_3$ and $r_3$, which are required to include at least one excitation within the active space$^{41,46,48}$. CC2 and CC3 scale as $O(N^5)$ and $O(N^7)$, respectively.

In the active-space CCSDt and EOM-CCSDt methods$^{49-60}$, the cluster operator $T'$ for CCSDt and the excitation operator for the $m$-th electronic state $R'$ for EOM-CCSDt are given by

$$
T' = T_1 + T_2 + t_3
$$

(1.14)
\[ R' = R_0 + R_1 + R_2 + r_3 \]  \hspace{1cm} (1.15)

Operators \( t_3 \) and \( r_3 \) are

\[ t_3(I) = \sum_{i \leq j \leq K} \sum_{A \leq b \leq c} i_{iK}^j \sum a_A^i a_b^j a_c^a K a_{j} a_i, \]  \hspace{1cm} (1.16)

\[ r_3(I) = \sum_{i \leq j \leq K} \sum_{A \leq b \leq c} r_{iK}^{j} \sum a_A^i a_B^j a_c^a K a_{j} a_i \]  \hspace{1cm} (1.17)

Capital letter indices denote orbitals from the active space. This form of the \( t_3 \) and \( r_3 \) operators defines variant I of CCSDt and EOM-CCSDt, or CCSDt(I) and EOM-CCSDt(I). The size of the operators can be reduced further by constraining two pairs of spin-orbitals to the active space

\[ t_3(II) = \sum_{i \leq J \leq K} \sum_{A \leq B \leq c} i_{iJ}^K \sum a_A^i a_B^j a_c^a K a_{j} a_i, \]  \hspace{1cm} (1.18)

\[ r_3(II) = \sum_{i \leq J \leq K} \sum_{A \leq B \leq c} r_{iJ}^K \sum a_A^i a_B^j a_c^a K a_{j} a_i, \]  \hspace{1cm} (1.19)

giving rise to the CCSDt(II) and the EOM-CCSDt(II) models. In CCSDt(III) and EOM-CCSDt(III), all the spin-orbitals in \( t_3 \) and \( r_3 \) belong to the active space, which results in the lowest memory and time requirements.

The active-space variant of EOM-CC(2,3), the EOM-CC(2,\( \tilde{3} \)) method, is based on the CCSD reference and includes internal and semi-internal triple excitations via the \( r_3 \) operator, as given by Eqs. (1.15) and (1.17).

The EOM-CCSD methods give an error in the range of 0.1–0.3 eV for excitation energies, and including triples reduces it to 0.01–0.02 eV, as was shown for small molecules\(^{61} \). A compilation of benchmarks by Schreiber et al.\(^{20} \) estimates errors in the excitation energies of singlet states for CC2, EOM-CCSD, CC3, and EOM-CCSDT:
mean absolute errors are 0.486 eV, 0.12 eV, 0.016 eV, 0.029 eV, and maximum errors are 1.08 eV, 0.23 eV, 0.047 eV, and 0.083 eV for the four methods, respectively.

Full-space EOM-CC(2,3) results closely follow those obtained with EOM-CCSDT\textsuperscript{45,46}. However, the EOM-CC(2,3) method for excitation energies, EOMEE-CC(2,3), is not size-intensive, which may result in increased error bars for larger molecules. Slipchenko and Krylov\textsuperscript{46} report encouraging results for the active-space EOM-CC(2,\~{3}), e.g., they observe errors on the order of 0.01 eV and 0.1 eV for singly and doubly excited states, respectively. Kowalski \textit{et al.}\textsuperscript{59} show that for states with the singly excited character in small benchmark systems, EOM-CCSDt(I), (II), and (III) give almost the same accuracy with the difference between EOM-CCSDt(I) and EOM-CCSDt(III) excitation energies not exceeding 0.1 eV. The active space for EOM-CC calculations with large basis sets has to be chosen with care, as diffuse orbitals often appear below the frontier valence orbitals.

\textbf{Non-iterative triples corrected EOM-CCSD(T)}

A calculation of accurate excitation energies requires a proper account for both non-dynamical and dynamical correlation. Methods that include active-space triple excitations like EOM-CC(2,\~{3}) or EOM-CCSDt are capable of recovering non-dynamical correlation arising from the doubly-excited character of a target state or a multi-configurational reference. A small magnitude of this correction indicates that the dynamical correlation component dominates in the overall effect. That can be accounted for accurately by including a complete set of higher excitations or less expensively through a non-iterative perturbative triples correction\textsuperscript{22,47,62–71}.

The completely renormalized EOM-CCSD with non-iterative triples [CR-EOM-CCSD(T)]\textsuperscript{68,69} is based on the methods of moments of coupled-cluster equations
(MMCC)\textsuperscript{72,73}. It includes an approximate treatment of triple excitations and has the $O(N^7)$ scaling like CCSD(T)\textsuperscript{12}.

The CR-EOM-CCSD(T) correction $\delta_m$ for the $m$-th electronic state is:

$$
\delta_m = \frac{\langle \Psi_m(3) | M_{m,3}(2) | \Phi_0 \rangle}{\langle \Psi_m(3) | \Psi_{EOM} \rangle},
$$

(1.20)

where the $M_{m,3}(2)$ operator corresponds to triply excited moments of the EOM-CCSD equations for the $m$-th state,\textsuperscript{72} $|\Psi_{EOM}^{m}\rangle$ and $|\Psi_{m}(3)\rangle$ are the EOM-CCSD and trial wave functions, respectively. The latter takes the form:

$$
|\Psi_{m}(3)\rangle = (P + Q_1 + Q_2 + Q_3 \left( R_{m,0} + R_{m,1} + R_{m,2} + \tilde{R}_{m,3} \right) e^{T_1 + T_2} |\Phi_0 \rangle,
$$

(1.21)

where $T_1, T_2, R_{m,0}, R_{m,1}, R_{m,2}$ are the cluster and excitation operators from the CCSD and EOM-CCSD wave functions, and the $\tilde{R}_{m,3}$ operator approximates the exact EOM-CCSDT $R_{m,3}$ operator. The $P$ and $Q_m$ ($m = 1, 2, 3$) operators are projection operators onto the reference Hartree–Fock determinant $|\Phi_0 \rangle$ and the subspace of all $m$-tuply ($m = 1, 2, 3$) excited configurations, respectively. The amplitudes defining the $\tilde{R}_{m,3}$ operator are expressed in terms of the triply excited moments\textsuperscript{68,69}. As demonstrated previously\textsuperscript{74,75}, good estimates of vertical excitation energies are obtained by adding these corrections to corresponding EOM-CCSD excitation energies.

Second-order perturbation treatment of the similarity transformed Hamiltonian $\hat{H}$ yields two variants of non-iterative triples energy corrections: EOM-CCSD(dT) and EOM-CCSD(fT), diagonal and Fock triples, respectively\textsuperscript{76}. The corrections are introduced for EOMEE-CCSD and EOMSF-CCSD methods, and the spin-flip variant is size-intensive.
Multi-reference configuration interaction

In MRCI methods, the wave function is expanded as:

$$\lvert \Psi^{\text{MRCI}} \rangle = \sum_{L=1}^{N_{\text{CSF}}} c_L \lvert \psi_L \rangle .$$

(1.22)

The basis functions of the expansion \{\lvert \psi_L \rangle \} are configuration state functions (CSFs). The CSFs, which are linear combinations of Slater determinants, are eigenfunctions of spin operators and have a correct spatial symmetry\(^7\). The MRCI energies and wave functions are obtained by diagonalizing the Hamiltonian in the basis of the CSFs.

MRCI calculations begin by determining zero-order wave functions and MOs in a CASSCF calculation. The CASSCF wave function includes configurations created by all possible excitations within an active orbital space. The coefficients of the expansion (1.22), \(c_L\), and the MOs are variationally optimized. Dynamical correlation is then described by MRCI by including single and double excitations from the reference CASSCF configurations. The MRCI method is very accurate, provided that all the important configurations are included in the expansion. This requirement is easily satisfied for small molecules, but as the size of the system increases, the expansion becomes prohibitively large, and truncations are necessary. The Davidson correction\(^8\) provides a simple formula for evaluating the contribution of the missing quadruple excitations to the energy computed with configuration interaction with up to double excitations.

Scaled opposite-spin CIS(D)

In the same manner as SOS-MP2 is introduced for correcting the ground state energy, SOS-CIS(D) is designed for excitation energies\(^9\). The computational scaling of SOS-CIS(D) is only \(O(N^4)\), which is a significant improvement over the \(O(N^5)\) scaling of
CIS(D). The accuracy of SOS-CIS(D) is very similar to that of CIS(D) for valence states, whereas the performance for the Rydberg states is improved. Based on a set of over forty various excited states in over twenty organic molecules, the mean signed error in the SOS-CIS(D) vertical excitation energy is 0.02 eV for valence states and −0.08 eV for Rydberg transitions. Limitations of SOS-MP2 and SOS-CIS(D) are the same as MP2 and CIS(D), respectively. For example, these methods fail when the ground-state wave function acquires a significant multiconfigurational character, as at a cis-trans isomerization transition state, and for excited states with considerable doubly excited character. Open-shell (e.g., doublet) states can also cause difficulties due to spin-contamination.

1.1.3 Density functional theory methods

Long-range-corrected density functionals

In long-range-corrected functionals, a range-separated representation of the Coulomb operator\(^{80,81}\) is used to mitigate the effects of the self-interaction error. The contribution from the short-range part is described by a local functional, whereas the long-range part is described using the exact Hartree-Fock exchange. The separation depends on a parameter \(\gamma\). In the BNL approach\(^{82}\), \(\gamma\) is optimized for each system using Koopmans-like arguments: \(\gamma\) is adjusted such that the HOMO energy equals the difference between the total BNL energies of the \(N\)- and \(N\)-electron systems. Initial benchmarks\(^{82,83}\) demonstrated an encouraging performance for excited states, and even such challenging systems as ionized dimers. In \(\omega\)PB97X\(^{84}\), \(\gamma\) and other parameters are optimized using standard training sets. Benchmark results have demonstrated consistently improved performance relative to non-long-range-corrected functionals.
1.2 Previous studies of the green fluorescent protein chromophore

Unique electronic properties of the green fluorescent protein (GFP) whose natural function is to convert blue light to green light have motivated a number of experimental and theoretical studies\textsuperscript{85,86} and have been exploited in numerous practical applications\textsuperscript{87–89}. Due to their fundamental and practical importance, studies of the structure and properties of photoreceptor proteins and their denatured chromophores constitute an important field of modern research. Moreover, GFP can be considered as a model for other fluorogenic unsymmetric methine dyes\textsuperscript{90–95}, and is of interest to organic photovoltaic materials. For example, the fluorescent protein motif has already inspired the creation of new organic photovoltaic sensitizers\textsuperscript{96} and other optoelectronic materials\textsuperscript{97}.

From the theoretical perspective, characterization of the electronic structure of isolated chromophores is the first step towards understanding their photochemical and photobiological properties in realistic environments. Modeling isolated species involves calculating the molecular parameters of the chromophores in the gas phase and in solution using quantum chemistry methods. This work presents accurate calculations of the properties of biological chromophores with ab initio methods using the model GFP chromophore, 4’-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) anion, as a benchmark system (Fig. 1.3). We also study the vertical excitation and electron detachment energies, discuss the electronic properties of the excited and detached states, the \textit{cis-trans} isomerization of the HBDI anion in the ground electronic state.
Figure 1.3: The chemical structure and atomic labels of the anionic form of the model GFP chromophore, 4’-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in the cis-conformation.

Studies of excited states of HBDI

Earlier experimental studies have characterized the absorption of HBDI in the native protein environment\textsuperscript{98}. The spectrum of the wild-type GFP has two broad absorption bands at 396 nm (3.13 eV) and 476 nm (2.60 eV) assigned to the neutral and anionic forms of the chromophore, respectively. The spectra in aqueous solution\textsuperscript{99,100} reveal strong pH sensitivity: the absorption maximum at neutral pH is at 370 nm (3.35 eV), whereas at pH=13 and pH=1 it is shifted to 426 nm (2.91 eV) and 396 nm (3.13 eV), respectively. The absorption of denatured wild-type proteins exhibit similar pH-dependence\textsuperscript{101}. The shifts were attributed to different protonation and deprotonation forms, as well as strong interaction with water. The latter is consistent with a large change (about 7 D) of the dipole moment upon excitation determined in Stark effect measurements taken in a buffered at pH=6.5 glycerol solution at 77 K\textsuperscript{102}. Therefore, the protonation and deprotonation of HBDI and solvent effects, which strongly affect its electronic properties, need to be accounted for in theoretical models. However, it is important to characterize the chromophore in the gas phase first in order to quantify the solvent effect separately.
Recently, a gas-phase action spectrum of anionic HBDI (as well as other protonated forms) using photodestruction spectroscopy of mass-selected ions injected into an electrostatic ion-storage ring was reported providing an important reference for theory\textsuperscript{99,100}. The spectrum shows an absorption band centered at 2.59 eV (479 nm), which extends from 2.4–2.8 eV (440–520 nm), as well as a minor peak around 2.3 eV (540 nm). The authors emphasized a striking similarity between the absorption bands in the gas phase and in the protein and suggested that the protein environment shields the chromophore from water and that the absorption in the protein is an intrinsic property of HBDI. While the role of the protein still needs to be investigated, this measurement facilitates more direct comparison between the gas-phase calculations of vertical excitation energies and the experimental absorption for benchmarking theoretical methods.

The large absorption band in the gas phase spectrum of the HBDI anion has been assigned as the $\pi\pi^*$ transition, however, the nature of the minor feature at 2.3 eV has not been discussed.

A variety of electronic structure techniques ranging from simple semiempirical approximations to high-level ab initio methods have been applied to simulate the properties of the cis-anionic form of HBDI\textsuperscript{103–109}. Selected representative results are summarized in Table 1.1.

These studies have identified the absorbing state of the HBDI anion as the $S_1$ state derived from a HOMO-LUMO excitation of the $\pi\pi^*$ character. LUMO, however, is a valence $\pi^*$-like orbital only in relatively small basis sets; including diffuse functions increases the number of molecular orbitals between the HOMO and the lowest $\pi^*$-like orbital. Though the bright state retains its $\pi\pi^*$ character, it is not, strictly speaking, a HOMO-LUMO transition in a realistic basis set.
Table 1.1: Selected theoretical estimates of the $S_0$-$S_1$ vertical excitation energy ($\Delta E$) and the corresponding wavelengths for the gas phase GFP chromophore. The experimental value is also given.

<table>
<thead>
<tr>
<th>Method to compute the excitation energy</th>
<th>Method to optimize ground-state geometry</th>
<th>$\Delta E$ (eV)</th>
<th>Wavelength (nm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDO-CI</td>
<td>PM3</td>
<td>2.77</td>
<td>448</td>
<td>Ref. 103$^a$</td>
</tr>
<tr>
<td>CIS/6-31G(d)</td>
<td>RHF/6-31G(d)</td>
<td>4.37</td>
<td>284</td>
<td>Ref. 105</td>
</tr>
<tr>
<td>TD-DFT(BP86)/6-31++G(d,p)</td>
<td>B3LYP/6-31++G(d,p)</td>
<td>2.94</td>
<td>421</td>
<td>Ref. 108$^b$</td>
</tr>
<tr>
<td>CASPT2/6-31G(d)</td>
<td>CASSCF(12/11)/6-31G(d)</td>
<td>2.67</td>
<td>465</td>
<td>Ref. 107</td>
</tr>
<tr>
<td>SAC-CI/DZV</td>
<td>B3LYP/6-31G(d)</td>
<td>2.22</td>
<td>558</td>
<td>Ref. 106</td>
</tr>
<tr>
<td>MRMP2 based on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA-CASSCF(16/14)/(aug)-cc-pVDZ</td>
<td>PBE0/aug-cc-pVDZ</td>
<td>2.47</td>
<td>501</td>
<td>Ref. 109</td>
</tr>
<tr>
<td>aug-MCQDPT2 based on</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA-CASSCF(16/14)/(aug)-cc-pVDZ</td>
<td>PBE0/aug-cc-pVDZ</td>
<td>2.54</td>
<td>489</td>
<td>Ref. 109</td>
</tr>
<tr>
<td>Experiment</td>
<td></td>
<td>2.59</td>
<td>479</td>
<td>Ref. 99, 100</td>
</tr>
</tbody>
</table>

$^a$ A close value (444 nm) was obtained in the later semiempirical calculations of the NDDO type (Ref. 104). $^b$ For an overview of previous TD-DFT calculations using different functionals and basis sets, see Table 1 of Ref. 108; only the value closest to the experimental excitation energy is presented here.
The first theoretical studies on the chromophore, dating back to the late 1990s, employed semiempirical methods based on the neglect-of-differential overlap approximation. They placed the $\pi\pi^*$ state at 2.77 eV, or 0.18 eV above the experimental absorption maximum. Ab initio calculations using configuration interaction with single excitations (CIS) in a small basis set grossly overestimated the excitation energy. These and other results from Table 1.1 reveal that accurate calculations of the vertical excitation energy for the bright $\pi\pi^*$ transition are challenging for quantum chemistry.

Time-dependent density functional theory (TD-DFT), which can be applied to very large systems, gave rise to high expectations in the field of photochemistry. However, the notorious self-interaction error results in an unphysical description of charge-transfer (CT) states, which are common in large molecules. In addition to artificially low excitation energies of real CT states, spurious CT states appear, spoiling the description of other states. The number of these false states increases steeply with the system size. In the case of fluorescent protein chromophores, TD-DFT has been reported to perform quite modestly, as discussed in Ref.108. In that study, which examined various functionals, the best agreement with the experiment was obtained for the BP86 functional. This value (2.94 eV or 421 nm), which is listed in Table 1.1, is still 0.35 eV (60 nm) away from the experiment. Overall, TD-DFT excitation energies range from 2.94 to 3.23 eV, the popular B3LYP functional yields 3.05 eV.

CASPT2, the complete active space self-consistent field (CASSCF) method with second-order perturbation theory (PT2) correction, has been applied to model photochemical properties of organic chromophores in various media. The CASPT2/6-31G(d) result from Table 1.1 agrees fairly well with the experimental value; however, the CAS distribution of 12 electrons over 11 orbitals is a truncation of the entire $\pi$-orbital active space, which requires the CASSCF(16/14) wave function.
for the GFP chromophore. Bravaya et al. performed calculations of the vertical ππ* transition energies for various forms of HBDI using a very expensive and highly correlated approach, i.e. state-averaged CASSCF(16/14) wave functions constructed in the full π-orbital active space augmented by perturbative corrections: multireference second-order Möller-Plesset perturbation theory (MRMP2) and an extended version of the multiconfigurational quasidegenerate perturbation theory (aug-MCQDPT2) (using ground-state equilibrium geometries optimized using DFT with the PBE0 functional and the (aug)-cc-pVDZ basis set: aug-cc-pVDZ on oxygen atoms and cc-pVDZ on all other atoms). The MRMP2 and aug-MCQDPT2 results are within 0.12 and only 0.05 eV from the experimental value, respectively (Table 1.1). These data suggest that one can compute the positions of absorption bands of biological chromophores with an accuracy of 10–20 nm (less than 0.1 eV) by applying perturbatively corrected CASSCF-based approaches. These techniques, however, are computationally demanding, and their execution requires advanced skills and extreme care, as the application of the method involves: (i) a careful selection of a large number of active space orbitals in fairly large basis sets; (ii) converging the state averaged CASSCF solutions corresponding to the ππ* transition, especially in realistic basis sets; (iii) a careful and often ambiguous treatment of perturbative corrections to the reference CASSCF solutions. Moreover, gradient calculations are only available for bare CASSCF wave functions. Thus, it is desirable to find a more robust approach of a comparable accuracy for wider applications in modeling the properties of biological chromophores.

Contrarily to the bright singlet ππ* state, little is known about the triplet states of the GFP chromophore. The fluorescence properties of GFP suggest that the intersystem crossing is not efficient, at least in the chromophore in the native protein environment. In
the gas-phase photocycle, however, triplet states can play an important role. For example, possible population trapping in the triplet has been suggested as a possible explanation for observed millisecond-long lifetimes of the photoexcited ions in the ion-storage ring experiments\textsuperscript{131}. Spin-forbidden relaxation channels have also been considered in the studies of GFP mutants\textsuperscript{132,133}.

**Cis-trans isomerization of HBDI**

The *cis-trans* isomerization (or Z/E diastereomerization) of the photoswitchable fluorescent chromophores plays an essential role in their photophysical properties. For example, some photoconversion mechanisms involve *cis-trans* isomerization\textsuperscript{134,135}. Kindling and blinking phenomena, as well as loss of fluorescence yield of bare chromophores in solution, are also related to this process. In a broader context, the photophysics of GFP is similar to that of other fluorogenic unsymmetric methine dyes\textsuperscript{90–95}, and is of interest with connection to organic photovoltaic materials.

The majority of the experimental and computational studies of this process focused on the excited state dynamics. However, the ground electronic state potential energy surface (PES) also needs to be considered for two main reasons. First, after photoisomerization has completed, the photoswitchable protein returns to its initial state to prepare for the next cycle. The recovery apparently takes place on the ground state PES. Second, the details of the ground-state isomerization can elucidate the chromophore’s rearrangements along the more complicated excited-state route. Despite previous studies of the *cis-trans* photoisomerization of GFP-like chromophores\textsuperscript{136–141} using quantum chemistry modeling\textsuperscript{103,104,126,142–145}, many questions remain unanswered.

The process of *cis-trans* isomerization of the HBDI anion in aqueous solution was investigated experimentally by He et al.\textsuperscript{136} who estimated the free energy differences
and the activation energies using the NMR technique. The authors stressed that the activation energy of 15.4 kcal/mol derived from their measurements in aqueous solution is in distinct disagreement with the results of calculations,\textsuperscript{103,104} which estimated that the barrier is above 21 kcal/mol. The relatively low value of the barrier (as compared to other isomerization reactions involving exocyclic double bonds) has also been emphasized in subsequent studies of the isomerization and several explanations have been suggested\textsuperscript{137,141}. For example, thermal isomerization studies of model GFP-like compounds\textsuperscript{137} suggested that different substituent groups may have a significant effect on the activation energy by changing the interaction between two resonance structures of the chromophore. Tolbert and coworkers considered mechanisms involving changes in chemical structure of the chromophore, e.g., addition/elimination pathway\textsuperscript{141}. No ab initio calculations have been reported so far to resolve this disagreement between experimental measurements and theoretical estimates and explain the low value of the barrier.

1.3 Previous studies of the excited states of uracil

Natural nucleobases—adenine, guanine, thymine, cytosine, and uracil—combine with residues of phosphoric acid and sugars to form nucleotides, the monomers of nucleic acids. Being UV chromophores, the nucleobases define a large portion of the RNA and DNA photochemistry and are used as model systems to study the properties of the polymers\textsuperscript{146–149}.

UV radiation is harmful to living organisms: when absorbed by the nucleic acid, it initiates excited-state dynamics that can result in structural damage. The process, which starts with an electronic excitation of a UV chromophore, can be quenched by radiationless relaxation to the ground state. Photoexcited natural nucleobases have lifetimes
of less than one picosecond, and that is believed to be responsible for the remarkable photostability of DNA. In a review, Crespo-Hernández et al.\textsuperscript{146} discuss mechanisms, pathways, and dynamics of the relaxation process.

The mechanism of the excited state population decay varies from nucleobase to nucleobase and depends on the order of the lowest electronically excited states, which correspond to a forbidden $n\pi^*$ and an allowed $\pi\pi^*$ transition. Areas of potential energy surfaces (PESs) where the ground and the excited states are degenerate or near-degenerate play a special role in the radiationless relaxation dynamics\textsuperscript{4,150,151}. The mechanism of the efficient radiationless decay through conical intersections in uracil was initially investigated using the multi-reference configuration interaction (MRCI) method by Matsika\textsuperscript{152}. Later results obtained with other methods\textsuperscript{153–158} are in qualitative agreement with MRCI.

The equilibrium structure of uracil is planar and has the $C_s$ point group symmetry. The first singlet excited state corresponds to the $n\pi^*$ transition and belongs to the $A''$ irreducible representation. The second singlet excited state, $A'$, arises from the $\pi\pi^*$ transition. Only the second transition has oscillator strength by symmetry. The next two valence singlet states are also of the $n\pi^*$ and $\pi\pi^*$ types. A Rydberg $A''$ state is close in energy to the second $n\pi^*$ state.

