Computational Studies of ThDP-Dependent Enzymes

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr.rer.nat.)

of the Georg-August-Universität Göttingen

within the doctoral program Chemistry

of the Georg-August University School of Science (GAUSS)

submitted by

Mirko Paulikat

from Wismar

Göttingen, 2018

Thesis