Absorption spectra of uracil in the gas phase\textsuperscript{159} and in solution\textsuperscript{155} show a broad feature centered at 244 nm (5.08 eV) and 259 nm (4.79 eV), respectively. These values have been interpreted as the vertical $\pi\pi^*$ excitation energy. The assignment was supported by complete active space self-consistent field (CASSCF)\textsuperscript{160} calculations with the perturbation theory correction (CASPT2) and time-dependent density functional theory
(TDDFT) with the PBE0 functional\textsuperscript{155,161}, which produce gas-phase vertical excitations of 5.0 eV and 5.26 eV, respectively. However, more reliable MRCI\textsuperscript{152}, and EOM-CCSD\textsuperscript{20,162} predict higher values of the vertical excitation energy for the bright $\pi\pi^*$ state. More recent CASPT2 calculations\textsuperscript{20} employing an empirical parameter to correct for “the known tendency of CASPT2 to slightly underestimate excitation energies”\textsuperscript{20}, yield a slightly higher excitation energy, which agrees better with EOM-CCSD, but moves away from the experimental absorption maximum. Thus, the excellent agreement between the earlier CASPT2 calculations\textsuperscript{160} and the experiment is rather coincidental. The authors of two recent benchmark studies\textsuperscript{20,162} note that EOM-CCSD energies are generally higher than CASPT2 and that inclusion of triple excitations through the iterative CC3 method\textsuperscript{47} usually brings them down. For smaller molecules and a non-augmented basis, an impressive correlation between CC3 and CASPT2 results has been observed\textsuperscript{20}. CC3 calculations\textsuperscript{162} of uracil in a polarized double-zeta quality basis produce an excitation energy of the $\pi\pi^*$ state 0.2 eV lower than EOM–CCSD and do not reveal any significant doubly excited character of the state. Unfortunately, no CC3 results with larger bases have been reported.
1.4 Chapter 1 references


A.M. Tokmachev, M. Boggio-Pasqua, M.J. Bearpark, and M.A. Robb. Photo-
stability via sloped conical intersections: A computational study of the pyrene


L.H. Andersen, H. Bluhme, S. Boyé, T.J.D. Jørgensen, H. Krogh, I.B. Nielsen,
S.B. Nielsen, and A. Svendsen. Experimental studies of the photophysics of

T.M.H. Greemers, A.J. Lock, V. Subramaniam, T.M. Jovin, and S. Völker. Pho-

G. Jung, C. Bräuchle, and A. Zumbusch. Two-color fluorescence correlation
spectroscopy of one chromophore: Application to the E222Q mutant of the green

D.M. Chudakov, V.V. Belousov, A.G. Zaraisky, V.V. Novoselov, D.B. Staroverov,
D.B. Zorov, S. Lukyanov, and K.A. Lukyanov. Kindling fluorescent proteins for

Chromophore environment provides clue to “kindling fluorescent protein” riddle.


B. Hager, B. Schwarzinger, and H. Falk. Concerning the thermal diastereomer-
ization of the green fluorescent protein chromophore. *Monatshefte für Chemie

D.C. Loos, S. Habuchi, C. Flors, J. Hotta, J. Wiedenmann, G.U. Nienhaus, and

J.-S. Yang, G.-J. Huang, Yi-H. Liu, and S.-M. Peng. Photoisomerization of
the green fluorescence protein chromophore and the meta- and para-amino ana-


Chapter 2: Electronic properties of the green fluorescent protein chromophore

2.1 Chapter 2 introduction

This chapter presents an approach to accurate calculations of the properties of biological chromophores with ab initio methods using the model GFP chromophore, 4’-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) anion, as a benchmark system (Fig. 2.1). It discusses the character of the bright $\pi\pi^*$ state of the HBDI anion.

Among the benchmarked computational methods are perturbation theory corrected multiconfigurational approaches, EOM-CCSD, as well as TD-DFT with two range-separated functionals, BNL$^1$ and $\omega$PB97X$^2$. We also characterize the lowest $\pi\pi^*$ triplet and report the vertical electron detachment energy (VDE). Based on the calculated VDE, we assign the minor feature as due to the photodetachment transition. This has important implications on the character of the bright state: the $\pi\pi^*$ transition is a resonance state embedded in an ionization continuum. The triplet state, however, lies below VDE. As the resonance state, the $\pi\pi^*$ singlet has a finite lifetime and can undergo autoionization due to coupling to the ionization continuum. Contrary to that, the triplet may have much longer lifetime. Thus, population trapping in the triplet state in the gas-phase photocycle seems to be required to explain millisecond kinetics of the fragments yield
Figure 2.1: The chemical structure and atomic labels of the anionic form of the model GFP chromophore, 4’-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) in the cis- (top) and trans- (bottom) conformations.

in the ion-storage ring experiments\(^3\). Moreover, the resonance nature of the \(\pi\pi^*\) state in the anionic GFP might be responsible for very different kinetics of the photofragment yield of the anionic and protonated GFP\(^3\).

### 2.2 Computational details

The equilibrium geometries of the HBDI anion were optimized by DFT with the PBE0 variant\(^4\) of the Perdew-Burke-Ernzerhof (PBE) hybrid functional\(^5\) and by CASSCF(14/12). The cc-pVDZ basis set\(^6\) was used in both calculations. After observing noticeable differences in vertical excitation energies computed using these two geometries, we reoptimized the equilibrium structure using MP2 with cc-pVTZ\(^6\), which
yields very accurate structures for well-behaved closed-shell molecules\textsuperscript{7}. MP2 calculations employed the resolution-of-the-identity (RI) technique. The CASSCF and PBE0 structures are $C_1$, whereas the RI-MP2 optimization produced a $C_s$ minimum.

Vertical excitation energies were computed by MRMP2, TDDFT with the BNL and $\omega$PB97X functionals, CIS, SOS-CIS(D), and EOMEE-CCSD. The VDE was computed at the CASSCF geometry as the energy of the Hartree–Fock HOMO (Koopmans’ theorem), by EOMIP-CCSD, and by using the BNL HOMO energy (as described below, this is equivalent to computing VDE as the difference between the total BNL energies of the anion and the neutral radical). The $\omega$PB97X and B3LYP Koopmans’ theorem and $\Delta E$ values are also given for comparison.

The cis-trans isomerization pathways of the chromophore were studied with the DFT and CASSCF methods. Solvent effects were estimated using continuum solvation models, as well as by explicit treatment of water molecules in a QM/MM scheme.

MRMP2 and D-PCM calculations were carried out with the PC GAMESS version\textsuperscript{8} of the GAMESS(US) quantum chemistry package\textsuperscript{9}. CIS, SOS-MP2, SOS-CIS(D), EOM-CCSD, and BNL calculations were performed with Q-Chem\textsuperscript{10}. GAMESS(US)\textsuperscript{9} was employed for C-PCM and SVPE computations. The QM/MM implementation is based on PC GAMESS\textsuperscript{8} and the Tinker molecular mechanics package\textsuperscript{11}.

### 2.2.1 Continuum solvation models

To simulate solvent effects on the chromophore’s cis-trans isomerization energy profile in aqueous solution, we employ three versions of the continuum solvation model\textsuperscript{12}: D-PCM, C-PCM, and SVPE. In the simplest approach, the dielectric polarizable continuum model (D-PCM)\textsuperscript{13}, the water solvent is treated as a continuous unstructured dielectric with a dielectric constant of 78.39. C-PCM\textsuperscript{14} is a version of PCM that takes
into account certain corrections in the boundary conditions of the electrostatic problem in accord with the conductor-like screening model. The surface and volume polarization for electrostatic (SVPE) approach employs the dielectric continuum model to incorporate the effects of volume polarization that arise from the solute charge density penetrating outside the solute cavity. This effect can be significant for reaction barriers, especially for ionic solutes.

2.2.2 Effective fragment potential-based QM/MM method

Solvent effects can be described explicitly in a combined quantum mechanical-molecular mechanical (QM/MM) approach based on the effective fragment potential (EFP) model. In this scheme, solvent molecules (water in our case) are represented by effective fragments in the MM-part. The fragments affect the Hamiltonian of the QM-part (HBDI anion) by their electrostatic potentials expanded up to octupole terms. The parameters of these one-electron electrostatic potentials, as well as contributions from interactions between polarizable effective fragments and the QM-region, are computed in preliminary ab initio calculations of the electronic densities of individual fragments. The exchange-repulsion potentials, which are combined with the electrostatic and polarizability terms, are obtained from preliminary ab initio calculations as well.

The original EFP approach treats interactions between solvent molecules as EFP-EFP interactions. Studies of chemical reactions in aqueous solution have shown that replacing the EFP-EFP terms by the empirically calibrated TIP3P potential yields a faster computational scheme for large water clusters around the solute.
2.3 Anionic form of HBDI

2.3.1 Molecular orbital framework

Fig. 2.2 shows two resonance structures of the HBDI anion. The interaction between these two structures result in charge delocalization between the two oxygen atoms and in scrambling CC bond orders, as discussed, for example in Ref. 40. Changes in bond-orders in the bridge region due to the resonance are believed to be important in thermal isomerization.\(^\text{20}\)

![Figure 2.2: Two resonance structures of the HBDI anion.](image)

The Natural Bond Orbital (NBO)\(^\text{21,22}\) charges on the phenoxy or the imidazolin oxygens (computed using the RI-MP2 densities) are \(-0.65\) and \(0.66\), respectively. The almost equal values of the charges suggest significant contributions from both resonance structures. As the result of the resonance interaction, the \(C_1P-C_1B\) bond gains double-bond character, whereas the order of \(C_1B-C_{1I}\) bond is reduced, and the bridge moiety acquires allylic character. It is interesting to compare the respective bond lengths with the values for the double and single CC bonds between \(sp^2\) hybridized carbons, e.g., in a planar and twisted ethylene (see Ref. 24 for explanation regarding the choice of the reference structures). Our best estimates of the \(C_1P-C_1B\) and \(C_1B-C_{1I}\) bond lengths (RI-MP2/cc-pVTZ) are \(1.394\) Å and \(1.378\) Å, respectively. The lengths of the CC bond in ethylene is \(1.333\) Å at the planar geometry (where the formal bond order is 2), and
1.470 Å at the twisted configuration, where the π-bond is completely broken. Thus, assuming a linear relationship between the bond order and bond length, one can assign 55% and 67% of a double-bond character to the $C_{1P}-C_{1B}$ and $C_{1B}-C_{1I}$ bonds, respectively. The NBO analysis assigns 1.8 electrons to a slightly asymmetric allylic three-center bond. The relative contributions of the $C_{1P}$ and $C_{1I}$ carbons are 38 and 62%, respectively, in a semi-quantitative agreement with the bond orders derived from the bond lengths.

Figure 2.3: Two molecular orbitals of the HBDI anion giving rise to the $\pi\pi^*$ state. The character of the orbitals can be explained by considering linear combination of two localized $\pi$-bonding orbitals.
The resonance interaction is also reflected by the shape of MOs. Fig. 2.3 shows two Hartree–Fock orbitals involved in the bright $\pi\pi^*$ and in the photodetachment transitions (HOMO and valence LUMO). These orbitals, which are traditionally referred to as $\pi$ and $\pi^*$, have quite complicated shapes and are delocalized over the entire molecule. Their character in the bridge region can be explained by considering two interacting $\pi$ orbitals, as shown in Fig. 2.3. The HOMO can be described as an out-of-phase combination of two localized $\pi$ bonds, whereas the large electron density on $C_{1B}$ in the LUMO can be derived from the in-phase combination. Of course, due to the delocalized character of the orbitals, this picture is just an approximation, but it allows one to see the origin of the large oscillator strengths and changes in charge distribution in the excited state, and also provides a useful framework for explaining the character of the transition state along $cis-trans$ isomerization coordinate.

### 2.3.2 Vertical electron detachment energy of the HBDI anion

Since the HBDI anion is a closed-shell system, it is stable in the gas phase and has a relatively large vertical electron detachment energy (VDE). However, as shown below, it does not support bound electronically excited singlet states, and the lowest valence excitation is embedded in an ionization continuum. Such resonance states are very common in molecular anions $^{25}$, and play an important role in dissociative electron attachment processes $^{26, 27}$. Thus, the broad character of the experimental action spectrum $^{28, 29}$ is at least partially due to the broadening of the resonance-like $\pi\pi^*$ state by its interaction with the ionization continuum. Finite lifetime and autoionization decay of this state should be taken into account when considering photoinduced dynamics and lifetime of the HBDI anion in the gas phase. In the condensed phase, solvent may stabilize the anion such that its excited states become bound. However, one-photon photodetachment
channel may still be relevant for the anionic forms, especially when the production of solvated electrons is considered\textsuperscript{30}. The resonance character of the \( \pi \pi^* \) state also has significant consequences in the electronic structure calculations\textsuperscript{25,31} of excited states, as described below.

The simplest estimate of VDE obtained by applying Koopmans’ theorem is 2.56–2.93 eV depending on the basis set, the largest value obtained with the 6-311(2+,+)G(2df,2pd) basis. The EOMIP method, which includes electronic correlation, can provide more reliable estimates of VDE. However, due to the size of the system, we are limited to relatively modest basis sets. In the 6-31G* basis, Koopmans’ VDE is 2.56 eV, while EOMIP-CCSD yields 2.05 eV. The larger 6-31+G* basis increases the energies to 2.91 and 2.48 eV, respectively. Thus, including correlation reduces VDE by 0.4–0.5 eV, and assuming the effect is consistent throughout basis sets, we estimate that the target EOMIP-CCSD/6-311(2+,+)G(2df,2pd) energy is 2.4–2.5 eV.

B3LYP/cc-pVDZ energy difference calculations yield 2.46 eV, which is considerably larger than the respective Koopmans value of 0.92 eV. The \( \omega \text{PB97X/cc-pVDZ} \) energy difference value is 2.39 eV, whereas the respective Koopmans IE is much higher (2.83 eV).

By construction, the BNL energy difference VDE is equal to the respective HOMO energy. Recent benchmarks\textsuperscript{32} demonstrated that BNL produces very accurate ionization energies. VDE calculated with BNL in the small cc-pVDZ basis is 1.99 eV (\( \gamma = 0.250 \)), and it increases to 2.52 eV in the 6-311G(2+,+)G(d,p) basis, for which \( \gamma = 0.228 \). Additional sets of diffuse orbitals or more extensive polarization do not affect this value, e.g. VDE calculated with BNL/6-311G(3+,2+)G(2df,2pd) is 2.53 eV.

The discrepancies between the Koopmans and \( \Delta E \) values have important implications for the excited state calculations, as the former value defines the onset of the
Thus, with B3LYP, the continuum begins 1.54 eV below its own VDE, whereas the situation with ωPB97X is reverse — the continuum states appear 0.44 eV above the respective VDE. BNL, by constriction, is internally consistent, and the continuum states in TDDFT calculations appear exactly at the respective VDE.

Thus, our estimate of VDE is 2.4–2.5 eV, within 0.1 eV from the maximum of the weak absorption feature at 2.3 eV. The remaining discrepancy between the two values might be due to the uncertainties in equilibrium geometries or possible vibrational excitation of the molecules in the experiment.

Although correlation has significant effect on VDE, the ionized state has Koopmans-like character, i.e. the leading EOM-IP amplitude corresponds to ionization from the HOMO (see Fig. 2.3) and equals 0.96. Thus the Hartree–Fock HOMO is a good approximation to a correlated Dyson orbital.

### 2.3.3 Vertical excitation energy and electronic properties of the singlet and triplet ππ* transitions of the HBDI anion

**Singlet ππ* state: Benchmark results**

Motivated by the discrepancies in previous theoretical estimates of the ππ* excitation energy of the HBDI anion (Table 1.1), we set out to benchmark other electronic structure methods, with the purpose of identifying a rigorous yet fairly inexpensive quantum chemistry approach that can be employed in condensed-phase applications. The experimental maximum of absorption is at 2.59 eV (479 nm), and the band’s full width at half maximum (FWHM) is 0.25 eV (45 nm). Assuming that the absorption maximum corresponds to the vertical transition of the lowest-energy isomer (which is not entirely
clear, as the temperature of the ions in the ring is unknown), one would like the computed wavelength to fall within 2.47–2.72 eV (456–502 nm). However, due to the resonance nature of the $\pi\pi^*$ state, calculating the excitation energies and oscillator strengths, as well as comparing them with the experimental spectrum, are not as straightforward as in the case of excited states lying below the electron detachment energy. In the following, we will use the term “detached states” to identify the electronic states that compose the continuum instead of the usual “ionized states” as the initial species is anionic.

The ground-state equilibrium geometry was optimized with PBE0/cc-pVDZ, CASSCF/cc-pVDZ, and RI-MP2/cc-pVTZ. Although the differences between the geometries are small (maximum bond length deviation is 0.03 Å and angles agree within 2 degrees), the $\pi\pi^*$ excitation energy computed using wavefunction-based methods differs by about 0.1 eV. This effect of geometry is consistent with previous calculations by Olsen\textsuperscript{34} for a similar system (p-hydroxybenzylidene-imidazolin-5-one, HBI). Of the three structures, the RI-MP2/cc-pVTZ one is the most accurate\textsuperscript{7}. Since most of the changes in electron density occur in the bridge region, it is interesting to compare the $C_{1p}$-$C_{1I}$ and $C_{1B}$-$C_{1I}$ bond lengths computed by different methods. The RI-MP2 values are 1.394 Å and 1.378 Å, which is very close to the PBE0 values of 1.404 Å and 1.385 Å. Due to the absence of dynamical correlation, CASSCF exaggerates the bond alternation giving 1.406 Å and 1.397 Å.

Calculations in small basis sets create discrete states from the continuum and artificially exclude the detached states from the picture. Assuming that the character of the resonance state of interest is well described, such calculations may provide a fairly good estimate of the position of this state in the continuum, however, it is difficult to predict how expanding the basis set will affect the resonance state\textsuperscript{25,31}. Moreover, one should anticipate broadening of the resonance state due to the interaction with the detached
states. Increasing the size of the one-electron basis brings the continuum states down. When the basis set is large enough to accommodate a detached electron, the lowest excited state will correspond to the detached state (this fact has been exploited in pilot implementations of EOMIP-CCSD based on the EOMEE-CCSD code by adding a very diffuse orbital to the basis to describe the ionized electron\textsuperscript{35,36}). It can be shown formally, that in the case of CIS and TD-DFT, the energy of the lowest of such CIS-IP states equals the HOMO energy, that is, the Koopmans theorem can be proven by considering configuration interaction of all singly detached determinants.

When the one-electron basis set is expanded with diffuse functions, the density of states rapidly increases with both CIS and TD-DFT (BNL), as illustrated in Fig. 2.4. At the same time, the lowest excited state converges to Koopmans’ VDE. Both methods split the oscillator strength of the bright $\pi\pi^*$ transition among multiple states, but BNL does that to a larger degree, which results in three to four states with close oscillator strengths (Table 2.1). That may be due to the remaining self-interaction error in BNL calculations.

The lowest excited CIS and TD-DFT states shown in Fig. 2.4 fall below the Koopmans estimate in large basis sets due to additional stabilization by the interaction with other CIS determinants that vanish once the detached electron is infinitely far from the core. Indeed, if a separate large non-interacting orbital is used instead of diffuse functions, the lowest excitation energy in such a system is exactly equal to the HOMO energy.

As the density of low-lying states that approximate the continuum increases and the oscillator strength gets redistributed, it becomes increasingly more difficult to compute or even identify the resonance state. By following the evolution of the states with the basis set increase, the limiting value of the resonance state can be extrapolated using the
Figure 2.4: Effect of increasing the number of diffuse functions in the basis set on the density of states and the convergence of the lowest excited and the bright $\pi\pi^*$ state. The calculations were performed with CIS (top) and TD-DFT/BNL (bottom). The basis set was varied from 6-311G(2pd,2df) to 6-311(4+,2+)G(2pd,2df).
Table 2.1: The interaction of the bright $\pi\pi^*$ state with the electron-detached continuum as reflected by the diminishing oscillator strength ($f_L$) in CIS and TD-DFT/BNL calculations. Also see Fig. 2.4.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Number</th>
<th>Energy, eV</th>
<th>$f_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CIS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(2df,2pd)</td>
<td>1</td>
<td>3.92</td>
<td>1.58</td>
</tr>
<tr>
<td>6-311(+,+G)G(2df,2pd)</td>
<td>2</td>
<td>3.81</td>
<td>1.50</td>
</tr>
<tr>
<td>6-311(2+,+G)G(2df,2pd)</td>
<td>6</td>
<td>3.66</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3.74</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.85</td>
<td>1.14</td>
</tr>
<tr>
<td>6-311(2+,2+)G(2df,2pd)</td>
<td>9</td>
<td>3.60</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>3.82</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>4.02</td>
<td>0.06</td>
</tr>
<tr>
<td>6-311(3+,2+)G(2df,2pd)</td>
<td>37</td>
<td>3.73</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>3.84</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>TD-DFT (BNL)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(2df,2pd)</td>
<td>2</td>
<td>3.44</td>
<td>1.47</td>
</tr>
<tr>
<td>6-311(+,+G)G(2df,2pd)</td>
<td>3</td>
<td>3.27</td>
<td>1.38</td>
</tr>
<tr>
<td>6-311(2+,+G)G(2df,2pd)</td>
<td>7</td>
<td>3.18</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3.24</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.38</td>
<td>0.52</td>
</tr>
<tr>
<td>6-311(2+,2+)G(2df,2pd)</td>
<td>8</td>
<td>3.12</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3.15</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>3.29</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>3.29</td>
<td>0.63</td>
</tr>
<tr>
<td>6-311(3+,2+)G(2df,2pd)</td>
<td>35</td>
<td>3.23</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>3.24</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>3.37</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>3.37</td>
<td>0.15</td>
</tr>
</tbody>
</table>
stabilization method by Taylor\textsuperscript{37}, as has been done in the calculations of resonance electron attached states\textsuperscript{25,31}. From the data in Table 2.1, we can estimate that the excitation energy converges to 3.8 eV with CIS and to 3.3 eV with TD-DFT/BNL. Overall, adding sets of diffuse functions results in lowering the excitation energy of the bright state by 0.1 eV.

The CIS and TD-DFT/BNL calculations demonstrate that the resonance state is embedded in an electron detached continuum, and the broad character of the experimental spectrum can be due to the interaction with the continuum. Thus, small basis set calculations of the vertical excitations energies of the $\pi\pi^*$ state can only provide a rough estimate of the position of the absorption maximum.

Introducing a non-iterative doubles correction via SOS-CIS(D) significantly lowers the energy of the $\pi\pi^*$ transition, while the detached state energies are affected less. We could not obtain SOS-CIS(D) results in heavily augmented bases sets due to the lack of the corresponding auxiliary bases for the RI procedure, which becomes unstable upon adding second set of the diffuse functions. The magnitude of correction is 0.9–1.2 eV for the $\pi\pi^*$ state and 0.1–0.3 eV for the detached states. MRMP2 based on sa-CASSCF(14/12) in the modest cc-pVDZ basis yields 2.51–2.61 eV for the $\pi\pi^*$ excitation energy. SOS-CIS(D)/cc-pVDZ gives 2.71–2.81 eV, which is consistent with the more rigorous MR-MP2 estimates.

At the CASSCF-optimized geometry, EOM-CCSD/6-31G* yields an excitation energy of 3.16 eV and an oscillator strength $f_L = 1.27$. Including only one set of diffuse functions lowers the energy to 3.04 eV ($f_L = 1.25$). The change in the oscillator strength is consistent with the observed drop of the EOM amplitude corresponding to the $\pi\pi^*$ transition from 0.85 to 0.74. Benchmarks on the electronically excited states of
closed-shell molecules have shown that the combined effect of increasing the basis set in EOM-CCSD calculations and the triples correction can be as large as 0.3–0.4 eV.

The computational cost of MRMP2 and EOM-CCSD does not allow us to use these methods to fully explore the basis set effect on the $\pi\pi^*$ excitation energy, e.g., via stabilization method. We anticipate a noticeable effect of improving the basis set on the energy of the $\pi\pi^*$ state. For example, with SOS-CIS(D), the excitation energy with the cc-pVTZ basis is lower by 0.13 eV relative to cc-pVDZ. The EOM-CCSD excitation energy also drops by as much as 0.2 eV upon expanding the basis set from 6-31G* to 6-31+G*. The oscillator strength calculated with CIS is about 10% lower when basis sets with diffuse functions are used compared to those without, demonstrating the beginning of the interactions with the continuum.

The $\pi\pi^*$ excitation energy obtained with MRMP2 based on state-averaged CASSCF(16/14), which spans the entire $\pi$-electron active space (Table 1), and a more compact (14/12) active space (Table 2.2) agree equally well with the experimental result: 2.47 eV (501 nm) and 2.52–2.61 eV (476–491 nm), respectively, versus 2.59 eV (479 nm). The most expensive aug-MCQDPT2/CASSCF(16/14) approach (Table 1) does not perform noticeably different than MRMP2/CASSCF(14/12). Therefore, one can rely on the fairly practical MRMP2 approach based on a reduced active space in the CASSCF wavefunction. Overall, the effect of contracting the active space is less than 0.1 eV, which is smaller than the uncertainty due to the equilibrium geometry and anticipated effects of extending the basis set beyond the double-zeta level.

In view of complexities associated with performing multireference perturbation theory and underlying CASSCF calculations, we find the results of the inexpensive
Table 2.2: Vertical $\pi\pi^*$ electronic excitation energies ($\Delta E$, in eV), corresponding wavelengths (nm) and oscillator strength ($f_L$) for the gas phase GFP chromophore (Fig. 1). The reference experimental value is 2.59 eV (479 nm).

<table>
<thead>
<tr>
<th>Method</th>
<th>Ground-state geometry optimized with</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PBE0/cc-pVDZ</td>
<td>CASSCF(14/12)/cc-pVDZ</td>
<td>RI-MP2/cc-pVTZ</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\Delta E$ $f_L$ nm</td>
<td>$\Delta E$ $f_L$ nm</td>
<td>$\Delta E$ $f_L$ nm</td>
<td></td>
</tr>
<tr>
<td>MRMP2 based on sa-CASSCF(14/12)/cc-pVDZ</td>
<td>2.52$^b$ 491$^b$</td>
<td>2.61$^c$ 476$^c$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD/6-31G(d)</td>
<td>3.08 1.26 402</td>
<td>3.16 1.27 392</td>
<td>3.12 398</td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD/6-31+G(d)</td>
<td>2.97 1.24 418</td>
<td>3.04 1.25 408</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOS-CIS(D)/cc-pVDZ</td>
<td>2.71 1.59$^a$ 457</td>
<td>2.81 1.61$^a$ 441</td>
<td>2.75 1.60$^a$ 451</td>
<td></td>
</tr>
<tr>
<td>SOS-CIS(D)/aug-cc-pVDZ</td>
<td>2.57 1.45$^a$ 482</td>
<td>2.67 1.45$^a$ 464</td>
<td>2.61 1.45$^a$ 475</td>
<td></td>
</tr>
<tr>
<td>SOS-CIS(D)/cc-pVTZ</td>
<td>2.58 1.54$^a$ 480</td>
<td>2.68 1.56$^a$ 463</td>
<td>2.62 1.54$^a$ 473</td>
<td></td>
</tr>
<tr>
<td>SOS-CIS(D)/aug-cc-pVTZ</td>
<td>2.58 1.37$^a$ 480</td>
<td>2.90 1.04$^a$ 427</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-DFT/BNL/cc-pVDZ</td>
<td>3.44 1.51 360</td>
<td>3.50 1.51 354</td>
<td>3.59 1.55 346</td>
<td></td>
</tr>
<tr>
<td>TD-DFT/BNL/6-311(2df,2pd)</td>
<td>3.22 1.38 385</td>
<td>3.27 1.38 379</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD-DFT/ωWB97x/cc-pVDZ</td>
<td>3.52 1.52 352</td>
<td>3.59 1.53 346</td>
<td>3.55 1.53 349</td>
<td></td>
</tr>
<tr>
<td>TD-DFT/ωWB97x/6-311(2df,2pd)</td>
<td>3.38 1.45 367</td>
<td>3.44 1.46 360</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Oscillator strength calculated with CIS. $^b$ At the equilibrium geometry computed with PBE0/(aug)-cc-pVDZ (diffuse functions only on oxygen atoms). $^c$ At the equilibrium geometry computed with CASSCF(12/11)/(aug)-cc-pVDZ (diffuse functions only on oxygen atoms).
SOS-CIS(D) method very encouraging: at the PBE0-optimized geometry, the SOS-CIS(D)/cc-pVTZ $\pi\pi^*$ excitation energy of 2.58 eV is within 0.01 eV from the experimental absorption maximum, and agrees very well with MRMP2 calculations.

The EOM-CCSD/6-31+G* excitation energy is 0.38 eV above the experimental maximum, which is outside the EOM-CCSD error bars. Analysis of the EOM amplitudes confirms singly excited character of the $\pi\pi^*$ state, thus, with an adequate basis set, EOM-CCSD error bars should not exceed 0.3 eV. In addition to anticipated basis set effects and interactions with the continuum states, other factors such as discrepancies due to equilibrium geometry, as well as the uncertainty in the experimental value (FWHM of the experimental absorption is 0.25 eV), make it difficult to arrive at a definite conclusion. More extensive calculations using larger basis sets (and ideally using stabilization graphs) and estimates of triples corrections (to account for dynamical correlation) are required to assess the performance of EOM-CCSD for this molecule. Basis set and triples effects alone have been shown to account for as much as 0.3 eV in excitation energies of $\pi\pi^*$ character.

Finally, we present TD-DFT results computed with the range-separated functionals. Using the BNL functional\textsuperscript{1} with the small cc-pVDZ basis set, the $\pi\pi^*$ state appears the second lowest, which is consistent with the respective VDE. At the CASSCF geometry, its excitation energy is 3.50 eV, and the oscillator strength is $f_L = 1.51$. The 6-311(+(+))G(2df,2pd) basis lowers the excitation energy to 3.27 eV, and the oscillator strength to 1.38. $\omega$PB97X\textsuperscript{2} gives similar values for the excitation energy and the oscillator strengths, i.e., 3.59 eV ($f_L = 1.51$) and 3.44 ($f_L = 1.46$) with the cc-pVDZ and 6-311(+(+))G(2df,2pd), respectively. Because the Koopmans continuum with $\omega$PB97X begins at higher energies (see section 2.3.2), the detached states do not appear below the excited states in these calculations.
Overall, it is unrealistic to expect accuracy better than 0.1 eV (20 nm) from computational protocols applicable to a molecule of this size even for non-resonance states, and the observed discrepancies between different methods confirm that. Moreover, when assessing the accuracy of computed values, one should keep in mind the finite width of the experimental absorption band. Thus, more calculations are necessary to provide a converged theoretical estimate, especially stabilization analysis. For practical applications, however, it is important that all the reliable theoretical methods agree with each other in that the origin of the intensity in the resonance state is due to $\pi\pi^*$ excitation. SOS-CIS(D) offers an inexpensive alternative to more rigorous multireference methods if single excitations are dominant in the wavefunction.

**Changes in electronic density in the singlet $\pi\pi^*$ state**

As one may expect from the molecular orbital character (Fig. 2.3) and the large oscillator strength, electronic excitation results in a significant redistribution of electronic density. A convenient measure of charge distribution is the permanent dipole moment. Although in charged system it is origin-dependent, the difference between the two dipole moments, $\Delta \mu = \mu_{gr} - \mu_{ex}$, is not. At the CIS level of theory, the value of $|\Delta \mu|$ is 0.6 D, and its direction is in the molecular plain pointing towards the bridge carbon. This value can be compared with the experimentally measured $\Delta \mu$, derived from Stark effect measurements in a buffered at pH=6.5 glycerol solution at 77 K\textsuperscript{38}. This work also reports the angle between $\Delta \mu$ and $\Delta \mu_{tr}$. Strikingly, the experimental value is 10 times larger than the computed one. Since $\Delta \mu$ is related to the changes in orbital occupations upon excitation, it is dominated by contributions from the leading excitation amplitudes, and should be reproduced fairly accurately at the CIS level. Thus, large discrepancy is likely to be due to the solvent effect. Indeed, polar solvents result in the increased dipole moment of the
Table 2.3: Vertical triplet $\pi\pi^*$ electronic excitation energies ($\Delta E$, in eV) of the HBDI anion in the gas phase.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>CIS</th>
<th>SOS-CIS(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>2.03</td>
<td>1.91</td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>2.02</td>
<td>1.88</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>2.03</td>
<td>1.86</td>
</tr>
</tbody>
</table>

solute. For example, dipole moment of water in bulk water is about 30% larger than in the gas phase. More polar charge distribution in the ground state in solvent is clearly seen from the respective NBO charges (see Table 2.5). Thus, for the difference of dipole moments of two states one may anticipate an enhanced effect.

**Triplet $\pi\pi^*$ state**

The vertical excitation energies at the RI-MP2/cc-pVTZ geometry of the lowest triplet state are summarized in Table 2.3. The analysis of the wave function confirms that the triplet is derived from the transitions between the same orbitals as the singlet (Fig. 2.3). As expected, all methods consistently place the triplet considerably below the singlet. The variations between the methods are smaller for the triplet state. Our best value (SOS-CIS(D)/cc-pVTZ) is 1.86 eV. The 0.76 eV gap between the singlet and the triplet does not suggest efficient intersystem crossing at this geometry. The triplet state is 0.3–0.4 eV below VDE and is, therefore, a bound electronic state. Thus, much longer lifetime is expected for this state (as compared to the singlet), not only because the radiationless relaxation to the ground state is a spin-forbidden process, but also because the autoionization channel is absent.
2.3.4 Cis-trans isomerization pathway of anionic HBDI

Gas phase

On the ground-state PES, conversion between the lower-energy cis-isomer and the higher-energy trans-isomer takes place when the two rings are rotated along the C1\textsubscript{I}-C1\textsubscript{B} bond as shown in Fig. 2.1. The dihedral angle N-C\textsubscript{1I}-C\textsubscript{1B}-H, which defines the reaction coordinate, is denoted as $\tau_{I}$.

The shape of the isomerization pathway can be characterized by: (i) the activation energy $E_{a}$ defined as the relative energy of the transition state (TS) with respect to the cis-isomer; and (ii) the energy difference $E_{tc}$ between the trans- and cis-forms. We began with gas-phase DFT calculations with the PBE0 functional and the 6-31+G(d,p) basis set and located stationary points that correspond to the cis-, trans- and TS structures. The latter is characterized by a single imaginary frequency of 725$i$ cm\textsuperscript{-1}. The intrinsic reaction coordinate (IRC) profile was computed by starting steepest descent pathways from the TS point in both directions along the Hessian eigenvector that corresponds to the imaginary frequency.

The results of this calculation are discouraging in two aspects: the cusp-like shape of the profile (Fig. 2.5), which is consistent with the large value of the imaginary frequency at the saddle point, and the value of the activation energy $E_{a}(\text{DFT}) = 34.5$ kcal/mol, which is more than twice higher than the experimental estimate of 15.4 kcal/mol in aqueous solution\textsuperscript{39}. The computed energy difference between the trans- and cis- structures $E_{tc}(\text{DFT}) = 2.3$ kcal/mol is in excellent agreement with the experimental estimate in solution\textsuperscript{39}, most likely due to error cancellation.
In the CASSCF energy profile (Fig. 2.5), the stationary points were fully optimized with CASSCF(12/11)/cc-pVDZ. The points in between representing the minimum energy path (MEP) were computed varying the value of the dihedral angle $\tau_I$ and minimizing the energy by relaxing all other degrees of freedom. Although this curve does not represent the true IRC path, it is expected to be a good approximation to it. Along with a smoother curvature in the vicinity of the saddle point (imaginary frequency $77i$ cm$^{-1}$), this profile yields a reasonable value of the trans-cis energy difference $E_{tc}(\text{CASSCF}) = 3.5$ kcal/mol and a much lower (and closer to the experimental estimate) value of the activation energy $E_a(\text{CASSCF}) = 22.5$ kcal/mol.

When calculating the energy profile in the CASSCF(12/11) approximation with a fairly large active space, we performed careful selection of the orbitals in order to avoid
dubious solutions of the variational problem. To better understand changes in the electronic structure along the isomerization pathway, we also computed the energy profile with the smallest active space CASSCF(2/2). Selection of active orbitals in this approach was performed on the base of previously optimized orbitals at the TS point. Then the descent in both directions towards minimum energy structures was easy to accomplish. As expected, the corresponding value of the activation energy, 25 kcal/mol, was slightly larger than that computed with CASSCF(12/11).

As discussed in more detail below, there is an increase in charge localization at the TS relative to that the minimum energy points.

The transition state located with SOS-MP2/cc-pVDZ is characterized by an imaginary frequency of $188i \text{ cm}^{-1}$. Starting from that point, the MEP was taken to the cis- and trans-configurations (Fig. 2.5). Both DFT and SOS-MP2 exhibit a cusp at the transition state, which is a manifestation of the multireference character of the ground-state wavefunction. The composition of the CASSCF wavefunction in the region discussed below confirms that.

Table 2.4 presents computed equilibrium geometry parameters in the bridging region: the C$_{1P}$-C$_{1B}$ and C$_{1I}$-C$_{1B}$ bond lengths and the $\tau_P$ (C$_{2P}$-C$_{1P}$-C$_{1B}$-H) and $\tau_I$ (N-C$_{1I}$-C$_{1B}$-H) dihedral angles (Fig. 2.1). The parameters optimized at the PBE0/6-31+G(d,p), CASSCF(12/11)/cc-pVDZ, and SOS-MP2/cc-pVDZ levels are in agreement with those obtained by Olsen and Smith$^{40}$ with SA3-CAS(4/3)/DZP, which stands for CASSCF with the (4/3) active space averaged over three states computed with the DZP basis set. Overall, the geometry parameters at the stationary points are rather insensitive to the level of theory: DFT, large active space state-specific CASSCF, small active space state-averaged CASSCF, and SOS-MP2 yield bond lengths that agree within 0.02 Å, and angles that agree within 2°.
Table 2.4: Geometry parameters (distances in Å, angles in degrees) at the stationary points for the gas-phase cis-trans chromophore isomerization.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Method</th>
<th>C$<em>{1P}$–C$</em>{1B}$</th>
<th>C$<em>{1B}$–C$</em>{1I}$</th>
<th>$\tau_P$</th>
<th>$\tau_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cis-isomer</td>
<td>PBE0/6-31+G(d,p)</td>
<td>1.404</td>
<td>1.385</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>CASSCF(12/11)/cc-pVDZ</td>
<td>1.406</td>
<td>1.383</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>SOS-MP2/cc-pVDZ</td>
<td>1.415</td>
<td>1.397</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>SA3-CAS(4,3)/DZP$^a$</td>
<td>1.408</td>
<td>1.382</td>
<td>0</td>
<td>180</td>
</tr>
<tr>
<td>Transition state</td>
<td>PBE0/6-31+G(d,p)</td>
<td>1.365</td>
<td>1.458</td>
<td>0.3</td>
<td>88.2</td>
</tr>
<tr>
<td></td>
<td>CASSCF(12/11)/cc-pVDZ</td>
<td>1.362</td>
<td>1.477</td>
<td>1.0</td>
<td>86.4</td>
</tr>
<tr>
<td></td>
<td>SOS-MP2/cc-pVDZ</td>
<td>1.374</td>
<td>1.476</td>
<td>0.0</td>
<td>87.6</td>
</tr>
<tr>
<td>Trans-isomer</td>
<td>PBE0/6-31+G(d,p)</td>
<td>1.403</td>
<td>1.392</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CASSCF(12/11)/cc-pVDZ</td>
<td>1.402</td>
<td>1.395</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SOS-MP2/cc-pVDZ</td>
<td>1.414</td>
<td>1.405</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SA3-CAS(4,3)/DZP$^a$</td>
<td>1.407</td>
<td>1.390</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$^a$ Ref. 40.
At both cis- and trans-equilibrium points, the molecule is essentially planar except that the methyl groups naturally have out-of-plane atoms. Analysis of the structures reveals that the difference between $C_{1P}-C_{1B}$ and $C_{1B}-C_{1I}$ is less than expected for the structure shown in Fig. 2.1. This can be explained by considering two resonance structures of the HBDI anion, see Fig. 2.2. The interaction between these two structures results in charge delocalization between two oxygen atoms and in scrambling CC bond orders as discussed, for example, in Ref. 40. The $C_{1P}-C_{1B}$ bond acquires a double-bond character, whereas the order of $C_{1B}-C_{1I}$ bond is reduced. Natural Bond Orbital (NBO)$^{21,22}$ charges and bond orders are consistent with computed bond lengths. Overall, CASSCF slightly exaggerates bond alternation relative to DFT or MP2, in favor of the canonical structure (Fig. 2.1).

At the TS, the two rings are nearly perpendicular to each other ($\tau = 87 – 88^\circ$), which disturbs the $\pi$ system and breaks the resonance interaction. The bond alternation pattern is reversed such that $C_{1P}-C_{1B}$ becomes shorter than $C_{1B}-C_{1I}$ suggesting that one of the two resonance structures becomes dominant. All the methods agree on the magnitude of the change relative to the equilibrium structures: $C_{1P}-C_{1B}$, which is longer at the cis- and trans-geometries, becomes shorter by about 0.04 Å at the TS, whereas the $C_{1B}-C_{1I}$ bond is about 0.07 Å longer at the saddle point. NBO charges (Table 2.5) and the molecular orbital picture (Fig. 2.6) reveal almost complete charge localization on the imidazolin ring.

Despite the similarity of the computed geometry parameters and charge distributions at the minima and TS computed with DFT and CASSCF, respective activation energies differ by more than 10 kcal/mol (34.5 versus 22.5 kcal/mol). Note that the corresponding cis-trans energy differences are very close: 2.3 and 3.5 kcal/mol, respectively. A close inspection of the charges from Table 2.5 reveals slightly more polar charge distribution.
Figure 2.6: Frontier valence molecular orbitals of the HBDI anion in the cis-form (left) and at the transition state of the cis-trans isomerization path (right).
Table 2.5: Cumulative natural charges on the fragments of the HBDI molecule: the phenyl and dimethylimidazolin rings and the CH bridge calculated with CASSCF(12/11) and DFT (PBE0 functional).

<table>
<thead>
<tr>
<th></th>
<th>CASSCF(12/11)/cc-pVDZ</th>
<th>PBE0/6-31+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cis</td>
<td>TS</td>
</tr>
<tr>
<td>Gas phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenyl</td>
<td>−0.61</td>
<td>−0.20</td>
</tr>
<tr>
<td>Bridge</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Imidazolin</td>
<td>−0.52</td>
<td>−0.93</td>
</tr>
<tr>
<td>Solutiona</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenyl</td>
<td>−0.85</td>
<td>−0.09</td>
</tr>
<tr>
<td>Bridge</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>Imidazolin</td>
<td>−0.29</td>
<td>−0.99</td>
</tr>
</tbody>
</table>

*a Calculated with QM/MM: EFP for the solvent-QM part interaction and TIP3P for the water-water interaction.

for the PBE0 density: the positive charge on the bridge moiety is 0.19, versus 0.13 at the CASSCF level. Because of the large energy penalty due to charge separation in the gas phase, small differences in ionicity may produce a large effect. Another notable difference is larger asymmetry in oxygen charges: at the CASSCF level the imidazolin oxygen is by 0.22e more negative than the phenyl oxygen, whereas at the DFT level this difference is reduced to 0.15. This suggests larger contribution of the second resonance structure in the CASSCF wave function, which can also contribute to the energy difference.

More ionic character of the PBE0 density and the cusp-like shape of the profile (Fig. 2.5) are due to multiconfigurational character of the wave function at the TS, which is not adequately described by DFT (or MP2). The CASSCF amplitudes
show almost equal weights of the two dominant configurations, \((\text{HOMO})^2\) and \((\text{HOMO})^1(\text{LUMO})^1\) at TS. The HOMO and LUMO at the TS differ considerably from those at the equilibrium geometry (shown in Fig. 2.3) — they become localized on the imidazolin and phenyl rings, respectively. Thus, \((\text{HOMO})^2\) and \((\text{LUMO})^2\) correspond to the two charge-localized configurations, and their interaction results in less ionic electron distribution. The ionicity is also reduced by \((\text{HOMO})^1(\text{LUMO})^1\).

A stability analysis of the Hartree-Fock wavefunction at the TS shows an RHF-UHF instability with a negative eigenvalue of \(-0.037\). The DFT/PBE0 solution, however, proves to be stable, which means that using the symmetry broken unrestricted solution to achieve better description of the barrier will not be useful in this case. Therefore, a multiconfigurational approach is necessary not only for describing excited-state isomerization of the GFP-like chromophores, but also for modeling isomerization in the ground state.

Although the CASSCF wave function is capable of capturing the multiconfigurational character of the wave function, it needs to be augmented by dynamical correlation to provide accurate energy differences. We included dynamical correlation correction via multireference perturbation theory (MRMP2) for the \(E_{tc}\) and \(E_a\) energies (at the CASSCF geometries). This yields an activation energy of \(E_a = 26.2\) kcal/mol, which is 3.7 kcal/mol higher than the CASSCF result.

The CASSCF results represent an improvement over DFT, however, there is still a considerable discrepancy between the theoretical (22–26 kcal/mol) and experimental (15.4 kcal/mol) values for \(E_a\), as noted by the authors of experimental studies\(^{20,39,41}\). Below we demonstrate that this discrepancy is resolved when solvent effects are taken into account.
Aqueous solution: Continuum solvation models

Continuum solvation models\textsuperscript{12} provide a reasonable starting point in modeling ground-state isomerization in aqueous solution. At first, two versions of the polarized continuum model (PCM), D-PCM\textsuperscript{13} and C-PCM\textsuperscript{14}, were applied to optimize the equilibrium geometry parameters of the cis- and trans-isomers and the TS configuration and to compute the relative energies $\Delta E_{tc}$ and $E_a$ at the PBE0/6-31+G(d,p) level. Both models produced similar results: $\Delta E_{tc} = 2.1$ (D-PCM) and 2.3 kcal/mol (C-PCM); $E_a = 33.5$ (D-PCM), 34.0 kcal/mol (C-PCM). The geometry parameters and the energies are close to the gas phase values obtained with the same DFT model. These results are fairly stable with respect to the basis set: upon expanding it to 6-311++G(2d,p), $E_a$ and $\Delta E_{tc}$ change by less than 1.5 kcal/mol and 0.7 kcal/mol, respectively. Table 2.6 summarizes solvent effects on $E_{tc}$ and $E_a$ computed using different approaches.

Next, we considered a new version of the continuum solvation model, SVPE\textsuperscript{15}, which has recently been implemented in GAMESS(US)\textsuperscript{9}, for the single point calculations of relative energies $\Delta E_{tc}$ and $E_a$ at the gas-phase geometry parameters. At the PBE0/6-31+G(d,p) level, we obtained a considerable reduction of the activation energy compared to the PCM results: $E_a = 24.6$ kcal/mol, while the energy $\Delta E_{tc} = 2.1$ kcal/mol was almost the same as in the PCM model. According to a comment by Chipman\textsuperscript{15}, the improvements introduced in SVPE do affect the reaction barrier estimated with the continuum solvation models, especially for charged substrates. The large solvent effect on $E_a$ (9.1 kcal/mol reduction relative to the gas-phase value) is consistent with the more ionic character of the TS (see Table 2.5).

Finally, we combined the SVPE model with the adequate description of the electronic structure of the solute and carried out the calculations of relative energies with the CASSCF(12/11)/cc-pVDZ description for the chromophore. The quantities, $\Delta E_{tc} = 2.6$
and \( E_a = 9.9 \) kcal/mol, can now be directly compared to the experimental free energy difference of 2.3 and 15.4 kcal/mol\(^{39}\). The reduction of about 15 kcal/mol in \( E_a \) when going from DFT to CASSCF results within the SVPE solvation model is consistent with the corresponding reduction (13 kcal/mol) in the gas-phase calculations. Moreover, DFT and CASSCF agree on the magnitude of the reduction of \( E_a \) due to solvent (9.1 kcal/mol and 12.6 kcal/mol, respectively). Thus, the absolute value of \( E_a \) is overestimated by DFT due to the multiconfigurational character of the electronic structure at the TS geometry, which is correctly captured by CASSCF.

**Aqueous solution: Explicit solvent molecules**

The relative energies \( \Delta E_{tc} \) and \( E_a \) for the HBDI anion inside a cluster of water molecules were computed using a QM/MM technique. To build the starting point, the chromophore molecule at the gas-phase TS geometry was placed in a sphere of 200 water molecules using the VMD computer program\(^{42}\). After running short molecular dynamics trajectories at various temperatures, the energy was minimized using molecular mechanics with the CHARMM force field\(^{43}\) by keeping the solute species frozen. Then the solvent outside the first solvation shell was removed with 49 water molecules completely covering the chromophore remaining in the system. This saddle point structure was re-optimized for all geometric degrees of freedom by QM/MM using DFT with the PBE0 functional and the 6-31+G(d,p) basis for the QM-part (chromophore), EFP\(^{16}\) for QM-solvent interactions, and the empirical TIP3P potential for water-water interactions. The steepest descent pathways taken in both directions from the optimized TS lead to the \textit{cis}- and \textit{trans}-isomers of the trapped chromophore (Fig. 2.7). The energies obtained with this method, \( \Delta E_{tc} = 5.0 \) kcal/mol and \( E_a = 26.0 \) kcal/mol, are consistent with the DFT/SVPE calculations (Table 2.6).
Figure 2.7: Cis-trans isomerization of the chromophore inside a cluster of water molecules. Relative energies of the stationary points are computed with QM(CASSCF(12/11)/cc-pVDZ)/EFP/MM(TIP3P).
Table 2.6: Cis-trans energy difference $\Delta E_{tc}$ and the energy barrier $E_a$ for the cis-trans isomerization of the HBDI anion calculated at various levels of theory for the chromophore molecule and solvent.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\Delta E_{tc}$, kcal/mol</th>
<th>$E_a$, kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gas phase</td>
<td>Solution</td>
</tr>
<tr>
<td>PBE0/6-31+G(d,p)</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>D-PCM</td>
<td>2.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>C-PCM</td>
<td>2.3</td>
<td>+0.0</td>
</tr>
<tr>
<td>SVPE</td>
<td>2.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>QM/MM$^a$</td>
<td>5.0</td>
<td>+2.7</td>
</tr>
<tr>
<td>CASSCF(12/11)/cc-pVDZ</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>SVPE</td>
<td>2.6</td>
<td>-0.9</td>
</tr>
<tr>
<td>QM/MM$^a$</td>
<td>2.1</td>
<td>-1.4</td>
</tr>
<tr>
<td>MRMP2/cc-pVDZ</td>
<td>3.7</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Calculated with QM/MM: EFP for the solvent-QM part interaction and TIP3P for the water-water interaction.
The energies of the cis-, trans- and TS structures were also calculated using CASSCF(12/11) for in the QM-part. The interactions between the chromophore molecule and solvent and between the water molecules were handled with EFP and TIP3P, respectively. In line with the gas-phase and dielectric continuum model results, the activation energy is lower at the CASSCF level relative to DFT: $\Delta E_{tc} = 2.1 \text{ kcal/mol}$, $E_a = 11.1 \text{ kcal/mol}$.

The relative energies can be further refined by including the dynamical correlation effects. In the gas phase, accounting for dynamical correlation via MRMP2 raises the CASSCF activation barrier by 3.7 kcal/mol. Applying that correction, we estimate the activation energy for the cis-trans isomerization of the GFP chromophore in aqueous solution to be 14.8 kcal/mol. The remaining discrepancy with the experimental value of $15.4 \text{ kcal/mol}^{39}$ can be partly attributed to the differences between the potential energy barrier and the free energy barrier. The present QM/MM model with only 49 explicit water molecules is not sufficient for describing the statistical state in solution. This water solvation shell completely covers the chromophore molecule and accounts for principal environmental effects, but precise free energy values along the isomerization pathway should be estimated by using more appropriate condensed phase models. As shown, e.g. in Ref. 108, no considerable changes in conclusions are expected if the dielectric continuum model is applied on top of the model with explicit water molecules.

2.4 Chapter 2 conclusions

In this work we exploit modern quantum chemical methods for calculations of the electronic properties of the GFP chromophore and compare the results to the gas-phase absorption spectrum obtained by photodestruction spectroscopy in the ion-storage ring$^{3,28,29}$. 
The experimental action spectrum of the denatured gas phase anionic GFP chromophore features a broad line (2.4–2.8 eV) with a maximum at 2.59 eV and a minor feature at 2.3 eV. Wavefunction-based and DFT calculations estimate VDE of 2.4–2.5 eV. Thus, we assign the minor peak as due to the photodetachment transition. Based on our estimate of VDE, the absorption band at 2.6 eV corresponds to the transition to the resonance state embedded in an electron detached continuum, and the broad character of the spectrum is at least partially due to the interaction with the continuum states.

The resonance nature of the $\pi\pi^*$ state suggest a finite lifetime, and that autoionization channel should be considered when modeling the anionic GFP photocycle. The triplet state is found to be well below the photodetachment threshold (vertical excitation energy 1.86 eV). Thus, the two states are expected to have very different lifetimes, which makes the suggested$^3$ population trapping in the triplet state even more essential for explaining slow fragmentation kinetics. The resonance nature of the $\pi\pi^*$ state in the anionic GFP might be responsible for very different behavior of the photofragment yield of the anionic and protonated GFP$^3$, however, more detailed electronic structure calculations are required in order to suggest a viable mechanism. An important question is how photoinduced isomerization and other structural changes$^{40,41}$ affect the relative states energies.

All wavefunction-based and TD-DFT methods agree on the nature of the transition lending the intensity to the resonance state, which is a bright $\pi\pi^*$ transition (HOMO-LUMO in a small basis set), however, quantitative agreement is more difficult to achieve. Most importantly, small basis set calculations discretize the ionization continuum, and the results of such calculations provide only a crude estimate of the energy of the resonance state. In order to account for basis set effects, the stabilization analysis can be
used; however, in view of the large size of the GFP chromophore molecule, we were able to only conduct it with the CIS and BNL methods.

Nevertheless, it is instructive to compare vertical excitation energies of the $\pi\pi^*$ state computed with different methods in a moderate basis set and with the experimental band maximum. While the CIS/aug-cc-pVTZ excitation energy is more than 1 eV off, perturbative inclusion of double excitations by SOS-CIS(D) yields a value which is within 0.1 eV from the experimental band maximum. The EOM-CCSD values computed in the modest 6-31+G* basis are within 0.38 eV from the experimental absorption maximum. An analysis of the EOM-CCSD wavefunction confirms the dominant one-electron character of the $\pi\pi^*$ state, however, perturbative inclusion of triple excitations and a larger basis set are required for a converged (with respect to the level of theory) EOM-CC value. Based on previous studies, a proper account of dynamical correlation by including triple excitations and increasing basis set can change the vertical excitation energy by as much as 0.3 eV. Basis set effects evaluated using inexpensive SOS-CIS(D) calculations affect vertical excitation energies by 0.14 eV upon the transition from cc-pVDZ to aug-cc-VTZ.

Additional uncertainties arise from the ground state geometry. For example, different choices of the ground-state geometry (optimized with CASSCF, DFT, and MP2) introduce an uncertainty of 0.1 eV in the vertical excitation energies. A relatively strong dependence of the excitation energy on the structure suggests additional broadening of the absorption band due to vibrational excitations of the chromophore.

The best MRMP2 estimate is within 0.07 eV from the experimental band maximum, however, the inclusion of basis set effects (using the SOS-CIS(D)/aug-cc-pVTZ estimate) increases the difference to 0.21 eV. Overall, the observed variations demonstrate
that it is unrealistic to expect an accuracy better than 0.1 eV from computational protocols applicable to a molecule of this size even for non-resonance excited states.

BNL/cc-pVDZ vertical excitation energy of the bright $\pi\pi^*$ transition is above MRMP2/cc-pVDZ value by 0.9 eV. Since the self-interaction error is considerably reduced in BNL, the photodetachment continuum is likely not contaminated by spurious low-lying charge-transfer states ubiquitous in TD-DFT calculations.

Calculations of the cis-trans isomerization pathway require wavefunctions that are flexible enough to reflect the multiconfigurational character of the transition state. CASSCF gives a qualitatively correct curve, whereas DFT and SOS-MP2 fail at the twisted geometries. The gas-phase CASSCF value of the barrier height is 10 kcal/mol higher than the experimental value, however, the inclusion of solvent effects brings it down to 9.9–11.1 kcal/mol, which agrees well with 15.4 kcal/mol derived from experimental measurements. Including dynamical correlation correction yields 14.8 kcal/mol, which is within 0.8 kcal/mol from the experiment. Good agreement between SVPE and QM/MM calculations further supports the validity of our results. Large solvent effect on $E_a$ is due to the more ionic character of the TS, in which negative charge is localized on imidazolin ring, in contrast to the equilibrium structure in which the negative charge is delocalized between the two rings. Our calculations help resolve previous disagreement between theory and experiment with respect to the GFP chromophore isomerization in aqueous solution.
2.5 Chapter 2 references


Chapter 3: The effect of oxidation on the electronic structure of the green fluorescent protein chromophore

3.1 Chapter 3 introduction

The unique properties of green fluorescent protein (GFP) exploited in novel bioimaging techniques have revolutionized many areas in life sciences\textsuperscript{1–3}. The photophysics of GFP and other wild-type and mutated photoactive proteins have been a subject of many experimental and theoretical studies\textsuperscript{4,5}. Yet, intrinsic electronic properties of the chromophores and their interactions with the protein matrix are not fully understood.

The absorption spectrum of the wild-type GFP has two broad (0.2–0.4 eV) bands at 3.13 eV (396 nm) and 2.60 eV (476 nm) assigned to a neutral and an anionic (deprotonated) forms of the chromophore, respectively. The fluorescence spectrum is also broad and has a maximum at 2.44 eV (508 nm). Detailed mechanistic studies have demonstrated that while both the neutral and the anionic forms of the GFP chromophore are responsible for absorption, fluorescence occurs from the anionic form (see, for example, recent reviews\textsuperscript{4,5} on the photophysics of GFP and references therein).
The absorption and fluorescence wave lengths and quantum yields can be controlled by structural modifications, and a number of GFP mutants and other photoactive proteins have been characterized\textsuperscript{5,6}. The changes in photophysics can be due to structural modifications of the chromophore itself or in the surrounding protein. Moreover, the photochemical properties of some chromophores from the GFP family can be affected by light giving rise to so-called optical highlighting techniques. By using light of varying intensity and duration, the fluorescence can be turned on and off (photoswitching), and its spectrum (both wave length and intensities) can be changed (photoconversion). Both reversible and irreversible photoconversion have been described in the literature\textsuperscript{5,7}.

Photoconversion phenomena are not yet well understood at the molecular level. The established mechanisms involve photoinduced isomerizations (e.g., cis-trans transformation), changes in protein microenvironment, as well as more significant structural changes such as bond breaking, proton and charge transfer between the protein and the chromophore\textsuperscript{5,7}. For example, the intensity of GFP fluorescence is increased by two orders of magnitude upon irreversible photoconversion by UV or visible light (4.9–2.6 eV) to the anionic form accompanied by the decarboxylation of an adjacent glutamine residue\textsuperscript{8,9}. That also leads to a change in the absorption spectrum: increased absorbance of blue light and diminished absorbance in the UV region. In another FP, Dronpa, photoconversion between a protonated dark and an anionic bright forms is reversible and involves cis-trans isomerization\textsuperscript{10}. A different photoconversion type is exemplified by Kaede FP, which can be irreversibly switched from green to red fluorescent form by the cleavage of a peptide bond leading to the extension of the \( \pi \)-system of the chromophore\textsuperscript{7}. 

82
Photoinduced redding of GFP is a photoconversion that has not yet been explained. It was observed\textsuperscript{11} that under anaerobic conditions illumination of a GFP mutant (GFP-mut2) with blue light (2.5–2.7 eV or 460–500 nm) for several seconds results in red fluorescence at 1.8–2.0 eV (610–680 nm). Similar behavior was observed for wild-type GFP and other proteins. GFP photoconversion was more efficient when using higher energy UV light (3.3–3.7 eV or 340–380 nm)\textsuperscript{11}. A more detailed study of Elowitz et al reported that illumination of GFP by blue light converts GFP to a stable (under anaerobic conditions) form that absorbs green light at 2.36 eV (525 nm) and emits green, yellow and red light (emission maxima at 2.21 eV, 2.10 eV and 2.07 eV)\textsuperscript{12}. Based on the measured timescale of brightening (0.7 sec), this study suggested that the photoactivation proceeds in two steps: blue light (2.54 eV, 488 nm) stimulates a fast transition to an excited intermediate which then decays slowly to the red-emitting GFP state (the second step can proceed in the dark).

Recently, photoinduced redding of GFP has been observed under aerobic conditions\textsuperscript{13}. In the presence of different oxidizing agents (e.g., K$_3$Fe(CN)$_6$, benzoquinone, etc), irradiation with 2.54 eV (488 nm) light results in the bleaching of the green fluorescence and the appearance of the red signal with excitation and emission maxima at 2.16 eV (575 nm) and 2.04 eV (607 nm), respectively. Measurements at different light intensities determined that the redding is a one-photon process. Based on kinetic data, which showed that the green fluorescence disappears faster than the red fluorescence builds up, a two-step mechanism involving oxidation of GFP was suggested\textsuperscript{13}. By testing a variety of electron acceptors, it was found that the redding was achieved using compounds with the standard reduction potential $E^\circ$ in the range of $+0.42$ to $-0.32$ V. The yield of the redox reaction between GFP and cytochrome c was found to be 1.7 suggesting that a two-electron oxidation of GFP is responsible for the redding (that is,
two reduced cytochrome c molecules are produced per one converted GFP molecule). The stability of the red form under the aerobic conditions points out that the mechanisms of anaerobic and oxidative redding are, most likely, different. Moreover, the excitation spectra of the two forms are different: the maxima are 2.36 and 2.16 eV (525 and 575 nm) for anaerobic and oxidative redding forms of GFP, respectively. Bogdanov et al. hypothesized that anaerobic redding proceeds by photoreduction producing a stable radical (which will be sensitive to the presence of oxygen). Their findings suggest that oxidative redding produces more stable species. Other GFP variants, i.e., AcGFP1, TagGFP, zFP506, amFP486, and ppuGFP2 also undergo oxidative redding, however, the blue and cyan mutants (that have notably different chromophores) of GFP (EBFP and ECFP) do not.

Theoretical modeling noticeably contributes to the studies of the GFP and related model systems (for a recent review, see Ref. 25). In this chapter we characterize the electronic structure of two different oxidized states of the deprotonated 4'-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) anion (Fig. 3.1). We consider singly and doubly oxidized forms and discuss excitation energies and oscillator strengths of the electronic transitions in these species. To avoid confusion, it is important to clarify the terminology. Here we refer to deprotonated HBDI as the HBDI anion, or simply anion. The (neutral) species derived by removing one electron from the anion is called the HBDI radical, and doubly oxidized deprotonated HBDI is called the HBDI cation. We also present the relevant ionization energies and estimate standard reduction potentials corresponding to these oxidation steps. The calculations are performed for the isolated gas-phase chromophores, however, possible effects of the solvent on reduction potentials are also discussed.
Figure 3.1: Structures of deprotonated HBDI (HBDI anion, top) and the cation (bottom). The structure of the radical is similar to that of the cation, with the unpaired electron being on the carbon that hosts positive charge in the cation.

The protein environment is crucial for the function of the chromophore. The rigid protein matrix restricts the chromophore range of motions and impedes twisting motion in the excited state, which strongly affects the \( S_1 - S_0 \) oscillator strength. For example, the chromophore in solution does not fluoresce, and the fluorescence is restored only in glass matrices\(^4\). The photoinduced conversion of the neutral to the anionic form also
involves the protein. It is facilitated by a hydrogen bond network\textsuperscript{26–28} and can be coupled to the decarboxylation of the glutamate 222 residue\textsuperscript{8,9}. The electrostatic field of the protein may induce spectral shifts, although gas-phase studies suggested that the effect of the protein environment on the excited states energies may be relatively small\textsuperscript{29,30}. Despite the role of the protein, an important step towards understanding of the function of GFP is characterization of the intrinsic properties of the isolated chromophore. By focusing on the electronic properties of the chromophore, we can quantify changes in absorption and fluorescence due to specific structural changes and to test different mechanistic ideas. Our calculations show that the absorption of the singly oxidized species (radical) is blue shifted with respect to the HBDI anion. However, the doubly oxidized system (cation), which has a stable closed-shell electronic structure, features red-shifted absorption at 2.02 eV. The proposed mechanism involves two-step oxidation proceeding via electronically excited states.

3.2 Computational details

Equilibrium geometries of closed-shell ground-state species (i.e., the HBDI anion and cation) were optimized by RI-MP2/cc-pVTZ. The ground state structure of the doublet radical was computed using \( \omega \text{B97X/6-311+(+G(2df,2pd))} \). The harmonic vibrational frequencies and IR intensities of the anionic and cationic forms of deprotonated HBDI were computed with RI-MP2/cc-pVTZ.

Excitation energies of the cation were computed using SOS-CIS(D) and EOM-EE-CCSD. SOS-CIS(D) calculations employed the cc-pVTZ and cc-pVDZ bases\textsuperscript{31}. EOM-CCSD calculations used the 6-311G* basis set\textsuperscript{32,33}. For the anion, SOS-CIS(D) excitation energies are in excellent agreement with more accurate MR-PT. The oscillator strengths for SOS-CIS(D) were computed using the underlying CIS wave functions,
whereas for EOMEE-CCSD, the transition density matrices computed using EOMEE wavefunctions were employed.

Electronic structure calculations for the open-shell radical species are more challenging due to spin contamination of the doublet reference. This effect can be mitigated by using better reference orbitals, e.g., those optimized for the MP2 wave functions, i.e., O2 method. We employed O2 orbitals in the SOS-CIS(D) calculations of the doublet states. Moreover, excited states in doublet radicals can also develop a notable doubly excited character causing poor performance of single-reference methods using the doublet reference. This can be avoided by employing the spin-flip (SF) methods utilizing a high-spin quartet reference, i.e., $(\pi_2)^1(\pi_1)^1(\pi^*)^1$. The SF approach mitigates spin contamination and provides more balanced description of the ground and relevant excited states. Our EOMSF-CCSD calculations (which employed B3LYP orbitals to control spin-contamination of the quartet reference) showed that the wave functions of the excited states are of singly-excited character (with respect to the doublet reference) thus validating the SOS-CIS(D) results. The $\langle S^2 \rangle$ values for the three lowest states of the doublet are 0.78, 0.79, and 0.81. The EOMSF-CCSD and SOS-CIS(D) calculations are in quantitative agreement. Finally, MR-PT calculations on the doublet states are also in agreement with EOMSF-CCSD and SOS-CIS(D).

MR-PT calculations employed a modified version of second-order multiconfigurational quasidegenerate perturbation theory (MCQDPT2). Originally, the corresponding programs were implemented in GAMESS (US) and PC GAMESS. Recently, serious bugs in the MCQDPT2 algorithm were noticed and resolved in the new variant called XMCQDPT2 as implemented in PC GAMESS/Firefly.

The XMCQDPT2 calculations used the complete active space self-consistent field (CASSCF) state-averaged (SA) electron density over several lowest-lying states. As
CASSCF does not account for dynamical electron correlation\textsuperscript{45}, the CASSCF state ordering is often different from the correct order predicted by higher-level calculations. This is the case for the doublet radical: the fourth lowest-lying state in the XMCQDPT2 calculation appears the seventh in CASSCF. Since it is one of the states of interest (i.e., the $D_2$ state whose orbital character is described below), we had to use an extensive state-averaging procedure including eight states in the consideration. The active space for SA-CASSCF was chosen as 12 electrons distributed over 11 orbitals, which is a truncated representation of the complete set of 14 $\pi$-orbitals forming the conjugated system of the chromophore. The MCQDPT2 based approaches have been shown to yield accurate results for a variety of chromophores\textsuperscript{41,46,47}. The underlying CASSCF wave functions are free of spin contamination.

The energy differences between different oxidized states were computed as follows. The vertical energy difference between the anion and the neutral radical, i.e., the first VDE, was computed by EOMIP-CCSD/6-311G* (2.33 eV) and corrected for basis set increase from 6-311G* to aug-cc-pVTZ using a Koopmans estimate (0.21 eV). Thus, our best estimate of VDE is 2.54 eV.

The vertical energy gap between the cation and the anion is computed as the difference between the RI-MP2/cc-pVTZ total energies of the respective species. The differences between the doublet and the cation were computed from the relative energies of the cation and the anion and the anion’s VDE. Finally, the adiabatic values were obtained from vertical gaps by using relaxation energies of the neutral and the cation, which were computed by RI-MP2/cc-pVTZ (cation) and $\omega$B97X/6-311(+/+)
(2df,2pd) (radical) using the geometries described above.

EOM and SOS-CIS(D) calculations were performed using Q-Chem\textsuperscript{48}. XMCQDPT2 calculations were carried out with PC GAMESS/Firefly\textsuperscript{43}. 

88
Table 3.1: Vertical excitation energies (eV) and oscillator strengths of the $\pi\pi^*$ states of the doublet HBDI radical. The values of $\langle S^2 \rangle$ are also shown.

<table>
<thead>
<tr>
<th>Method</th>
<th>D1</th>
<th>$\langle S^2 \rangle$</th>
<th>D2</th>
<th>$\langle S^2 \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS-CIS(D)/cc-pVDZ$^a$</td>
<td>1.64 (0.02)</td>
<td>1.04</td>
<td>3.53 (1.35)</td>
<td>0.90</td>
</tr>
<tr>
<td>SOS-CIS(D)/cc-pVTZ$^a$</td>
<td>1.52 (0.02)</td>
<td>1.01</td>
<td>3.37 (1.34)</td>
<td>0.89</td>
</tr>
<tr>
<td>EOM-SF-CCSD/6-311G*$^b$</td>
<td>1.93 (0.01)</td>
<td>0.79</td>
<td>4.14 (1.04)</td>
<td>0.81</td>
</tr>
<tr>
<td>XMCQDTP2/6-311G*$^c$</td>
<td>1.80 (0.01)</td>
<td></td>
<td>3.27 (0.66)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Using O2 orbitals. Oscillator strength are computed at the CIS level. $^b$ Using B3LYP orbitals. $^c$ Based on SA(8)-CASSCF(12/11).

3.3 Results and discussion

Relevant MOs and electronic states of the gas-phase HBDI anion are shown in Figs. 3.2 and 3.3. Gas-phase photodestruction spectra show the absorption maximum of the anionic GFP chromophore at 2.59 eV$^{29,30}$, which is remarkably close to the absorption of wild-type GFP at 2.61 eV. Most of the electronic structure studies agree on the character of the bright state$^{14–17, 19, 21–23, 49}$, which is of a $\pi \to \pi^*$ type (Fig. 3.2), although there is a spread in reported excitation energies. Our recent theoretical study revealed that the bright state is, in fact, a resonance state embedded in a photodetachment continuum. These calculations determined that the vertical detachment energy (VDE) of the HBDI anion is 2.4–2.5 eV, which is below the bright state. A recent experimental study$^{50}$, which also employed electronic action spectroscopy, reported the vertical excitation energy of 2.60 eV and determined VDE to lie within 2.48–2.8 eV, in excellent agreement with our computational predictions.
As evident from Fig. 3.3, the electronic excitation using 2.6 eV (and higher) produces an excited state that can photodetach an electron producing a doublet neutral radical. Thus, our first step is to characterize electronic transitions of this form. Three lowest electronic states of the doublet and the corresponding excitation energies and oscillator strengths are shown in Fig. 3.4 and summarized in Table 3.1. The ground state of the doublet is derived by the ionization from the anion’s HOMO ($\pi_1$ from Fig. 3.2). The two excited states are of a mixed character and include almost equal contributions from the $\pi_1 \rightarrow \pi^*$ and $\pi_2 \rightarrow \pi^*$ excitations. Consistently with the allylic character of the three relevant MOs, both excitations have large transition dipole moments, and their plus and minus combinations give rise to the bright and dark states at 4.14 eV ($f_i = 1.04$) and
The bright absorbing $\pi_1 \pi^*$ state at 2.62 eV is a resonance state embedded in a photodetachment continuum beginning at 2.54 eV. The respective MOs are shown in Fig. 3.2.

1.93 eV ($f_i = 0$). The SOS-CIS(D) and MR-PT calculations agree with EOM-SF-CCSD on the character of the excited states and yield slightly lower excitation energies. Overall, the bright state of the doublet is strongly blue-shifted relative to the anionic form. Therefore, the doublet radical is an unlikely candidate for the red GFP form. The energy spacing between the two excited doublet states is 2.21 eV and the oscillator strength is non-negligible (0.01). Thus, red-shifted electronic transitions between $D_2$ and $D_1$ are possible.

Doubly ionized HBDI anion (i.e., the cation) is a well-behaved closed-shell system derived by removing two electrons from $\pi_1$. Its electronic states are summarized in
Table 3.2: Vertical excitation energies (eV) and oscillator strengths of the excited states of the HBDI cation at its equilibrium geometry.

<table>
<thead>
<tr>
<th>Method</th>
<th>$^1A'$ ($f_i$)</th>
<th>$^1A''$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOS-CIS(D)/cc-pVDZ$^a$</td>
<td>2.10 (1.20)</td>
<td>2.14</td>
</tr>
<tr>
<td>SOS-CIS(D)/cc-pVTZ$^a$</td>
<td>2.02 (1.17)</td>
<td>2.09</td>
</tr>
<tr>
<td>EOM-CCSD/6-311G*</td>
<td>2.53 (0.78)</td>
<td>2.10</td>
</tr>
<tr>
<td>XMCQDPT2/6-311G*$^b$</td>
<td>2.27 (0.84)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$ Oscillator strength are computed at the CIS level. $^b$ Based on SA(2)-CASSCF(12/11).

Figure 3.5: Electronic states of the doubly ionized HBDI anion. The respective MOs are shown in Fig. 3.2.

Fig. 3.5 and in Table 3.2. The lowest singlet excited state ($^1A'$) is at 2.02 eV (SOS-CIS(D)/cc-pVTZ level) and has a large oscillator strength. It is red-shifted relative to the $S_0 \to S_1$ transition in the anion by 0.6 eV, which is similar in magnitude to the observed experimental shift in absorption. The EOM-EE-CCSD excitation energy is higher, but the shift relative to the EOM-EE value of the $S_1$ excitation energy in the anion (at 3.1 eV at the EOM-EE/6-311G* level) is very close to the SOS-CIS(D) one. Thus, this species is a possible candidate for the red fluorescent form of GFP produced in oxidative redding experiments$^{13}$.

The red shift is consistent with the changes in the character of the electronic transition, which is a $\pi_2 \to \pi_1$ excitation, as opposed to the $\pi_1 \to \pi^*$ transition in the anion.
Thus, in terms of the anionic orbitals, the $S_1$ state of the cation is HOMO $- 1 \rightarrow$ HOMO excitation. Because of a relatively dense spectrum of the occupied orbitals (which becomes even denser in the positively charged system), the gap between HOMO $- 1$ and HOMO is smaller than the gap between the HOMO and the LUMO. Thus, double ionization is expected to result in a red shift in the lowest electronic transition.

Interestingly, the $\pi_2 \rightarrow \pi^*$ state of the cation is much higher in energy (approximately at 3.16 eV, $f_l = 0.0261$). The closest electronic state is a dark state, $1A''$ derived from the HOMO $- 3 \rightarrow$ HOMO transition (in terms of the anionic orbitals).
Fig. 3.6 shows an overall energy diagram for the forms of HBDI. Adiabatically, the doubly oxidized form is 9.98 eV above the ground state of the anion. The respective value of VDE (corresponding to removing two electrons) is 10.37 eV. The adiabatic IE from the ground state of the doublet radical is 7.59 eV (computed using the anions’s VDE value of 2.54 eV and 0.15 eV relaxation energy of the neutral radical). Another relevant value is the energy gap between the excited states ($D_1$ and $D_2$) of the doublet radical and the cation. Using the same values of vertical detachment and relaxation energies, and 1.52 eV and 3.37 eV for the vertical $D_0 \rightarrow D_{1,2}$ excitation energies, we estimate the ionization energy of the electronically excited doublet radical as 6.07 eV and 4.22 eV, respectively. Finally, the energy gap between the excited singlet state of the anion and the two excited states of the radical are 1.29 eV and 3.14 eV, respectively.

To evaluate whether two-electron oxidation of the anion is possible using the electron acceptors employed in Ref. 13, it is necessary to estimate standard reduction potentials corresponding to the detachment and ionization transitions described above. Rigorous calculations of the oxidation/reduction potentials require free energy calculations, however, a simpler approach based on the correlation between detachment/ionization energies and the standard reduction potentials can be employed:

$$E^o = (-2.59 \pm 0.26) + (0.56 \pm 0.03) \cdot \text{VIE},$$

where $E^o$ corresponds to the following half-reaction:

$$A^+ + e = A$$

and VIE is the ionization energy of $A$. This equation was derived by comparing experimental values of the reversible oxidation potentials (in DMF or ACN solvent) for a set of
Table 3.3: Standard oxidation potentials corresponding to different one-electron oxidation transitions computed using gas-phase detachment and ionization energies (see text).

<table>
<thead>
<tr>
<th>Transition</th>
<th>IE, eV</th>
<th>$E^\circ$, V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_0 \rightarrow$ cation</td>
<td>7.59</td>
<td>1.66 ± 0.49</td>
</tr>
<tr>
<td>$D_1 \rightarrow$ cation</td>
<td>6.07</td>
<td>0.81 ± 0.44</td>
</tr>
<tr>
<td>$D_2 \rightarrow$ cation</td>
<td>4.22</td>
<td>−0.23 ± 0.39</td>
</tr>
<tr>
<td>$S_1 \rightarrow D_1$</td>
<td>1.29</td>
<td>−1.87 ± 0.30</td>
</tr>
<tr>
<td>$S_1 \rightarrow D_2$</td>
<td>3.14</td>
<td>−0.83 ± 0.35</td>
</tr>
</tbody>
</table>

14 organic molecules with their gas-phase ionization energies. Applying this equation, we arrive at the values of $E^\circ$ summarized in Table 3.3. Eq. (3.1) was derived using IEs in the range of 5.6–9.0 eV and its application to lower values (such as $S_1 \rightarrow D_{1,2}$ transitions from Table 3.3) assumes that the trend can be extrapolated. Thus, estimations for the lower IEs are of a semiquantitative value only.

Oxidation of the ground-state doublet radical ($D_0$) corresponds to $E^\circ = 1.66$ V, which is much larger than $E^\circ$ of the strongest oxidizing agent employed in Ref. 13 ($K_3Fe(CN)_6$, $E^\circ = 0.42$ V). However, the oxidation of the electronically excited neutral radical corresponds to $E^\circ = 0.81$ and $E^\circ = −0.12$ V for the $D_1$ and $D_2$ states, respectively. Thus, electronically excited doublet radical (in the $D_2$ state) can be oxidized even by the weakest oxidizing agent from Ref. 13 ($E^\circ = −0.32$ V), whereas the oxidation of $D_1$ is less likely. Finally, the electronically excited anion can be easily oxidized producing excited states of the doublet ($E^\circ = −1.87$ and $E^\circ = −0.83$ V for $D_1$ and $D_2$ states, respectively). Note that the $S_1 \rightarrow D_0$ transition is spontaneous in the gas phase and does not require the presence of oxidizing agents due to the autoionizing nature of $S_1$. By
comparing electronic configurations of the doublet states (see Fig. 3.4) and the $S_1$ state (Fig. 3.3), we note that these transitions include significant one-electron character.

Thus, a possible mechanism of the oxidative redding emerging from the present study is

![Mechanism diagram](image)

The first step involves photoexcitation, and blue light is sufficient to generate this transition. The second and third steps are one-electron oxidation steps with $E^\circ = -0.88$ V, which is within the reach of the oxidizing agents employed in Ref. 13. The two-electron oxidation is consistent with the stoichiometric results from Ref. 13. Moreover, the closed-shell character of the cation is consistent with the chemically stable nature of the red form of GFP. Moreover, the closed-shell character of the cation is consistent with the relatively chemically stable nature of the red form of GFP. Although the chemical structure of the cation suggests high reactivity with respect to nucleophilic agents such as water, the protein may be sufficiently effective in shielding this fragile species. Finally, the absorption of the cationic form is red-shifted by 0.6 eV, and the resulting value of 2.02 eV is in a good agreement with the experimental excitation energy of 2.12 eV. The structural differences between the chromophores of the FPs that undergo oxidative redding (AcGFP1, TagGFP, zFP506, amFP486, and ppuGFP2), and those that do not (EBFP and ECFP) provide additional support to the proposed mechanism. The chromophores in the former group contain a phenolic moiety that can be easily converted to the anionic deprotonated form and thus oxidized, whereas in EBFP and ECFP the imidazolin moiety is connected to nitrogen-containing aromatic fragments that are expected to be basic rather than acidic.
Whereas the present model is consistent with the available experimental information, it cannot rule out possible involvement of the protein via proton transfer or other chemical changes (such as covalent bond breaking leading to the extension of the π-system of the chromophore\(^7,24\)). Thus, it would be desirable to observe oxidative reddening in the isolated chromophore, e.g., in solution. However, it is essential for the proposed mechanism that the excited states involved in oxidation, \(S_1\) of the anion and \(D_{1,2}\) of the radical, are relatively long-lived. Thus, it may be problematic to observe oxidative reddening of isolated HBDI in solution, where the internal conversion is markedly faster than in the protein (which is responsible for the loss of fluorescence), except for low-temperature glass matrices (in which fluorescence is restored). Finally, the protein environment may change the predicted energetics by stabilizing different states differently. However, the observed strong red shift in the doubly oxidized species originates in strong changes in the electronic structure of the chromophore, and it is unlikely that the energies of these transitions will be affected by the protein much stronger than the energies of the transitions in the anion. Our preliminary QM/MM calculations, which treated the chromophore and the nearest amino acid side chains as the quantum subsystem and the rest of the model protein system as the collection of effective fragments, do not show dramatic structural changes in the chromophore binding site for the anionic and cationic species. We also note that the present model does not address the nature of the anaerobic reddening of GFP, which will be discussed elsewhere.

**Ionization-induced structural changes**

As discussed in Chapter 2, the equilibrium structure the HBDI anion can be explained by considering two resonance structures shown in Fig. 3.1 (upper panel). Consequently, the negative charge is delocalized between the phenolic and imidazolin oxygens and the
Table 3.4: Selected geometric parameters (bond lengths in Å, angles in degrees) of the HBDI anion, and its singly and doubly oxidized forms.

<table>
<thead>
<tr>
<th>Species</th>
<th>C\textsubscript{P}–C\textsubscript{B}</th>
<th>C\textsubscript{I}–C\textsubscript{B}</th>
<th>C\textsubscript{P}–C\textsubscript{B}–C\textsubscript{I}</th>
<th>O\textsubscript{P}–C</th>
<th>O\textsubscript{I}–C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anion\textsuperscript{a}</td>
<td>1.393</td>
<td>1.385</td>
<td>134.8</td>
<td>1.248</td>
<td>1.236</td>
</tr>
<tr>
<td>Cation\textsuperscript{a}</td>
<td>1.386</td>
<td>1.389</td>
<td>127.8</td>
<td>1.226</td>
<td>1.197</td>
</tr>
<tr>
<td>Radical\textsuperscript{b}</td>
<td>1.413</td>
<td>1.369</td>
<td>129.6</td>
<td>1.228</td>
<td>1.209</td>
</tr>
</tbody>
</table>

\textsuperscript{a} RI-MP2/cc-pVTZ \textsuperscript{b} \omega B97X/6-311(++,+)G(2df,2pd)

C\textsubscript{I}–C\textsubscript{B} and C\textsubscript{P}–C\textsubscript{B} bond orders are scrambled. The analysis of the structural parameters and the wave functions demonstrated that the phenolic form dominates, i.e., the C\textsubscript{P}–C\textsubscript{B} bond is longer than C\textsubscript{I}–C\textsubscript{B}. Double ionization eliminates the phenolic structure entirely producing the quinoid structure shown in Fig. 3.1 (bottom). The positive charge located on the imidazolin carbon is stabilized by electron-donating methyl group (the NBO analysis\textsuperscript{52} suggests considerable delocalization of the hole over imidazolin moiety). The structure of the radical can be described as the quinoid structure with the unpaired electron on the imidazolin’s methyl-substituted carbon.

Table 3.4 summarizes important geometric parameters of the anion, neutral, and the cation. Consistently with the above analysis, both CO bonds become shorter in the cation, and the C\textsubscript{P}–C\textsubscript{B} bond becomes shorter than C\textsubscript{I}–C\textsubscript{B}. However, the bond alternation remains small due to conjugation. Moreover, because of the two resonance structures present in the anion, the overall structural differences between the anion and the cation are less than could be expected by considering only the dominant phenolate structure.

The ionization-induced structural changes strongly affect vibrational frequencies, which may be exploited in the experimental characterization of the red form of GFP, e.g., by using IR or Raman spectroscopy as in Refs. 9, 53 (a detailed summary of the vibrational spectroscopy studies of GFP can be found in Ref. 54). Fig. 3.7 shows the
Figure 3.7: Infrared spectra of the anionic and cationic forms of deprotonated HBDI computed with RI-MP2/cc-pVTZ.

computed IR spectra of anionic and cationic deprotonated HBDI. The anion spectrum is dominated by a band at 1691 cm$^{-1}$, which is an asymmetric CO stretch strongly coupled with an asymmetric bridge vibration. In the cation, this mode changes its character to an almost pure bridge vibration, loses intensity and is blue-shifted. Instead, those modes gain intensity that involve the imidazolin moiety hosting a positive charge. The two most intense bands are predominantly CH$_3$ vibrations at 1534 and 1435 cm$^{-1}$, followed by lower frequency bands involving skeletal imidazolin vibrations.

3.4 Chapter 3 conclusions

We characterized changes in the electronic structure of the deprotonated HBDI anion due to one- and two-electron oxidation processes. One-electron oxidation produces a doublet radical with strongly blue-shifted absorption, whereas the absorption in the cation is red-shifted by almost 0.6 eV with respect to the anion. From the electronic structure point of view, the cation has a closed-shell singlet ground state, and is expected to be relatively stable chemically.
Our results suggest that the doubly oxidized species (the deprotonated HBDI cation) may be responsible for the oxidative redding of GFP, which was recently reported\textsuperscript{13}. The proposed mechanism involves two-step oxidation via electronically excited states and is consistent with the available experimental information\textsuperscript{13}, i.e., it is a one-photon process including two-electron oxidation resulting in chemically stable species. Moreover, all proteins that undergo oxidative redding contain a phenolic moiety and are expected to have similar oxidative properties, whereas the two proteins that do not exhibit redding (EBFP and ECFP) have nitrogen-containing aromatic fragments that are less efficient electron donors. This suggests that the oxidative redding is an intrinsic property of the chromophore, although chemical changes involving the protein cannot be ruled out. Moreover, the irreversible nature of the oxidative redding\textsuperscript{13} suggests that the protein is likely to change in response to the changes in the chromophore’s oxidation state.

The closed-shell character of the cation is consistent with the relatively chemically stable nature of the red form of GFP. Although the chemical structure of the cation suggests high reactivity with respect to nucleophilic agents such as water, the protein may be sufficiently effective in shielding this fragile species.

Ionization-induced structural changes of the chromophore result in significant changes in vibrational spectra (the frequencies as well as the character of the modes and IR intensities) which may exploited in the experimental validation of the proposed mechanism.

We would like to emphasize that simple molecular orbital considerations predict red-shifted absorption for any doubly-oxidized species. Thus, while the predicted values of the absorption maximum may change once the protein environment is taken into account (or at a higher level of theory), the overall trend remains unchanged. This mechanism of
redding is distinctly different from previously characterized ones in which redding was achieved by extending the π-system of the chromophore.
3.5 Chapter 3 references


Chapter 4: Electronically excited states of uracil

4.1 Chapter 4 introduction

This chapter presents calculations of the electronic structure and the excited states of uracil using high-level MRCI and EOM-CC methods with the purpose of obtaining reliable and accurate theoretical estimates of vertical excitation energies. We investigate the dependence of the excitation energies on the basis set, the possible doubly-excited character of the states, as well as dynamical correlation effects using several EOM-CC methods with triple excitations\textsuperscript{1–5}.

Our best estimates of the vertical $n\pi^*$ and $\pi\pi^*$ excitations are 5.0 eV and 5.3 eV, respectively, with estimated error bars of less than 0.1 eV. We attribute the 0.2 eV difference between the computed vertical excitations and the maximum of the experimental absorption to a strong vibronic interaction between the lowest $A''$ and $A'$ states leading to intensity borrowing by the forbidden transition. Experimental observations of Kong \textit{et al.}\textsuperscript{6} suggest the strong vibronic coupling between the states. The EOM-CCSD/aug-cc-pVTZ excitation energy of the $n\pi^*$ state is only 0.23 eV higher than our best estimate, which is well within a conservative estimate of the EOM-CCSD error bars\textsuperscript{7}. Thus, no evidence of a significant overestimation of the excitation energies in
uracil by EOM-CCSD has been observed. We also investigate solvent effects, and pro-
vide accurate estimates for vertical excitation energies of triplet and other singlet states.

4.2 Computational details

Density functional theory (DFT) with the B3LYP\textsuperscript{8} functional and the 6-
311G(2df,2pd)\textsuperscript{9,10} basis set are employed for ground-state structure optimization.

Pople’s 6-31G\textsuperscript{11} and 6-311G\textsuperscript{9} basis sets with polarization and diffuse functions are
used in the study. Their performance is compared with Dunning’s cc-pVDZ and aug-
cc-pVTZ\textsuperscript{12} bases, as well as the NASA Ames atomic natural orbital (ANO)\textsuperscript{13,14} basis
set contracted as double-ζ and augmented with diffuse functions taken from the aug-
cp-pVDZ\textsuperscript{12} basis set. In many calculations, mixed basis sets are employed: a higher
quality basis is used on heavy atoms. The mixed basis sets are designated with a for-
ward slash; for example, 6-311+G(d)/6-31G(d,p) means that 6-311+G(d) is used on the
carbon, nitrogen, and oxygen atoms, and 6-31G(d,p) is used on the hydrogens. All basis
sets include the pure angular momentum spherical harmonics (5 d-functions).

The lowest eight core orbitals are frozen in all the calculations. Where specified, the
highest eight virtual orbitals are frozen as well.

Active spaces are denoted as \((n,m)\), which stands for \(n\) electrons in \(m\) orbitals.

The EOM-CC(2,3) and EOM-CCSDt calculations are performed with two active
spaces. The first one contains 10 electrons in 8 orbitals: five highest occupied (two
\(a'\) in-plane lone pairs and three \(a''\) \(\pi\) orbitals) and three lowest virtual orbitals (one \(a'\)
orbital and two \(a''\) \(\pi^*\) orbitals). We denote this active space as \((10,8)\). The second active
space extends \((10,8)\) by including two more lowest virtual orbitals, both of which are \(a'\).
This active space is denoted \((10,10)\).
Two different MRCI expansions are used. In both, the eight core and the highest eight virtual orbitals are frozen. The first expansion, denoted MRCI1, is based on a (14,10) active space, which consists of two $a'$ in-plane lone pair orbitals, five occupied $a'' \pi$ orbitals, and three virtual $a'' \pi^*$ orbitals. The MOs are obtained from a state-averaged CASSCF(14,10) calculation with five states included in the average. Single excitations outside the active space are allowed from the active and all $\sigma$ orbitals. The expansion consists of 16–32 million CSFs, depending on the basis set.

The second expansion (MRCI2) uses a smaller (12,9) active space, which includes only one $a'$ lone pair orbital and the same eight $a''$ orbitals as in (14,10). A previous study$^{15}$ showed that (12,9) and (14,10) produce very similar energies of the first two excited states, which are the most important for the photophysical properties of uracil. MRCI2 is used to refine those energies. The MOs are obtained from state-averaged CASSCF(12,9) with five states included in the average. Single and double excitations are allowed from the active space and only single excitations from the $\sigma$ orbitals. The Davidson correction is also computed for these MRCI2 energies. This MRCI2 expansion consists of about 100–330 million CSFs. Because of the size of the expansion, only the first two excited states are calculated at the MRCI2 level. Moreover, since the active space has only one $a'$ orbital, only one $A''$ state can be computed.

Several iterative and non-iterative EOM-CC methods, including EOM-CCSD, active-space EOM-CCSDt, and CR-EOM-CCSD(T) are fully interfaced with the molecular mechanics module of NWChem$^{16}$, which enables excited-state simulations of molecules in realistic environments within the quantum-mechanical/molecular-mechanical (QM/MM) framework. Excited-state studies of biologically relevant species have indicated need for balanced inclusion of correlation effects, as well as effects of
a fluctuating environment\textsuperscript{17–19}. In the present work, a similar approach is employed to compute the excitation energies of uracil in the aqueous solution.

Solvent effects are evaluated by the QM/MM approach\textsuperscript{20} as implemented in NWChem. The uracil molecule at the equilibrium ground-state geometry is embedded in a 30 Å-wide cubic box containing 887 water molecules. The QM region consists of the uracil molecule, and the rest of the system is treated at the MM level using the SPC/E water model\textsuperscript{21}. To ensure proper solvent structure around uracil, the solvent part of the system is first optimized, equilibrated over the course of 50 ps of molecular dynamics simulations, and then re-optimized again. During this simulation, the electrostatic field of uracil is represented by a set of fixed effective charges obtained from an electrostatic potential (ESP) fitting using B3LYP calculations with the 6-31G(d) basis set. Molecular dynamics equilibration is performed at the constant temperature of 298.15 K with a 15 Å cutoff. The resulting configuration is then used to calculate the vertical excitation energies using the B3LYP and EOM-CCSDt levels of theory with the 6-31G(d) basis set. The (10,10) active space is used in the QM part.

EOM-CCSD and EOM-CC(2,3) calculations are performed with Q-Chem\textsuperscript{22}, CRI-EOM-CCSD(T), EOM-CCSDt, and some EOM-CCSD computations are carried out with NWChem\textsuperscript{16,23}. The COLUMBUS\textsuperscript{24,25} suite of programs is used for MRCI calculations.

### 4.3 Results and discussion

#### 4.3.1 Structure and electronic states of uracil

In the ground electronic state, uracil is a planar molecule. Fig. 4.1 defines atomic labels in its diketo tautomer. Table 4.1 gives equilibrium geometrical parameters optimized
Figure 4.1: Structure of uracil defining atom labels referred to in Table 4.1.

by DFT and compares them to experimental values\textsuperscript{26} obtained by averaging 32 uracil residue structures found in a crystallographic database. The standard deviation of the experimental data is about 0.01 Å for distances and about 1.0° for angles. The differences between the calculated and experimental parameters do not exceed 0.03 Å for bond lengths and 2.5° for angles. The equilibrium structure also agrees well with the CC2/aug-cc-pVQZ optimized geometry reported by Fleig et al.\textsuperscript{27}: the maximum difference between the bond lengths is 0.015 Å.

Frontier MOs calculated with RHF/aug-cc-pVDZ are depicted in Fig. 4.2. 26 $a'$ and 27 $a'$ are occupied in-plane lone pairs on oxygen atoms. They lie below 28 $a''$ and 29 $a''$, two occupied π orbitals. Two virtual π* orbitals, 34 $a''$ and 35 $a''$, contain contributions from orbitals on all the heavy atoms. There are two low-lying diffuse orbitals: the lowest unoccupied MO 30 $a'$ and a higher orbital 38 $a'$.

The ππ* states correspond to transitions from the π orbitals 28 $a''$ and 29 $a''$ to the π* orbitals 34 $a''$ and 35 $a''$ (Table 4.2). Transitions originating from the lone pair orbitals 26 $a'$ and 27 $a'$ form the nπ* states. The transition from HOMO to the lowest diffuse orbital gives rise to a low-lying Rydberg state.
Figure 4.2: Frontier molecular orbitals of uracil at the ground state equilibrium geometry. Orbital energies are calculated with RHF/aug-cc-pVDZ.
Table 4.1: Equilibrium geometry of uracil. The calculated structure is optimized with B3LYP/6-311G(2df,2pd). Experimental values are obtained by averaging dimensions found in crystal structuresa. Atomic labels are defined in Fig. 4.1, distances in Å, angles in degrees.

<table>
<thead>
<tr>
<th>Bond</th>
<th>Calc. b</th>
<th>Exp.</th>
<th>Angle</th>
<th>Calc. b</th>
<th>Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₁–C₂</td>
<td>1.3908</td>
<td>1.379</td>
<td>C₆–N₁–C₂</td>
<td>123.57</td>
<td>121.3</td>
</tr>
<tr>
<td>C₂–N₃</td>
<td>1.3801</td>
<td>1.373</td>
<td>N₁–C₂–N₃</td>
<td>112.92</td>
<td>114.8</td>
</tr>
<tr>
<td>N₃–C₄</td>
<td>1.4092</td>
<td>1.383</td>
<td>C₂–N₃–C₄</td>
<td>128.15</td>
<td>127.0</td>
</tr>
<tr>
<td>C₄–C₅</td>
<td>1.4559</td>
<td>1.440</td>
<td>N₃–C₄–C₅</td>
<td>113.40</td>
<td>114.7</td>
</tr>
<tr>
<td>C₅–C₆</td>
<td>1.3426</td>
<td>1.338</td>
<td>C₄–C₅–C₆</td>
<td>119.97</td>
<td>119.2</td>
</tr>
<tr>
<td>C₆–N₁</td>
<td>1.3706</td>
<td>1.380</td>
<td>C₅–C₆–N₁</td>
<td>121.99</td>
<td>122.8</td>
</tr>
<tr>
<td>C₂–O₂</td>
<td>1.2084</td>
<td>1.218</td>
<td>O₂–C₂–N₃</td>
<td>124.43</td>
<td>122.0</td>
</tr>
<tr>
<td>C₄–O₄</td>
<td>1.2110</td>
<td>1.227</td>
<td>O₄–C₄–C₅</td>
<td>126.18</td>
<td>125.4</td>
</tr>
<tr>
<td>N₁–H₁</td>
<td>1.0060</td>
<td></td>
<td>H₁–N₁–C₂</td>
<td>115.03</td>
<td></td>
</tr>
<tr>
<td>N₃–H₃</td>
<td>1.0112</td>
<td></td>
<td>H₃–N₃–C₄</td>
<td>116.20</td>
<td></td>
</tr>
<tr>
<td>C₅–H₅</td>
<td>1.0769</td>
<td></td>
<td>H₅–C₅–C₆</td>
<td>121.92</td>
<td></td>
</tr>
<tr>
<td>C₆–H₆</td>
<td>1.0808</td>
<td></td>
<td>H₆–C₆–N₁</td>
<td>115.42</td>
<td></td>
</tr>
</tbody>
</table>

a Ref. 26. b Nuclear repulsion energy $E_{nuc} = 357.159959$ hartree.

Table 4.2 lists leading electronic configurations as well as EOM singles amplitudes obtained in an EOM-CCSD/aug-cc-pVDZ calculation. The total norm of singly excited amplitudes $R²₁$ is about 0.9 for all the excited states, which suggests that they do not carry a significant doubly excited character. This conclusion is also supported by small weight of triple excitations in the EOM-CCSD(2,3) wave function, as discussed below.
Table 4.2: Leading electronic configurations in the EOM-CCSD/aug-cc-pVDZ wave function. \( r \) is the weight of a configuration, \( R_1^2 \) is the norm of the EOM singles amplitudes.

<table>
<thead>
<tr>
<th>State</th>
<th>Configurations</th>
<th>( r )</th>
<th>( R_1^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X^1A' )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 1^1A''(\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(34 , \alpha'')^2 )</td>
<td>0.22</td>
<td>0.91</td>
</tr>
<tr>
<td>( 1^1A' (\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(35 , \alpha'') )</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>( 2^1A''(\text{Ry}) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(30 , \alpha'') )</td>
<td>0.68</td>
<td>0.93</td>
</tr>
<tr>
<td>( 2^1A' (\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(34 , \alpha'') )</td>
<td>0.34</td>
<td>0.89</td>
</tr>
<tr>
<td>( 2^1A' (\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(35 , \alpha'') )</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>( 1^3A' (\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(34 , \alpha'') )</td>
<td>0.30</td>
<td>0.95</td>
</tr>
<tr>
<td>( 1^3A'' (\pi\pi^+) )</td>
<td>( (26 , \alpha')^2(27 , \alpha')^2(28 , \alpha'')^2(29 , \alpha'')^2(35 , \alpha'') )</td>
<td>0.46</td>
<td></td>
</tr>
</tbody>
</table>

4.3.2 Effect of the one-electron basis set

The upper panel of Fig. 4.3 shows changes in the excitation energies of singlet and triplet states as the basis set on heavy atoms is expanded with diffuse and polarization functions. Changes in the oscillator strength are shown in the lower panel.

States within 5.5 eV from the ground state do not exhibit a significant dependence as the basis set is augmented: the excitation energy changes by less than 0.05 eV (Table 4.3). The change is appreciable for higher states (\( 1^1A' \) and upward). Among the studied states, the low-lying Rydberg \( 2^1A'' \) state is affected the most. At least one
Figure 4.3: Excitation energies (upper panel) and oscillator strengths (lower panel) of the lowest excited states of uracil calculated with EOM-CCSD and various one-electron basis sets.
Table 4.3: Vertical EOM-CCSD excitation energies of uracil (eV). Ground state CCSD energies are shown in hartree, oscillator strengths are given in parentheses. The singlet CCSD reference wave function was used for both singlet and triplet EOM calculations.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>Ground state</th>
<th>$1^1A''$</th>
<th>$1^1A'$</th>
<th>$2^1A''$</th>
<th>$3^1A''$</th>
<th>$2^1A'$</th>
<th>$1^3A'$</th>
<th>$1^3A''$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(nπ*)</td>
<td>(ππ*)</td>
<td>(Rydberg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(d)/</td>
<td>−413.814 317</td>
<td>5.28</td>
<td>5.84</td>
<td>6.71</td>
<td>6.97</td>
<td>3.92</td>
<td>4.99</td>
<td></td>
</tr>
<tr>
<td>6-31G</td>
<td></td>
<td>($12.35 \times 10^{-5}$)</td>
<td>(0.2203)</td>
<td>($8.532 \times 10^{-6}$)</td>
<td>(0.05573)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311G(d)/</td>
<td>−413.846 864</td>
<td>5.26</td>
<td>5.84</td>
<td>6.69</td>
<td>6.96</td>
<td>3.92</td>
<td>4.97</td>
<td></td>
</tr>
<tr>
<td>6-31G(d,p)</td>
<td></td>
<td>($12.27 \times 10^{-5}$)</td>
<td>(0.2219)</td>
<td>($8.269 \times 10^{-6}$)</td>
<td>(0.05418)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311+G(d)/</td>
<td>−413.864 729</td>
<td>5.22</td>
<td>5.70</td>
<td>6.16</td>
<td>6.59</td>
<td>6.80</td>
<td>3.89</td>
<td>4.95</td>
</tr>
<tr>
<td>6-31G(d,p)</td>
<td></td>
<td>($0.984 \times 10^{-5}$)</td>
<td>(0.2367)</td>
<td>($1.254 \times 10^{-6}$)</td>
<td>(0.05851)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311(2+G)(d)/</td>
<td>−413.866 067</td>
<td>5.22</td>
<td>5.70</td>
<td>6.01</td>
<td>6.58</td>
<td>6.79</td>
<td>3.89</td>
<td>4.95</td>
</tr>
<tr>
<td>6-31G(d,p)</td>
<td></td>
<td>($0.794 \times 10^{-5}$)</td>
<td>(0.2354)</td>
<td>($0.695 \times 10^{-6}$)</td>
<td>(0.05542)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311(2+G)(d,f)/</td>
<td>−414.008 904</td>
<td>5.21</td>
<td>5.71</td>
<td>6.12</td>
<td>6.58</td>
<td>6.82</td>
<td>3.89</td>
<td>4.94</td>
</tr>
<tr>
<td>6-31G(d,p)</td>
<td></td>
<td>($0.984 \times 10^{-5}$)</td>
<td>(0.2385)</td>
<td>($0.598 \times 10^{-6}$)</td>
<td>(0.05677)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-ANO-DZ</td>
<td>−413.953 957</td>
<td>5.25</td>
<td>5.60</td>
<td>6.05</td>
<td>6.76</td>
<td>3.86</td>
<td>4.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000)</td>
<td>(0.2110)</td>
<td>(0.002773)</td>
<td>(0.04666)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-cc-pVDZ</td>
<td>−413.822 857</td>
<td>5.22</td>
<td>5.58</td>
<td>6.00</td>
<td>6.57</td>
<td>6.72</td>
<td>3.87</td>
<td>4.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>($0.026 \times 10^{-5}$)</td>
<td>(0.2089)</td>
<td>($0.002660$)</td>
<td>($0.754 \times 10^{-6}$)</td>
<td>(0.04569)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-cc-pVTZ</td>
<td>−414.131 547</td>
<td>5.23</td>
<td>5.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
set of diffuse functions is required to describe it correctly, and with a second set of diffuse functions its excitation energy drops by 0.15 eV. Fleig et al.\textsuperscript{27} report similar effects of improving the basis set. They also show that the CC2 excitation energies for the four lowest valence singlet states of uracil change by 0.01 eV by extending the basis set from aug-cc-pVTZ to aug-cc-pVQZ, whereas the energy of the lowest Rydberg state changes by 0.05 eV. Thus, we consider our aug-cc-pVTZ energies of the valence states converged within approximately 0.01 eV with respect to the one-electron basis set.

The oscillator strength of the allowed $\pi\pi^*$ transitions changes very slightly as the basis set is expanded. In the case of the dark $n\pi^*$ transitions, the oscillator strength drops by an order of magnitude upon addition of a set of diffuse functions on heavy atoms.

We also note the excellent performance of the augmented ANO double-zeta and aug-cc-pVDZ bases, which produce excitation energies for the two lowest states within 0.02 eV from aug-cc-pVTZ (see Table 4.3), whereas the CC2 energies from Ref. 27 change by 0.08–0.11 eV between the aug-cc-pVDZ and aug-cc-pVTZ bases.

Overall, both polarization and diffuse functions are required to obtain the correct order and converged energy of the excited states. The basis set effects can account for as much as 0.25 eV for the $\pi\pi^*$ state, while the $n\pi^*$ state is less sensitive. We will use the aug-ANO-DZ and aug-cc-pVTZ basis sets in the best estimations of excitation energies.

As summarized in Table 4.3, EOM-CCSD calculations with the 6-311(2+)G(df)/6-31G(d,p), aug-ANO-DZ, and aug-cc-pVTZ bases give excitation energies of 5.21–5.25 eV for the lowest $n\pi^*$ state and 5.59–5.71 eV for the $\pi\pi^*$ transition.
4.3.3 Effects due to including higher excitations

Methods with explicit active-space triples, EOM-CC(2,3) and EOM-CCSDt, are used to estimate the effect of non-dynamical correlation. Results obtained with the 6-31G(d) basis set are presented in Table 4.4. Due to the high computational cost, only the energies of the first two transitions can be calculated with EOM-CCSDt(I). Including active-space triples affects the EOM-CCSD vertical excitation energies by no more than 0.15 eV for the lowest two singlet states.

The EOM-CC(2,3) expansion is similar to that of EOM-CCSDt(I), only different reference wave functions are used for the two methods: CCSD and CCSDt, respectively. Unlike EOM-CCSDT, EOM-CC(2,3) is not size-intensive\(^3\). This can be the cause of the opposite direction of the excitation energy shift produced by the two methods: EOM-CC(2,3) transition energies are blue-shifted, while EOM-CCSDt(I) energies are red-shifted with respect to EOM-CCSD. The resulting large discrepancy of about 0.2 eV (5.42 eV versus 5.20 eV and 5.96 eV versus 5.80 eV for the \(n\pi^*\) and \(\pi\pi^*\) transitions, respectively) is attributed to the same reason. However, both methods agree on the energy difference between the excited states. The transition energy between the first two excited singlet states (\(1^1A'' \rightarrow 1^1A'\)) is in the 0.54–0.61 eV range for all the explicit triples methods, slightly down from the 0.63 eV of EOM-CCSD. Thus, both models with active-space triples confirm that there is no significant doubly excited character in the lowest valence states of uracil, in agreement with the small basis CC3 results from Ref. 27.

Table 4.5 shows contributions from the singles, doubles, and triples parts of the excitation operator to the EOM-CC(2,3)/6-31G(d) wave function in the (10,10) active space. The excitation operator \(R\) is described by Eqns. (1.10) and (1.12). The norm of the doubly excited with respect to the reference part \(R_2\) is 0.16 and 0.13 for the \(n\pi^*\) and
Table 4.4: Effect of triple excitations on vertical excitation energies. Shift in excitation energies due to triple excitations is shown in parentheses. All energies are in eV.

<table>
<thead>
<tr>
<th>Method</th>
<th>$1^1A''$</th>
<th>$1^1A'$</th>
<th>$3^1A''$</th>
<th>$2^1A'$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(n\pi^*)$</td>
<td>$(\pi\pi^*)$</td>
<td>$(n\pi^*)$</td>
<td>$(\pi\pi^*)$</td>
</tr>
<tr>
<td><strong>6-31G(d) basis set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD</td>
<td>5.30</td>
<td>5.93</td>
<td>6.73</td>
<td>7.05</td>
</tr>
<tr>
<td>(10,8) active space</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CC(2,3)</td>
<td>5.42 (0.12)</td>
<td>5.96 (0.03)</td>
<td>6.93 (0.20)</td>
<td>6.92 (−0.13)</td>
</tr>
<tr>
<td>EOM-CCSDt(III)</td>
<td>5.30 (0.00)</td>
<td>5.91 (−0.02)</td>
<td>6.73 (0.00)</td>
<td>7.04 (−0.01)</td>
</tr>
<tr>
<td>EOM-CCSDt(II)</td>
<td>5.31 (0.01)</td>
<td>5.89 (−0.04)</td>
<td>6.76 (0.03)</td>
<td>7.02 (−0.03)</td>
</tr>
<tr>
<td>EOM-CCSDt(I)</td>
<td>5.20 (−0.10)</td>
<td>5.80 (−0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10,10) active space</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CC(2,3)</td>
<td>5.44 (0.14)</td>
<td>5.98 (0.05)</td>
<td>6.95 (0.22)</td>
<td>6.93 (−0.12)</td>
</tr>
<tr>
<td>EOM-CCSDt(III)</td>
<td>5.30 (0.00)</td>
<td>5.91 (−0.02)</td>
<td>6.73 (0.00)</td>
<td>7.04 (−0.01)</td>
</tr>
<tr>
<td>EOM-CCSDt(II)</td>
<td>5.31 (0.01)</td>
<td>5.89 (−0.04)</td>
<td>6.76 (0.03)</td>
<td>7.02 (−0.03)</td>
</tr>
<tr>
<td>CR-EOM-CCSD(T)</td>
<td>5.19 (−0.11)</td>
<td>5.65 (−0.28)</td>
<td></td>
<td>6.77 (−0.28)</td>
</tr>
<tr>
<td><strong>aug-cc-pVTZ basis set</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD</td>
<td>5.23</td>
<td>5.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR-EOM-CCSD(T)</td>
<td>5.00 (−0.23)</td>
<td>5.25 (−0.34)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\pi\pi^*$ states, respectively; the norm of $R_3$ does not exceed 0.01 for both. Thus, there is no evidence of a multi-configurational reference or a doubly excited character, for which explicit higher excitations are required. Therefore, the perturbative correction can be applied to recover the effect of all triple excitations.
Table 4.5: Norms of the amplitudes $R_{m,k}$ from Eq. (1.15) in the EOM-CC(2,3)/6-31G(d) wave function of uracil. The (10,10) active space is used. Total norm is not unity in virtue of biorthogonal properties of EOM-CC.

<table>
<thead>
<tr>
<th>State</th>
<th>$R_{0}^2$</th>
<th>$R_{1}^2$</th>
<th>$R_{2}^2$</th>
<th>$R_{3}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{1}A'$</td>
<td>0.99</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>$1^{1}A''$</td>
<td>0.00</td>
<td>0.83</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>$1^{1}A'$</td>
<td>0.00</td>
<td>0.86</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>$2^{1}A''$</td>
<td>0.00</td>
<td>0.86</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>$2^{1}A'$</td>
<td>0.00</td>
<td>0.78</td>
<td>0.21</td>
<td>0.02</td>
</tr>
</tbody>
</table>

In the EOM-CCSDt(III) → (II) → (I) series of calculations (Table 4.4), the active-space constraint for the excitations given by Eqns. (1.14)–(1.19) gradually relaxes, giving rise to a partial account for dynamical correlation. The increasing magnitude of the correction in the series indicates a considerable effect of dynamical correlation. The reduced subset of triples is not capable of fully accounting for dynamical correlation, and we employ CR-EOM-CCSD(T) for that purpose, which indeed produces a larger correction of 0.11–0.28 eV. A comparable correction (0.2 eV) for the lowest $\pi\pi^*$ state was obtained at the CC3 level in a similar quality basis set\textsuperscript{27}. Overall, the vertical excitation energies calculated with CR-EOM-CCSD(T)/6-31G(d) for the $n\pi^*$ and $\pi\pi^*$ transitions are 5.19 eV and 5.65 eV, respectively.

Since dynamical correlation converges slowly with the basis set, it is important to evaluate the effect of triple excitations in a large basis set. In the aug-cc-pVTZ basis, the CR-EOM-CCSD(T) excitation energies of the two lowest states are below the respective EOM-CCSD values by 0.23–0.34 eV, which is appreciably larger than the 6-31G(d) values.
The observed tendency of triples to lower EOM-CCSD excitation energies is consistent with recent studies\textsuperscript{27,28}.

### 4.3.4 MRCI results

The results of MRCI calculations are presented in Table 4.6. Due to its high computational requirements, MRCI\textsubscript{2} was only used for the first two singlet excited states. As reported previously\textsuperscript{15}, the MRCI\textsubscript{2} expansion, which includes more correlation than MRCI\textsubscript{1}, does not change the excitation energies dramatically. That indicates that the $\sigma$–$\pi$ correlation plays a more important role than double excitations from the active space.

The MRCI expansions used in these calculations cannot describe Rydberg states since there are no Rydberg orbitals included in the active space. So even when the basis set has diffuse functions, the Rydberg states will be too high in energy, and that is why the second $A''$ state in the MRCI results is always $n\pi^*$, even though in the coupled cluster calculations Rydberg states appear below the second $n\pi^*$.

With the 6-31G($d,p$) basis set, the MRCI\textsubscript{2} excitations are 5.16 eV and 5.89 eV for the first two singlet states compared to 5.30 eV and 5.93 eV at the EOM-CCSD/6-31G($d$) level. Thus the MRCI\textsubscript{2} energies are blue-shifted by 0.14 eV and 0.04 eV for the two states compared to EOM-CCSD. With the 6-311+G($d$)/6-31G($d,p$) basis set, the MRCI\textsubscript{2} energies become 4.87 eV and 5.70 eV. The corresponding excitation energies using EOM-CCSD and the same basis set [6-311+G($d$)/6-31G($d,p$)] are 5.22 eV and 5.70 eV. The two methods agree well on the $\pi\pi^*$ energy, but the disagreement for $n\pi^*$ has now increased to 0.35 eV. This is mainly because the MRCI\textsubscript{2} energy of the lowest singlet shows a surprisingly large effect on the basis set. With the aug-ANO-DZ basis set the energy of $\pi\pi^*$ is 5.65 eV, which is very similar to the EOM-CCSD value of
Table 4.6: MRCI excitation energies (eV) of the four lowest singlet excited states of uracil. Oscillator strengths are given in parentheses.

<table>
<thead>
<tr>
<th>Basis set</th>
<th>CSFs</th>
<th>$1^1A''$</th>
<th>$1^1A'$</th>
<th>$3^1A''$</th>
<th>$2^1A'$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>($n\pi^*$)</td>
<td>($\pi\pi^*$)</td>
<td>($n\pi^*$)</td>
<td>($\pi\pi^*$)</td>
</tr>
<tr>
<td><strong>MRCI1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G($d,p$)</td>
<td>16,627,710</td>
<td>5.09 ($14.90 \times 10^{-5}$)</td>
<td>5.79 (0.186)</td>
<td>6.62 (0.000)</td>
<td>6.57 (0.0478)</td>
</tr>
<tr>
<td>6-311+G($d$)/6-31G($d,p$)</td>
<td>28,159,230</td>
<td>5.02 ($3.200 \times 10^{-5}$)</td>
<td>5.65 (0.221)</td>
<td>6.87 ($5.100 \times 10^{-5}$)</td>
<td>6.44 (0.0357)</td>
</tr>
<tr>
<td>aug-ANO-DZ</td>
<td>32,483,550</td>
<td>5.01 ($0.500 \times 10^{-5}$)</td>
<td>5.53 (0.180)</td>
<td>6.49 ($1.600 \times 10^{-5}$)</td>
<td>6.35 (0.0384)</td>
</tr>
<tr>
<td><strong>MRCI2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G($d,p$)</td>
<td>112,036,212</td>
<td>5.16</td>
<td>5.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311+G($d$)/6-31G($d,p$)</td>
<td>254,019,060</td>
<td>4.87</td>
<td>5.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-ANO-DZ</td>
<td>320,019,060</td>
<td>5.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MRCI2 + Davidson correction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-31G($d,p$)</td>
<td></td>
<td>5.09</td>
<td>5.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-311+G($d$)/6-31G($d,p$)</td>
<td></td>
<td>4.76</td>
<td>5.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aug-ANO-DZ</td>
<td></td>
<td></td>
<td>5.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.60 eV. Unfortunately, we could not obtain the energy of the \( n\pi^* \) state at that level due to technical difficulties, so we could not explore further the dependence of the \( n\pi^* \) state on the basis set. The basis set dependence of the \( n\pi^* \) state at the CC level and at the MRCI1 level is much smaller. We do not understand this MRCI2 dependence well at present, but it may be an effect of the smaller active space used in MRCI2.

The Davidson correction lowers the excitation energy of the \( \pi\pi^* \) state by 0.29–0.49 eV, indicating the importance of higher than double excitations in the wave function. The magnitude of the effect is consistent with the EOM-CCSD and CR-EOM-CCSD(T) results (Table 4.4). The \( n\pi^* \) state is not affected as much by the Davidson correction, which lowers the energy by about 0.1 eV.

### 4.3.5 Solvent effects

TDDFT calculations of the isolated uracil molecule with the B3LYP functional and the 6-31G(d) basis set give excitation energies of 4.73 eV and 5.37 eV for the \( n\pi^* \) and \( \pi\pi^* \) states, respectively (Table 4.7). Account for effects of surrounding water molecules via QM/MM increases the energies by 0.65 eV and 0.02 eV, respectively.

QM/MM calculations that employ EOM-CCSDt(II) with the 6-31G(d) basis and the (10,10) active space yield first two excitation energies of 5.75 eV and 5.96 eV, which are 0.44 eV and 0.07 eV above the respective gas-phase excitation energies.

Solvent effects are in a qualitative agreement with previous TDDFT/PBE0 and TDDFT/B3LYP calculations, which predict a larger positive shift for the \( n\pi^* \) state and a much smaller shift for the \( \pi\pi^* \) state\(^{29–31}\). In our calculations, the excitation energy of the \( n\pi^* \) state increases by 0.5 eV, and the excitation energy of the \( \pi\pi^* \) state remains almost the same. Taking into account previous calculations, we estimate the effect on the \( \pi\pi^* \) state as \( \pm 0.1 \) eV.
Table 4.7: Vertical excitation energies (eV) of uracil in the gas phase and water solution calculated with QM/MM. Energy shifts due to the solvent are shown in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>Environment</th>
<th>$1^1A''$ $n\pi^*$</th>
<th>$1^1A'$ $\pi\pi^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDDFT/B3LYP/6-31G(d)</td>
<td>Gas phase</td>
<td>4.73</td>
<td>5.37</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>5.38 (0.65)</td>
<td>5.39 (0.02)</td>
</tr>
<tr>
<td>EOM-CCSDt(II)/6-31G(d)$^a$</td>
<td>Gas phase</td>
<td>5.31</td>
<td>5.89</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>5.75 (0.44)</td>
<td>5.96 (0.07)</td>
</tr>
<tr>
<td>TD-DFT/PBE0/6-311+G(2d,2p)$^b$</td>
<td>Gas phase</td>
<td>4.80</td>
<td>5.26</td>
</tr>
<tr>
<td></td>
<td>Water</td>
<td>5.27 (0.47)</td>
<td>5.14 (−0.12)</td>
</tr>
</tbody>
</table>

$^a$ (10,10) active space. $^b$ Ref. 29.

Our calculations do not account for solvent-specific interactions and possible tautomerization. Previous cluster studies of DNA bases with several water molecules$^{32–36}$ do not suggest a large effect on the excitation energies due to solvent-specific interactions. For example, Yoshikawa and Matsika$^{36}$, who investigated excited states of uracil–water isomers, report shifts between $−0.03$ eV and $+0.2$ eV for the $n\pi^*$ state and between $−0.1$ eV and $+0.1$ eV for the $\pi\pi^*$ state.

There is no explicit QM treatment of the hydrogen bond interaction between uracil and surrounding water molecules in our calculations. Since excitation is localized on the uracil moiety, we expect the average solvent electric field to have a larger effect on the excitation energies than hydrogen bonding. That is captured by the QM/MM calculations through the classical description of the solvent. To quantify the errors due to the classical as opposed to the fully QM description of the first solvation shell, we perform an additional TDDFT QM/MM calculation with one of the waters coordinating with an oxygen atom included into the QM part. Resulting excitation energies of 5.38 eV and
5.42 eV for the \( n\pi^* \) and \( \pi\pi^* \) states show that the errors due to the classical description are less than 0.01 eV and 0.03 eV for the two states, respectively.

### 4.3.6 Best theoretical estimate of vertical excitation energies in uracil

EOM-CCSD with large augmented basis sets gives vertical excitation energies of about 5.25 eV and 5.60 eV for the \( n\pi^* \) and \( \pi\pi^* \) states, respectively (Table 4.8). Including triple excitations non-iteratively via CR-EOM-CCSD(T) lowers both energies by 0.2–0.3 eV to 5.0 eV and 5.3 eV. That is our best estimate of the excitation energies. The \( \pi\pi^* \) transition energy computed with MRCI with the Davidson correction is 5.3 eV, which is in an excellent agreement with the EOM value. We estimate the error bars for the theoretical values to be ±0.05 eV.

TDDFT/PBE0 agrees with the EOM methods on the \( \pi\pi^* \) energy, but underestimates the energy of the \( n\pi^* \) transition by 0.2 eV. TDDFT/B3LYP underestimates the excitation energies of both states even more. The CASPT2/ANO values\(^{37}\) are below our best estimates by 0.5 eV (\( n\pi^* \)) and 0.3 eV (\( \pi\pi^* \)). More recent CASPT2 calculations\(^{38}\) employing a larger active space and a smaller basis are closer to the EOM-CC estimates, however, in view of strong basis set dependence, this agreement is accidental. Empirically-corrected CASPT2/TZVP values\(^{28}\) for the two singlet states are below our estimates by 0.1 eV and 0.02 eV. However, a relatively modest basis set was used, and it is not clear how well empirically corrected CASPT2 would perform in a larger basis.

The CC2/TZVP excitation energies\(^{28}\) differ from our estimates by \(-0.09\) eV and \(+0.22\) eV for the \( n\pi^* \) and \( \pi\pi^* \) states, respectively. Consistently with our findings, larger basis CC2 calculations\(^{27}\) yield considerably different values, which are \(-0.2\) eV and
Table 4.8: Best estimates of the vertical excitation energies (eV) of the lowest singlet and triplet states compared with previously published results.

<table>
<thead>
<tr>
<th>Method</th>
<th>$1^1A''$</th>
<th>$1^1A'$</th>
<th>$1^3A'$</th>
<th>$1^3A''$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOM-CCSD/aug-cc-pVTZ</td>
<td>5.23</td>
<td>5.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR-EOM-CCSD(T)/aug-cc-pVTZ</td>
<td>5.00</td>
<td>5.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD/aug-ANO-DZ</td>
<td>5.25</td>
<td>5.60</td>
<td>3.86$^a$</td>
<td>4.98$^a$</td>
</tr>
<tr>
<td>CR-EOM-CCSD(T)/aug-ANO-DZ</td>
<td>5.06</td>
<td>5.30</td>
<td>5.17</td>
<td></td>
</tr>
<tr>
<td>MRCISD+Q(12,9)/aug-ANO-DZ</td>
<td></td>
<td></td>
<td></td>
<td>5.32</td>
</tr>
<tr>
<td>CASPT2(2,10)/Roos ANO$^b$</td>
<td>4.54</td>
<td>5.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASPT2(14,10)/6-31G(d,p)$^c$</td>
<td>4.93</td>
<td>5.18</td>
<td>3.80</td>
<td>4.71</td>
</tr>
<tr>
<td>CASPT2/TZVP$^d$</td>
<td>4.90</td>
<td>5.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDDFT/PBE0/6-311+G(2d,2p)$^e$</td>
<td>4.80</td>
<td>5.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDDFT/B3LYP/aug-cc-pVTZ$^f$</td>
<td>4.64</td>
<td>5.11</td>
<td>3.38</td>
<td>4.26</td>
</tr>
<tr>
<td>TDDFT/B3LYP/aug-cc-pVTZ$^g$</td>
<td>4.62</td>
<td>5.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDDFT/B3LYP/6-311++G(d,p)$^h$</td>
<td>4.67</td>
<td>5.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIS(2$'$)/cc-pVDZ$^i$</td>
<td>5.11</td>
<td>5.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI-CC2/aug-cc-pVQZ$^j$</td>
<td>4.80</td>
<td>5.35</td>
<td>3.95</td>
<td>4.61</td>
</tr>
<tr>
<td>CC2/TZVP$^d$</td>
<td>4.91</td>
<td>5.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOM-CCSD/TZVP$^d$</td>
<td>5.11</td>
<td>5.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ The singlet CCSD reference wave function was used. $^b$ Ref. 37. $^c$ Ref. 38. $^d$ Ref. 28. $^e$ Ref. 29. $^f$ Ref. 39. $^g$ Ref. 30. $^h$ Ref. 40. $^i$ Ref. 41. $^j$ Ref. 27.

+0.05 eV from our estimates. These differences are well within the CC2 error bars (see Sec. 3.2).

The present data, as well as previous studies,$^{27,28}$ demonstrate that a large basis set and an accurate account of dynamical correlation are both essential for converged results.
Fig. 4.4 shows a UV absorption spectrum of uracil vapor obtained at 228°C and the EOM-CCSD/aug-ANO-DZ and CR-EOM-CCSD(T)/aug-ANO-DZ vertical excitation energies depicted as a stick spectrum using the EOM-CCSD/aug-ANO-DZ intensities. The experimental spectrum shows a broad feature peaking at 242 nm (5.08 eV), which is 0.17 eV below the CR-EOM-CCSD(T)/aug-cc-pVTZ vertical excitation energy for the bright \( \pi\pi^* \) state. As elaborated in the benchmark study of Schreiber et al., there are numerous reasons complicating the interpretation of the observed absorption maxima. We attribute the difference to strong vibronic coupling of the \( n\pi^* \) and \( \pi\pi^* \) states. Intensity borrowing by the lower-lying dark state can affect the overall shape of the absorption, shifting it toward lower energies. Strong vibronic interaction is consistent with the broad featureless shape of the spectrum, as well as with experimental observations of Kong and coworkers, however, more detailed investigations are necessary for a definite conclusion.
4.4 Chapter 4 conclusions

A series of calculations of energies and oscillator strengths of transitions to electronically excited states of uracil show that using high-quality basis sets (i.e., aug-cc-pVTZ or aug-ANO-DZ) with diffuse functions is essential for converged results.

Active-space EOM-CC calculations with explicit triple excitations demonstrate that the excited states do not exhibit a doubly excited character, and therefore EOM-CCSD provides a good zero-order wave function for including remaining dynamical correlation via perturbative treatment of triple excitations. The latter affects excitation energies by as much as 0.3 eV. Thus, the EOM-CCSD errors for the excitation energies of uracil do not exceed the conventional estimates of the EOM-EE-CCSD error bars.

Our best estimate of excitation energies calculated with CR-EOM-CCSD(T)/aug-cc-pVTZ places the $n\pi^*$ singlet excited state at $E(1^1A''')=5.00$ eV, and the $\pi\pi^*$ singlet state at $E(1^1A')=5.25$ eV above the ground state. An excellent agreement between high-level MRCI calculations, which place the $\pi\pi^*$ state at 5.32 eV, and EOM-CC further supports our conclusions. A solvent (water) affects the excitation energies of the two states by +0.5 eV and less than 0.1 eV, respectively.

We conclude that the maximum of the broad UV spectrum of uracil does not correspond to the position of the vertical $\pi\pi^*$ excitation, possibly due to strong vibronic coupling with the lower-lying dark $n\pi^*$ state. Our results indicate that previously reported CASPT2 and TDDFT calculations underestimate the excitation energies.
4.5 Chapter 4 references


Chapter 5: Direct location of the minimum point on intersection seams of potential energy surfaces with equation-of-motion coupled-cluster methods

5.1 Chapter 5 introduction

The dynamics of chemical reactions is often interpreted in terms of nuclear motions on a single potential energy surface (PES), which is based on the (non-relativistic) Born-Oppenheimer (BO), or adiabatic, approximation\textsuperscript{1}. The BO approximation separates slow and fast degrees of freedom: the nuclei move in the mean field created by the electrons, which adjust their positions instantaneously. This is valid for many chemical reactions, however there is a plethora of processes for which the single-surface approximation breaks down\textsuperscript{2}. Examples of such nonadiabatic processes, which involve transitions between different electronic states, include organic photochemistry\textsuperscript{3} with a
Figure 5.1: The potential energy surface slices for the ground and the first excited states of a substituted ethylene. There is an energy barrier that prevents the molecule from rotating freely around the double bond. The top of the barrier on the ground state PES corresponds to the minimum on the excited state PES.

significant share of biologically important chemical reactions\textsuperscript{4}, as well as spin-forbidden processes\textsuperscript{5}.

Consider, for instance, the isomerization of retinal, an important step in the chemistry of vision\textsuperscript{6}. A prototype reaction\textsuperscript{7}, the \textit{cis-trans} photoisomerization of substituted ethylene, is sketched in Fig. 5.1. Upon electronic excitation, the system slides downhill towards the minimum on the excited state PES, which corresponds to a local maximum on the lower PES. At that point, the two surfaces approach each other, and the probability of a transition back to the ground state increases. Once on the ground state PES, the system relaxes to either the \textit{cis}- or \textit{trans}-product.

A set of points of confluence of two adiabatic PESs forming a continuous hyperline is called an intersection seam\textsuperscript{8}. A nonadiabatic transition from one electronic state to another is more likely to happen near the crossing and is enabled by electronic couplings
between the states. The nature of the coupling can be different: electronic states of the same multiplicity and symmetry can be coupled through the derivative coupling, while spin-orbit interaction facilitates the transition between states of different multiplicities. While the nonadiabatic transition can occur at any nuclear configuration, its probability is higher if the energy gap between the states is small.

A full description of a nonadiabatic process requires a calculation of nuclear dynamics that brings the system from the Franck-Condon region towards the point where the two states are nearly degenerate, and the probability of the nonadiabatic transition is larger (see, for example, Ref. 9, 10 and references therein). To describe the transition, both electronic and nuclear degrees of freedom should be considered.

Alternatively, one can gain considerable mechanistic insight and even quantitative information about the probability of nonadiabatic transitions from the geometry and energy of the minimum energy point on the intersection seam (minimum energy crossing point, MECP, MSX)\textsuperscript{11–19}. The MECP can be thought of as a characteristic point of the process, just like the transition state in a thermal chemical reaction. The energy of the MECP can help assess the possibility of the process, and with the use of the nonadiabatic transition state theory\textsuperscript{20}, the rate of the nonadiabatic process can be estimated. The geometry of MECP suggests which modes are most efficient in facilitating the transition\textsuperscript{21}.

Several methods of finding the MECP have been proposed over the past years\textsuperscript{11–13,15–19}. The crossing optimization procedure is very similar to that of finding the minimum on a PES, with the difference that constraints must be enforced to ensure the degeneracy of the states of interest. The algorithm of locating the MECP on a conical intersection seam is the same as for crossings of different symmetry or spin potential surfaces, only the constraints (and the dimensionality of the seam) are different.
Many methods employ the Lagrange multiplier technique to constrain the optimization. Some approaches require the calculation of a nonadiabatic coupling vector as well, which is rather demanding and have been implemented only for a handful of wavefunctions. For example, Ragazos et al.\textsuperscript{12} and Manaa et al.\textsuperscript{15} introduced analytical formalisms for the calculation of the nonadiabatic coupling vector at the multi-reference self-consistent field (MCSCF) and the multi-reference configuration interaction (MRCI) levels of theory and used it for seam minimization. Levine et al.\textsuperscript{19} avoid the laborious evaluation of the coupling vector by using a tuning parameter. This implementation works with the time-dependent density functional theory (TD-DFT) for excited states.

This chapter reports an implementation of the projected gradient method\textsuperscript{16} within the equation-of-motion coupled-cluster (EOM-CC) family of methods. The algorithm does not require the use of the Lagrange multiplier method and enables the constraints in the minimization procedure by projecting the energy gradient. The method was implemented within the Q-Chem quantum chemistry package\textsuperscript{22}.

5.2 Theory

5.2.1 Adiabatic approximation

The full non-relativistic Hamiltonian of a molecular system $\mathbf{H}$ depends explicitly on both nuclear and electronic coordinates, $\mathbf{R}$ and $\mathbf{r}$, respectively:

$$\mathbf{H}(\mathbf{r}, \mathbf{R}) = T_n(\mathbf{R}) + T_e(\mathbf{r}) + V_{ee}(\mathbf{r}) + V_{ne}(\mathbf{r}, \mathbf{R}) + V_{nn}(\mathbf{R})$$

$$= T_n(\mathbf{R}) + H_e(\mathbf{r}; \mathbf{R}),$$
where $T_e$ and $T_n$ are electronic and nuclear kinetic energy operators, respectively, $V_{ee}$, $V_{ne}$, and $V_{nn}$ are potential energy operators for electron-electron, nuclei-electron, and nuclei-nuclei interactions. The electronic Hamiltonian $H_e(r;R)$ is defined as the full Hamiltonian minus the nuclear kinetic energy operator and thus depends explicitly on the electronic coordinates and only parametrically on the nuclear coordinates.

Eigenfunctions of the full Hamiltonian satisfy the Schrödinger equation:

$$H(r, R)\Psi_i(r, R) = \varepsilon_i \Psi_i(r, R).$$  \hfill (5.2)

The solutions of the so-called electronic Schrödinger equation are defined as follows:

$$H_e(r; R)\phi_i(r; R) = E_i \phi_i(r; R),$$  \hfill (5.3)

and may be used as a basis to represent the full wavefunction $\Psi(r, R)$:

$$\Psi(r, R) = \sum_j \phi_j(r; R)\chi_j(R),$$  \hfill (5.4)

where $\chi_j(R)$ are expansion coefficients called nuclear wave functions.

Substituting representation (5.4) into the full Schrödinger equation (5.2) and multiplying by $\phi_k$ from the left, we arrive at:

$$\sum_{k'} \langle \phi_k | H_e + T_n | \phi_{k'} \rangle_r \chi_{k'} = \varepsilon \sum_{k'} \langle \phi_k | \phi_{k'} \rangle_r \chi_{k'},$$  \hfill (5.5)
where the subscript $r$ denotes integration over the electronic coordinates. After expanding $T_n$ and using 5.3:

$$(\varepsilon - E_k - T_n - T''_{kk}) \chi_k = \sum_{k' \neq k} (T'_{kk'} + T''_{kk'}) \chi_{k'},$$

(5.6)

where:

$$T_n = -\sum_A \frac{1}{2M_A} \frac{\partial^2}{\partial R_A^2},$$

(5.7)

$$T'_{ij} = -\sum_A \frac{1}{M_A} \left\langle \phi_i \left| \frac{\partial}{\partial R_A} \right| \phi_j \right\rangle \frac{\partial}{\partial R_A},$$

(5.8)

$$T''_{ij} = -\sum_A \frac{1}{2M_A} \left\langle \phi_i \left| \frac{\partial^2}{\partial R_A^2} \right| \phi_j \right\rangle.$$

(5.9)

A solution to the coupled set of equations (5.6) yields $\chi_j(R)$, and therefore the solution to the full problem (5.2). If the off-diagonal terms on the right hand side of (5.6) vanish, the overall wavefunction becomes just a single product of the electronic and nuclear parts, and the nuclear motions on different potential surfaces become uncoupled. This is the essence of the Born-Oppenheimer approximation: complete adiabatic separation of the nuclear and electronic motions (see, for example, Ref. 1).

The approximation therefore requires that the coupling terms in (5.6) are small and can be neglected. Thus, it is important to understand when they may become significant. Using perturbation theory, one can rewrite the electronic part of the nonadiabatic coupling terms in (5.8) as:

$$\left\langle \phi_i \left| \frac{\partial}{\partial \mathbf{R}} \right| \phi_j \right\rangle = \frac{\left\langle \phi_i \left| \frac{\partial H_e}{\partial \mathbf{R}} \right| \phi_j \right\rangle}{E_j - E_i},$$

(5.10)
which suggests that the coupling between two electronic states becomes large when they are nearly degenerate. In such situations, the adiabatic approximation breaks down, and transitions between different electronic states become possible.

5.2.2 Crossing location and minimization

Although the set of equations (5.6) is infinite, the energy denominator in (5.10) suggests that the coupling between only two electronic states needs to be taken into account in the regions of proximity of their PESs, and the couplings with other states can be neglected.

The description of conical intersections and the algorithms for locating MECPs are based on a first order perturbative treatment of the electronic Schrödinger equation in the vicinity of the intersection. Recently, second-order treatment that accounts for the curvature of the seam, in addition to the first-order linear terms, has been reported.

At the intersection point \( \mathbf{R} \), the electronic Hamiltonian has the following matrix form in the so-called “crude adiabatic basis” \( \{ \phi_i(\mathbf{r}; \mathbf{R}_0) \} \), eigenstates of the electronic Hamiltonian at a nearby point \( \mathbf{R}_0 \):

\[
\mathbf{H}_e(\mathbf{R}) = \begin{pmatrix}
H_{11} & H_{12} \\
H_{12} & H_{22}
\end{pmatrix},
\]

(5.11)

where \( H_{ij}(\mathbf{R}) = \langle \phi_i(\mathbf{r}; \mathbf{R}_0) | \mathbf{H}_e(\mathbf{R}) | \phi_j(\mathbf{r}; \mathbf{R}_0) \rangle_r \).

After finding the eigenvalues of the Hamiltonian (5.11), one can readily write the conditions on the matrix elements that ensure that the states are degenerate:

\[
\begin{cases}
H_{11} - H_{22} = 0, \\
H_{12} = 0.
\end{cases}
\]

(5.12)
The above result is the essence of the Neumann–Wigner theorem, or the noncrossing rule\textsuperscript{25}, which states that two conditions have to be imposed on the Hamiltonian in order for the states to be degenerate and, therefore, limits the dimensionality of the intersection seam to $N - 2$, where $N$ is the number of internal degrees of freedom. If the coupling term is zero due to the symmetry restrictions (e.g., between the states of different spatial symmetry or different multiplicity), the dimensionality of the seam is $N - 1$.

By applying degenerate perturbation theory (see, for example, Ref. 23) to the eigenproblem with Hamiltonian (5.11), conditions (5.12) become:

\[
\begin{cases}
\frac{\partial}{\partial R} \left[ E_1(R) - E_2(R) \right] \bigg|_{R=R_0} \delta R = 0, \\
\left\langle \phi_1(r; R) \left| \frac{\partial}{\partial R} \phi_2(r; R) \right\rangle \right|_r \bigg|_{R=R_0} \delta R = 0.
\end{cases}
\]

By defining the gradient difference vector:

\[
g(R_0) = \frac{\partial}{\partial R} E_1(R) \bigg|_{R=R_0} - \frac{\partial}{\partial R} E_2(R) \bigg|_{R=R_0},
\]

and the derivative coupling vector:

\[
h(R_0) = \left\langle \phi_1(r; R) \left| \frac{\partial}{\partial R} \phi_2(r; R) \right\rangle \right|_r \bigg|_{R=R_0},
\]

we arrive at:

\[
\begin{cases}
g(R_0) \cdot \delta R = 0, \\
h(R_0) \cdot \delta R = 0.
\end{cases}
\]

Non-zero $g$ and $h$ vectors define a so-called $g$-$h$, or branching, plane\textsuperscript{26,27}. The degeneracy of the states is preserved in the subspace complement to the $g$-$h$ plane and is lifted
only along the plane. Thus, MECP can be found by following the gradient on one of the PES under constraint (5.16).

In this work, we implemented a reduced version of the projected gradient method\textsuperscript{16} for locating the minimum energy point on an intersection. The method requires two vectors, the gradient difference $g$ as defined in (5.14), the derivative coupling $h$ as in (5.15), and the projection of the gradient on the higher PES onto the subspace orthogonal to the branching plane:

$$G_2 = P \frac{\partial}{\partial R} E_2,$$

(5.17)

where $P$ is the projector that performs the required operation. The gradient difference vector is multiplied by the energy difference in order to yield zero on the seam:

$$G_1 = \frac{E_2 - E_1}{\sqrt{|g|^2}} g.$$

(5.18)

The total gradient used in the optimization procedure is defined as:

$$G = G_1 + G_2.$$  

(5.19)

Following $G_1$, one arrives to the seam, while $G_2$ is a gradient along the seam towards MECP.

The advantage of this technique is that it can be readily used with any existing optimization algorithm. The disadvantage is that it requires both gradients and nonadiabatic couplings.

Analytical calculations of the $h$ vector have been implemented for the MCSCF and MRCI methods\textsuperscript{12,15}. For EOM-CC, however, only the mathematical procedure has been developed\textsuperscript{28}, and no implementation has been reported so far. In this work we consider the situations when the nonadiabatic derivative coupling vector is zero by symmetry, e.g.
the spatial parts of the wavefunctions of the states carry different irreducible representations or the states are of different multiplicities. In this case, only analytic gradients are required\textsuperscript{29}.

5.3 Computational details

All calculations were performed with the Q-Chem\textsuperscript{22} package. Various versions of the EOM-CC method were employed for excited states calculations. The 6-31G basis set\textsuperscript{30} was used for \(N_3^+\) and NO\(_2\), and 6-31G* basis\textsuperscript{31} — for the \textit{para}-benzylene diradical. To validate the code, MECPs were located by generating dense numerical grids for \(N_3^+\) and NO\(_2\). Then the algorithm was tested by setting the calculations with different starting geometries and employing different tolerance criteria.

5.4 Benchmarks and examples

5.4.1 Cyclic trinitrogen cation

The cyclic trinitrogen cation is a closed-shell system with a manifold of low-lying near-degenerate excited states\textsuperscript{32,33}.

The equilibrium ground state geometry of \(N_3^+\) is an equilateral triangle \((D_{3h})\), whereas low-lying excited states undergo Jahn-Teller distortions to \(C_{2v}\). The \(C_{2v}\) point group symmetry will be used to label the excited states.

Both HOMO and LUMO in the ground state orbital picture of the cation are doubly degenerate. Therefore, four near-degenerate electronic states are derived upon the excitations from the highest occupied lone pair orbitals to the lowest unoccupied \(\pi^*\) orbitals. Among those, the \(2^1A_2\) and \(1^1B_1\) adiabatic states have PESs that form two intersection
seams. One of the seams lies along the nuclear space coordinate that preserves the $D_{3h}$ point group symmetry (this is the Jahn-Teller seam), and the other one is found in the subspace where the molecule is distorted from the equilateral triangular geometry to $C_{2v}$. There are 3 symmetry-identical MECP of the $C_{2v}$ type.

This interesting feature — 4 very closely located conical intersections — was characterized by Dillon and Yarkony$^{33}$ and makes the intersection to appear glancing rather than conical$^{32}$.

As in the recent study$^{32}$, we employed EOMEE-CCSD to describe the excited states of $N_3^+$. As long as the molecule preserves the $C_{2v}$ symmetry, the wavefunctions of the electronic states of interest belong to different irreps, $A_2$ and $B_1$, and therefore the derivative coupling matrix element vanishes. The intersection seam between the states is two-dimensional.

To validate the algorithm, the $2^1A_2/1^1B_1$ crossing seams were first scanned in order to find the minimum energy points. The locations of these points, as well as the points of triple degeneracy, are listed in Table 5.1.

To assess the performance of the algorithm, we attempted various geometries as the starting points for the optimization procedure. Usually the algorithm locates the minimum energy points on the $2^1A_2/1^1B_1$ seams in less than 20 iterations (Table 5.2). Just like the local minimum on a global PES found with an optimization algorithm depends on the starting point, the seam located with the projected gradient method depends on the choice of the initial molecular geometry.

Another interesting feature of the manifold of the electronic states in $N_3^+$ is additional accidental degeneracies. The $1^1A_2$ and $2^1A_2$ are degenerate at a point in the internal coordinate space which happens to correspond to the molecule having the $D_{3h}$ point group symmetry. Since $2^1A_2$ and $1^1B_1$ are rigorously degenerate in $D_{3h}$, the point
Table 5.1: Geometries and total energies of the intersection seam minima and points of triple degeneracy in $N_3^+$ calculated by EOM-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees, all energies are given in a.u.

<table>
<thead>
<tr>
<th>$R$(NN)</th>
<th>$\alpha$(NNN)</th>
<th>$E(1^1A_2)$</th>
<th>$E(2^1A_2)$</th>
<th>$E(1^1B_1)$</th>
<th>$E(2^1B_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{3h}$ seam minimum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4556</td>
<td>60.00</td>
<td>-162.836070</td>
<td>-162.822635</td>
<td>-162.822635</td>
<td>-162.816033</td>
</tr>
<tr>
<td>$C_{2v}$ seam minimum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4476</td>
<td>60.78</td>
<td>-162.837652</td>
<td>-162.821903</td>
<td>-162.821903</td>
<td>-162.815864</td>
</tr>
<tr>
<td>Points of triple degeneracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6671</td>
<td>60.00</td>
<td>-162.795662</td>
<td>-162.795661</td>
<td>-162.799368</td>
<td>-162.795663</td>
</tr>
<tr>
<td>1.6041</td>
<td>60.00</td>
<td>-162.812016</td>
<td>-162.807857</td>
<td>-162.807860</td>
<td>-162.807858</td>
</tr>
</tbody>
</table>

becomes a point of a triple $1^1A_2/2^1A_2/1^1B_1$ intersection. The $1^1B_1$ and $2^1B_1$ electronic states have a point of degeneracy as well, which is therefore the point of a triple $2^1A_2/1^1B_1/2^1B_1$ intersection.

The $1^1A_2/2^1A_2/1^1B_1$ and $2^1A_2/1^1B_1/2^1B_1$ triple intersections were located at $D_{3h}$ geometries with the N–N distance $R$(NN) = 1.6671 Å and $R$(NN) = 1.6041 Å, respectively. Scans of the PESs around the points of intersections are shown in Fig. 5.2.

Note that the triple intersections are found at the $D_{3h}$ symmetry and in fact are crossings of PESs belonging to different irreducible representations, even though it is not immediately clear from the $C_{2v}$ notation. Therefore, the problem with complex energy values$^{34}$ does not arise.
Table 5.2: Results of energy minimization of the $^2A_2/^1B_1$ intersection seam in $N_3^+$ calculated by EOM-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees.

<table>
<thead>
<tr>
<th>Starting geometry</th>
<th>Iterations</th>
<th>Final geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R(\text{NN})$</td>
<td>$\alpha(\text{NNN})$</td>
<td>$R(\text{NN})$</td>
</tr>
<tr>
<td>Tolerance $a\Delta g = 3 \times 10^{-4}$ (Q-Chem default)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>60.00</td>
<td>6</td>
</tr>
<tr>
<td>1.4200</td>
<td>60.00</td>
<td>4</td>
</tr>
<tr>
<td>1.4600</td>
<td>70.00</td>
<td>16</td>
</tr>
<tr>
<td>1.5400</td>
<td>50.00</td>
<td>5</td>
</tr>
<tr>
<td>1.6000</td>
<td>90.00</td>
<td>8</td>
</tr>
<tr>
<td>Tolerance $\Delta g = 3 \times 10^{-5}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>60.00</td>
<td>7</td>
</tr>
<tr>
<td>1.4200</td>
<td>60.00</td>
<td>5</td>
</tr>
<tr>
<td>1.4600</td>
<td>70.00</td>
<td>16</td>
</tr>
<tr>
<td>1.5400</td>
<td>50.00</td>
<td>6</td>
</tr>
<tr>
<td>1.6000</td>
<td>90.00</td>
<td>9</td>
</tr>
<tr>
<td>Tolerance $\Delta g = 1 \times 10^{-5}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>60.00</td>
<td>7</td>
</tr>
<tr>
<td>1.4200</td>
<td>60.00</td>
<td>5</td>
</tr>
<tr>
<td>1.4600</td>
<td>70.00</td>
<td>17</td>
</tr>
<tr>
<td>1.5400</td>
<td>50.00</td>
<td>6</td>
</tr>
<tr>
<td>1.6000</td>
<td>90.00</td>
<td>9</td>
</tr>
</tbody>
</table>

$^a$Tolerance on the maximum gradient component. The optimization procedure converges if this and one of the two criteria — energy change $\Delta E = 10^{-6}$ a.u. and maximum displacement $\Delta q = 1.2 \times 10^{-3}$ Å — are satisfied. $^b$Nuclear repulsion energy $E_{nr} = 53.441794$ a.u., total energy $E_{tot} = -162.822635$ a.u. $^c$Nuclear repulsion energy $E_{nr} = 53.528406$ a.u., total energy $E_{tot} = -162.821900$ a.u.

### 5.4.2 Nitrogen dioxide

The PESs of the ground and first excited states of the NO$_2$ molecule, $X^2A_1$ and $A^2B_2$, form an intersection$^{35}$. Crossing points have been located for different N–O bond distances with the O–N–O angle between 107° and 108°$^{36}$. 

146
When the molecule is constrained to the $C_{2v}$ point group symmetry, the spacial parts of the wavefunctions of the states belong to different irreps, $A_1$ and $B_2$, and the derivative coupling matrix element between them vanishes.

If the symmetry is lowered further to the $C_s$ point group symmetry, and the N–O bond lengths are allowed to vary independently, both the ground state and first excited state wavefunctions become of the $A'$ symmetry. The derivative coupling matrix element between the two states is no longer zero, and the intersection becomes truly conical.

The calculations were performed using the EOMIP-CCSD method. The closed-shell CCSD wavefunction for NO$_2^-$ was used as the reference for the EOM-IP calculations.

To locate and minimize the seam numerically, the two surfaces were first scanned. The N–O distance was varied from 1.20 Å to 1.40 Å, and at each point the O–N–O angle was varied from 90$^\circ$ to 130$^\circ$, as shown in Fig. 5.3. These results are summarized in Table 5.3. The minimum of the seam was located at $R$(NO) = 1.3045 Å, $\alpha$(ONO) = 106.75$^\circ$. 

Figure 5.2: $D_{3h}$-constrained scans of adiabatic PESs of N$_3^+$ around the points of intersections calculated by EOM-CCSD/6-31G.
Figure 5.3: The seam of the $X^2A_1/A^2B_2$ intersection in nitrogen dioxide. At every N–O distance, the PESs of both states were scanned to find the O–N–O angle at which the states are degenerate.

Table 5.3: Points of the $C_{2v}$ constrained seam of the $X^2A_1/A^2B_2$ intersection in nitrogen dioxide. All bond lengths are given in angstroms, all angles are given in degrees, all energies are given in a.u.

<table>
<thead>
<tr>
<th>$R$(NO)</th>
<th>$\alpha$(ONO)</th>
<th>$E$ ($X^2A_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2000</td>
<td>106.027</td>
<td>-204.223693</td>
</tr>
<tr>
<td>1.2200</td>
<td>106.125</td>
<td>-204.233934</td>
</tr>
<tr>
<td>1.2400</td>
<td>106.241</td>
<td>-204.241419</td>
</tr>
<tr>
<td>1.2600</td>
<td>106.376</td>
<td>-204.246503</td>
</tr>
<tr>
<td>1.2800</td>
<td>106.531</td>
<td>-204.249495</td>
</tr>
<tr>
<td>1.3000</td>
<td>106.706</td>
<td>-204.250673</td>
</tr>
<tr>
<td>1.3045</td>
<td>106.748</td>
<td>-204.250713</td>
</tr>
<tr>
<td>1.3200</td>
<td>106.901</td>
<td>-204.250279</td>
</tr>
<tr>
<td>1.3400</td>
<td>107.117</td>
<td>-204.248528</td>
</tr>
<tr>
<td>1.3600</td>
<td>107.353</td>
<td>-204.245608</td>
</tr>
<tr>
<td>1.3800</td>
<td>107.611</td>
<td>-204.241686</td>
</tr>
<tr>
<td>1.4000</td>
<td>107.889</td>
<td>-204.236910</td>
</tr>
</tbody>
</table>
Table 5.4: Results of $C_{2v}$-constrained energy minimization of the $X^2A_1/A^2B_2$ intersection seam in NO$_2$ calculated by EOMIP-CCSD/6-31G. All bond lengths are given in angstroms, all angles are given in degrees.

<table>
<thead>
<tr>
<th>Starting geometry</th>
<th>Iterations</th>
<th>Final geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$(NO)</td>
<td>$\alpha$(ONO)</td>
<td>$R$(NO)</td>
</tr>
<tr>
<td>Tolerance$^a$ $\Delta g = 3 \times 10^{-4}$ (Q-Chem default)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>100.00</td>
<td>13</td>
</tr>
<tr>
<td>1.3000</td>
<td>90.00</td>
<td>14</td>
</tr>
<tr>
<td>1.3000</td>
<td>120.00</td>
<td>15</td>
</tr>
<tr>
<td>Tolerance $\Delta g = 3 \times 10^{-5}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>100.00</td>
<td>15</td>
</tr>
<tr>
<td>1.3000</td>
<td>90.00</td>
<td>18</td>
</tr>
<tr>
<td>1.3000</td>
<td>120.00</td>
<td>17</td>
</tr>
<tr>
<td>Tolerance $\Delta g = 1 \times 10^{-5}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2000</td>
<td>100.00</td>
<td>17</td>
</tr>
<tr>
<td>1.3000</td>
<td>90.00</td>
<td>19</td>
</tr>
<tr>
<td>1.3000</td>
<td>120.00</td>
<td>18</td>
</tr>
</tbody>
</table>

$^a$Tolerance on the maximum gradient component. The optimization procedure converges if this and one of the two criteria — energy change $\Delta E = 10^{-6}$ a.u. and maximum displacement $\Delta q = 1.2 \times 10^{-3}$ Å — are satisfied. $^b$Nuclear repulsion energy $E_{nr} = 61.603681$ a.u., total energy $E_{tot} = -204.250712$ a.u.

Various starting geometries were used in order to test the performance of the algorithm. Table 5.4 contains the initial points attempted and the results obtained.

### 5.4.3 para-Benzyne diradical

The para-benzyne diradical has the $^1A_g$ ground state and a low-lying $^3B_{1u}$ triplet excited state$^{37–42}$. In both states, the equilibrium structure is of the $D_{2h}$ point group symmetry.
Figure 5.4: The singlet and the $M_z=0$ triplet target states of the para-benzyne diradical derived from the high-spin triplet reference by spin-flip.

Figure 5.5: Geometrical parameters used to describe the para-benzyne diradical in Table 5.5.

The ground singlet state and the excited triplet state are derived from distributions of two electrons over the lower $b_{1u}$ and the higher $a_g$ frontier molecular orbitals. These states can be accurately described by the SF method\textsuperscript{43} as spin-flipping excitations form the high-spin triplet reference state, as shown in Fig. 5.4.

The equilibrium geometries of the para-benzyne diradical were calculated at the EOMSF-CCSD/6-31G* level. The geometries agree with the SF-DFT ones from Ref. 40, as shown in Table 5.5.

Since the adiabatic excitation energy of the triplet state is only about 0.03 eV, one could anticipate a low-energy $^1A_g/^3B_{1u}$ intersection. The ground state equilibrium geometry found in\textsuperscript{40} was used as the starting point for locating the minimum of the crossing seam. The convergence was achieved in about 50 iterations. The progress of the intersection optimization is shown in Fig. 5.6.
Table 5.5: Equilibrium geometries and the MECP of the $^1A_g$ and $^3B_{1u}$ electronic states calculated by EOMSF-CCSD/6-31G*. Geometrical parameters are defined in Fig. 5.5. All bond lengths are given in angstroms, all angles are given in degrees.

<table>
<thead>
<tr>
<th>Method</th>
<th>$R$(CC$_1$)</th>
<th>$R$(CC$_2$)</th>
<th>$R$(CH)</th>
<th>$\alpha$(CCC$_1$)</th>
<th>$\alpha$(CCC$_2$)</th>
<th>$\alpha$(CCH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1A_g$ ground state</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-DFT(50,50)/6-31G*$_a$</td>
<td>1.419</td>
<td>1.358</td>
<td>1.077</td>
<td>124.8</td>
<td>117.6</td>
<td>123.6</td>
</tr>
<tr>
<td>CASSCF(8,8)/6-31G*$_b$</td>
<td>1.411</td>
<td>1.383</td>
<td></td>
<td>124.0</td>
<td>118.0</td>
<td></td>
</tr>
<tr>
<td>EOMSF-CCSD/6-31G*$_c$</td>
<td>1.424</td>
<td>1.374</td>
<td>1.087</td>
<td>124.4</td>
<td>117.8</td>
<td>123.0</td>
</tr>
</tbody>
</table>

| $^3B_{1u}$ excited state|       |             |         |                    |                    |               |
| SF-DFT(50,50)/6-31G*    | 1.398     | 1.370       | 1.079   | 126.8              | 116.6              | 122.3         |
| CASSCF(8,8)/6-31G*      | 1.403     | 1.387       |         | 125.3              | 117.4              |               |
| EOMSF-CCSD/6-31G*$_d$   | 1.406     | 1.383       | 1.088   | 126.2              | 116.9              | 121.9         |

| $^1A_g$/$^3B_{1u}$ intersection| |     |         |                    |                    |               |
| EOMSF-CCSD/6-31G*$_e$     | 1.388     | 1.366     | 1.101   | 137.4              | 111.3              | 124.0         |

$_a$Reference$^{40}$.  
$_b$Reference$^{44}$.  
$_c$Nuclear repulsion energy $E_{nr} = 185.259520$ a.u., total energy $E_{tot} = -230.178701$ a.u.  
$_d$Nuclear repulsion energy $E_{nr} = 185.220983$ a.u., total energy $E_{tot} = -230.171962$ a.u.  
$_e$Nuclear repulsion energy $E_{nr} = 187.472024$ a.u., total energy $E_{tot} = -230.154762$ a.u.
Figure 5.6: Convergence of the $^1A_g/ ^3B_{1u}$ intersection optimization in the para-benzyne diradical by EOMSF-CCSD/6-31G*. In the upper part, the relative energies of both states are plotted for each iteration. The corresponding energy gaps are shown in the lower part. Convergence is achieved in about 50 iterations.
Figure 5.7: Relative energies of the $^1A_g$ and $^3B_{1u}$ states of *para*-benzyne diradical at their equilibrium structures and MECP. Both at the singlet and triplet equilibrium geometry, the singlet is below the triplet. The two states intersect 0.65 eV above the ground state minimum.

The MECP was located 0.65 eV above the ground state equilibrium energy as shown in the energy diagram in Fig. 5.7. The nuclear configuration at this point is significantly distorted compared to the ground state geometry. Note that although this distortion increases the splitting between the frontier MOs due to increased through-space interaction between the radical centers, it stabilizes the $a_g$ orbital, due to the favorable through-bond interactions.\textsuperscript{45}

5.5 Discussion: Performance of the algorithm and details of implementation

Our implementation of the projected gradient algorithm employs the minimization technique used in Q-Chem: the Newton-Raphson method in conjunction with the eigenvalue following algorithm.\textsuperscript{46} An approximate guess Hessian matrix used for the optimization
is updated using the BFGS formula. We modified the EOM-CCSD/CCSD code\textsuperscript{29} to allow simultaneous analytical calculation of the gradients for two states. The two components of the gradient (5.19) used in the minimization procedure are evaluated as given by (5.17) and (5.18).

We found that the convergence of the algorithm strongly depends on the relative absolute values of the two gradient components. If one of the components has a much greater absolute value than the other, after the step is taken, the smaller gradient changes its value significantly at the new point. Often, this leads to surges and oscillations of the energy during the minimization that spoils the convergence. For example, the optimization of the $^1A_g/r^3B_1u$ intersection in the \textit{para}-benzyne diradical fails if the $G_1$ and $G_2$ vectors are taken with the same weights, as in (5.19). We employed a modified version of the algorithm when the two components of the total gradient can be scaled with coefficients specified by the user. When the $G_2$ vector is scaled such that its absolute value is approximately the same as that of the $G_1$ vector, the procedure converges as shown in Fig. 5.6.

In this context, the Lagrange multiplier algorithm (e.g., Ref. 14) seems to be more robust, as it adjusts the coefficients automatically during the optimization procedure, thus insuring proper treatment of the problem. This, however, should not be considered a serious drawback of the projected gradient algorithm as with user-controlled coefficients comparable performance is achieved.

The scaling problem is related to the quality of the Hessian. In the straightforward implementation the approximate Hessian for the upper PES is used, which is not the Hessian of the function which is being optimized. The issue could be rectified by improving Hessian update procedure. Alternatively, scaling coefficients can be used in order to improve convergence. By default, the components of the gradient are not
scaled. However, the tests showed that scaling the $G_2$ vector to one-tenth of its original absolute value significantly improves the convergence.

Yarkony$^{47}$ and Chachiyo and Rodriguez$^{17}$ addressed this problem by modifying the algorithms as follows. Rather than optimizing the energy difference, they minimize the average energy between the states, which demonstrated better performance.

### 5.6 Chapter 5 conclusions

We implemented a reduced variant of the projected gradient method$^{16}$ within the equation-of-motion coupled-cluster (EOM-CC) family of methods in the Q-Chem quantum chemistry package$^{22}$. The algorithm employs analytic gradients$^{29}$ and assumes zero derivative couplings. Thus, it is applicable for minimizing the crossing seams between states of different symmetry and/or multiplicity. The method was applied to locate the MECPs between: (i) the excited states of $N_3^+$ described by EOMEE-CCSD; (ii) the two lowest states of NO$_2$ described by EOMIP-CCSD; and (iii) the singlet and triplet states of the *para*-benzyne diradical described by EOMSF-CCSD.

The benchmarks demonstrated that the convergence depends on the quality of the initial guess, as well as local steepness/curvature of the seam, e.g., the convergence is poor when the magnitude of the two gradient components is very different. Scaling of the two components of the gradient was shown to improve the convergence.
5.7 Chapter 5 references


[37] E. Kraka and D. Cremer. CCSD(T) investigation of the bergman cyclization of
enediyne. relative stability of o-, m-, and p-didehydrobenzene. J. Am. Chem. Soc.,

[38] T.D. Crawford, E. Kraka, J.F. Stanton, and D. Cremer. Problematic p-benzyne:

[39] L.V. Slipchenko and A.I. Krylov. Singlet-triplet gaps in diradicals by the spin-flip

[40] Y. Shao, M. Head-Gordon, and A.I. Krylov. The spin-flip approach within time-
dependent density functional theory: Theory and applications to diradicals. J.


[42] V. Vanovschi, A.I. Krylov, and P.G. Wenthold. Structure, vibrational frequencies,
ionization energies, and photoelectron spectrum of the para-benzyne radical anion.

method for a description of excited states, bond-breaking, diradicals, and trirad-

[44] A.E. Clark and E.R. Davidson. Model studies of hydrogen atom addition and
abstraction processes involving ortho-, meta-, and para-benzenes. J. Am. Chem.

[45] K.K. Baldridge, T.R. Buttersby, R.V. Clark, and J.S. Siegel. Does \( \pi - \sigma - \pi \)
through bond coupling significantly increase C-C bond lengths? J. Am. Chem.


[47] D.R. Yarkony. Marching along ridges. an extrapolatable approach to locating
Chapter 6: Future work

As exemplified by the calculations here as well as other works\textsuperscript{1–3}, it is possible to produce theoretical estimates of real life experimental data on large systems using high quality modeling techniques. The accuracy of such estimates depends directly on the amount of detail included in the model. Typical biochemical reactions involve a relatively small center, where chemical bonds are actually broken and created, and surrounding solvent or protein environment. While the environment does not participate in the reaction directly, it influences virtually every aspect of the process.

Hybrid quantum mechanical/molecular mechanical (QM/MM) techniques are designed to give a handle on such complex systems. Their principle is to employ \textit{ab initio} theory for the reaction center and describe the surroundings using less sophisticated and usually much cheaper methods.

The QM/MM approach has been used successfully for single point calculations and simple PES explorations that can yield basic energetics of chemical reactions. However, modeling vibration spectra, free energy, reaction dynamics, and many other studies require extensive PES sampling. For example, to calculate vibrational states one needs multidimensional PES scans, and \textit{ab initio} molecular dynamics (AIMD) and Monte Carlo simulations call for random points on PES.
Figure 6.1: Large biological systems such as this active center of GFP require treatment of the QM core and the MM environment using hybrid QM/MM methods.

Using CC and EOM methods for the quantum subsystem in QM/MM for the purpose of sampling PES has a prohibitive cost. Because we desire to keep the attractive properties of the methods, rather than substituting them with cheaper methods, we propose to increase their efficiency to the extent where direct AIMD calculations on multiple PES become feasible.

The most challenging problem of CC and EOM theories is their computational complexity. CCSD and EOMEE-CCSD scale as $N^6$, where $N$ is the size of the molecule. That means that a dimer requires roughly 64 times more computation effort than the original system. Our approach is to attack the scaling by performing CC and EOM calculations using reduced-rank (compressed) tensors.
Local and reduced-rank methods. The physical principle underlying reduced scaling electronic structure methods is the locality of electronic correlation. Previous attempts to use it have proven successful resulting in near linear-scaling local CCD\(^4\) and CCSD\(^5\) implementations. However, the approach has two serious shortcomings. One is that the equations in the entire hierarchy of CC methods have to be rederived and reprogrammed. The other is the doubt that the same principle of locality can be applied to both ground and excited states equally.

Tensor decomposition techniques can be used to overcome the problems of the local methods. They put a solid mathematical base under the same physical idea, and with the new tensor library the original CC and EOM equations need not be reimplemented.

The reduced-rank tensor representations are based on tensor decompositions in which \(d\)-dimensional tensors are factorized into a sum of vectors or lower-dimensionality tensors\(^6\text{--}11\). The two main strategies are canonical product (CP)\(^12\) and Tucker (or HOSVD)\(^13\) decompositions. The TT (Tree-Tucker, Three-dimensional Tensor) decomposition\(^14,15\) combines the qualities of CP and HOSVD and resolves some of their shortcomings.

Canonical product decomposition. Following a recent proposal supported by encouraging preliminary results\(^12\), we first consider the CP or canonical decomposition (CANDECOMP) format\(^6\). In this format, a four-dimensional tensor \(A(w,x,y,z)\) (i.e., two-electron integrals or \(T_2\) amplitudes) is expanded as:

\[
A = \sum_{r=1}^{R} a^w_r \otimes a^x_r \otimes a^y_r \otimes a^z_r, \quad (6.1)
\]

where one-dimensional vectors \(a^\mu_r\) are the representing vectors and the summation length \(R\) is called rank of the representation.
The trivial CP decomposition is obtained as follows: the first representing vector codes the index of the first dimension as a unit vector, the second one does the same for the second dimension, and the last representing vector contains the values corresponding to such multi-index. Obviously, the rank of such decomposition is \( n^3 \) (for the two-electron integrals), and the storage requirements \( (R = n^3) \) sets of four \( n \)-dimensional vectors need to be stored) are increased relative to the full four-index tensor \( (n^4 \) elements). However, if the rank can be significantly reduced, then the storage requirements \( (4 \cdot n \cdot R) \) of the decomposed tensor can become much smaller and linear scaling can, in principle, be achieved.

Recently, an efficient (and easily parallelizable) rank-reduction algorithm (based on the accelerated gradient method) have been developed by Hackbusch and coworkers \(^{12}\) and applied to two-electron integrals and MP2 amplitudes. They demonstrated that ranks of these basic quantities can be significantly reduced by this algorithm giving rise to \( n^{1.2} - n^{2.5} \) depending on the thresholds used. The scaling of their algorithm depends on the ratio of the initial and final ranks. Thus, the cost of the rank reduction step can be decreased if one starts with a representation which is more compact than the trivial decomposition. The authors \(^{12}\) suggested that using RI or Cholesky decomposed integrals as a starting point of the rank reduction procedure can considerably increase the efficiency of this most time-consuming step.

Once the integrals are represented in the CP format, the contractions with any two-index tensors (MO coefficients \( C \) or \( T_1 \) amplitudes) can be trivially performed at the
reduced cost. For example, the integral transformation (from AO to MO basis) assumes the following form:

$$\langle ab \mid cd \rangle = \sum_{\mu \nu \lambda \sigma} C_{\mu a} C_{\nu b} C_{\lambda c} C_{\sigma d} \sum_r \xi^\mu_r \otimes \xi^\nu_r \otimes \xi^\lambda_r \otimes \xi^\sigma_r$$

$$= \sum_r \left( \sum_{\mu} C_{\mu a} \xi^\mu_r \right) \otimes \left( \sum_{\nu} C_{\nu b} \xi^\nu_r \right) \otimes \left( \sum_{\lambda} C_{\lambda c} \xi^\lambda_r \right) \otimes \left( \sum_{\sigma} C_{\sigma d} \xi^\sigma_r \right)$$

$$= \sum_r v^a_r \otimes v^b_r \otimes v^c_r \otimes v^d_r$$ \hspace{1cm} (6.2)

Thus, the complexity of the transformation is reduced to $O(n \cdot N_{\text{virt}} \cdot R)$, to be compared with the $O(n^4 \cdot N_{\text{virt}}^2 + n^3 \cdot N_{\text{virt}} + n^2 \cdot N_{\text{virt}}^2 + n \cdot N_{\text{virt}}^3)$ scaling of the conventional algorithm. Most importantly, the rank does not change upon this transformation.

The complexity of contractions between two four-index CP tensors is also reduced, however, such contractions do change the rank, i.e., the rank of the product obtained by contracting two tensors of ranks $R'$ and $R''$ equals $R' \cdot R''$, as demonstrated by the following example:

$$\sum_{ab} t_{ij}^{ab} \langle kl \mid ab \rangle = \sum_{ab} \left( \sum_{s=1}^{S} t^i_s \otimes t^j_s \otimes t^a_s \otimes t^b_s \right) \left( \sum_{r=1}^{R} v^i_r \otimes v^j_r \otimes v^a_r \otimes v^b_r \right)$$

$$= \sum_{q=1}^{R S} t^i_q [i] t^j_q [j] v^i_q [k] v^j_q [l] \left( \sum_{a} t^a_q [a] v^a_q [a] \sum_{b} t^b_q [b] v^b_q [b] \right)$$

$$= \sum_{q=1}^{R S} t^i_q \otimes t^j_q \otimes v^i_q \otimes v^j_q$$ \hspace{1cm} (6.3)

Thus, the rank reduction procedure should be applied after each such tensor operation.

A minor complication arises due to the terms involving energy denominators, which can be efficiently dealt with by the Laplace transform based on approximating $1/x$ by exponentials sums\textsuperscript{16–18}, as done in the Laplace MP2 method\textsuperscript{16–21}.
In sum, both the storage requirement and the cost of contractions between tensors represented in the low-rank CP format is dramatically reduced compared to the original tensors. However, the contractions (except when one of the tensors is a two-index tensor) lead to a rank increase. Thus, to preserve the low-rank representation, a rank reduction procedure should be performed after each contraction.

**Rank reduction.** The rank-reduction procedure amounts to the minimization problem of the difference between an original tensor $A$ and its lower-rank approximation, $\tilde{A}$:

$$||A - \tilde{A}|| \leq \varepsilon \quad (6.4)$$

where $\varepsilon$ determines the accuracy. The accelerated gradient (AG) algorithm\textsuperscript{22} has been successfully employed in Ref. 12. The most time-consuming step in the AG algorithm is the calculation of the gradient. The complexity of this step is $O(d \cdot \tilde{R} \cdot n \cdot (\tilde{R} + R))$ where $d$ is the dimension of the tensor, and $R$ and $\tilde{R}$ are the ranks of $A$ and $\tilde{A}$, respectively.

Thus, the cost of this step depends critically on the initial rank. The authors of Ref. 12, who employed the trivial decomposition as the initial step for rank reduction, have suggested that a better starting point can be provided by using RI or Cholesky decomposed integrals.

To further reduce the cost of the rank reduction algorithm, the sliced factorization procedure has been developed\textsuperscript{12}, in which the initial reduction is performed for smaller blocks of data. This algorithm can be trivially parallelized.

We plan to begin by implementing the parallelized AG procedure including sliced reduction following the algorithm described in Ref. 12. We then will explore the following approaches for improving the starting point for the rank reduction of the integrals:
(i) RI decomposed integrals; (ii) Cholesky decomposed integrals; (iii) using localized MOs; (iv) using frozen natural orbitals (FNO)\textsuperscript{23}.

**HOSVD decomposition.** There are multiple ways to define SVD for more than two dimensions. In this study we use the Tucker decomposition as high-order SVD

\begin{equation}
A = \sum_{t_1=1}^{R} \sum_{t_2=1}^{R} \sum_{t_3=1}^{R} \sum_{t_4=1}^{R} g(t_1, t_2, t_3, t_4) a(w, t_1) a(x, t_2) a(y, t_3) a(z, t_4)
\end{equation}

This form assumes a multidimensional core which is expanded by matrices $a(x, t)$. $R$ is the rank of Tucker decomposition. Because the core has the same dimensionality as the original tensor, the number of parameters in this decomposition does not have a reduced scaling. In the equation above, even after the reduction of the rank, the number of parameters grows as $O(n^4)$. Therefore, an improvement is required to use Tucker decomposition for reduced-scaling methods.

**TT decomposition.** In the TT decomposition, a target tensor $A(w, x, y, z)$ is represented as

\begin{equation}
A = \sum_{\alpha_1, \alpha_2, \alpha_3, \alpha_4} a_w^{\alpha_1} a_x^{\alpha_2} a_y^{\alpha_3} a_z^{\alpha_4}
\end{equation}

TT decomposition involves three-index tensors at most, thereby enabling reduced-scaling representation of higher-order tensors with appropriate rank reduction. In addition, the decomposition algorithm is based on partial SVD of unfolding matrices, which is better positioned to solve the rank reduction problem than the iterative algorithms used in CP decomposition.

* * *
The suite of CC and EOM methods, including calculations of energies, properties and gradients, is implemented in Q-Chem 4 using the new library for tensor algebra (see Appendix). The design of the library allows us to reuse the CC and EOM expressions, but introduce new types of tensors that are based on the decompositions.

This project has three milestones:

1. Enable the tensor library to store and use tensors in a readily available decomposition format, for example RI. While this step does not reduce the scaling of the method, it improves the performance by reducing the amount of the data set.

2. Introduce decomposition algorithms and corresponding tensor formats. Use the infrastructure created in Step 1 to use these new compressed formats.

3. Optimize specific tensor operations to take advantage of the compressed formats. If rank reduction is achieved successfully by the decompositions, this step will lead to a reduced-scaling implementation of the methods.
6.1 Chapter 6 references


Bibliography


R. Baer and A.I. Krylov. in preparation.


181


Appendix: Tensor algebra library

Efficient object-oriented implementation of CC and EOM-CC methods. Programmable expressions in wave function methods are dominated by tensor contractions (generalized matrix multiplications). For example, the CCD correlation energy is given by

$$E_{CCD} = \frac{1}{4} \sum_{i,j,a,b} t_{i,j}^{a,b} \langle i,j||a,b \rangle$$

where $t_{i,j}^{a,b}$ is the two-body cluster operator, and $\langle i,j||a,b \rangle$ is the electron repulsion integrals in the molecular orbital basis. Both are four-index antisymmetric tensors.

The challenge of implementing many-body methods lies in the complexity of the expressions, which consist of numerous tensor products. For example, the CCSD cluster amplitude equations involve more than thirty contractions between the amplitudes ($T_1$ and $T_2$) and the integrals. All of them have to be programmed correctly and efficiently.

Object-oriented computer languages such as C++ allow the programmer to manage the complexity and code such equations in a compact and readable form that facilitates validation and debugging. To take full advantage of this philosophy, we developed a set of expression templates that use overloaded C++ operators. They allow one to code programmable equations using an intuitive syntax.
For example, the second-order Møller-Plesset perturbation theory (MP2) electronic correlation energy is computed as

\[ E_{MP2} = \sum_{ia} t_i^a f_{ia} + \frac{1}{4} \sum_{ijab} t_{ij}^{ab} \langle ij||ab\rangle, \]

where

\[ t_i^a = \frac{f_{ia}}{\varepsilon_i - \varepsilon_a} \text{ and } t_{ij}^{ab} = \frac{\langle ij||ab\rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \]

The C++ function that computes the energy using the tensor algebra library may look similar to this.

```cpp
double mp2_energy(btensor<2> &t1, btensor<4> &t2,
                 btensor<2> &f_ov, btensor<4> &i_oovv) {

  letter i, j, a, b;

  return dot_product(t1(i|a), f_ov(i|a)) +
         0.25 * dot_product(t2(i|j|a|b), i_oovv(i|j|a|b));
}
```

The expression templates are designed so that most common programming mistakes are detected at compilation time rather than runtime thus improving the robustness of the code. When expressions are executed, they are translated into a series of block tensor operations. Since the conversion takes place at runtime, the library can take full advantage of its computational setup (details unknown at compilation time such as memory and disk availability, type and number of processors, etc.)

The level of abstraction offered by the expression templates allows us to introduce changes in the low-level library code without modifying the expressions. Once coded using this interface, all methods immediately take advantage of improvements in the
library. One example of such improvement is transition to a new computational platform. Once the tensor algebra library features the ability to use certain hardware or parallelism, all codes automatically take advantage of that. Other examples are new types of symmetry, different representations of tensors such as algebraic decompositions, or exploiting sparsity in local correlation approaches.

**Library for tensor algebra in electronic structure calculations.** We implemented the principles presented here in the form of a C++ tensor algebra library. It supports virtual memory (utilizing disk for storage when there is not enough RAM), symmetries found in electronic structure calculations and parallel computation on capable hardware. The library is now a part of Q-Chem.

The tensor library has a layered structure (Fig. A.1):

- The interface is a tool that enables the entry of programmable expressions by providing functions and overloaded arithmetic operators for tensor operations such as addition, scaling, contraction, dot and direct products. It also serves a secondary
purpose of preliminary optimization of the expression, which is converted to a sequence of block tensor operations.

- Block tensor operations are a set of routines that produce the symmetry and numerical result of basic and general block tensor linear algebra. Operations such as addition of multiple block tensors, general contraction of two block tensors, products of two block tensors are used as building blocks to compute complex expressions and implemented as individual classes. These routines first calculate the symmetry of the result, and then compute the numerical data accounting for symmetry and avoiding redundancies. Each unique block of the resulting block tensor is computed by invoking simple tensor operations on the blocks of the arguments. In parallel environments the blocks of the result can be calculated simultaneously using shared input data. Concurrency is implemented using MPI, OpenMP, native threads and other similar techniques.

- Simple tensor operations work on individual dense blocks of block tensors and perform tensor linear algebra on multidimensional arrays. Depending on the nature of the operation, it can compute the result directly or invoke lower-level numerical libraries.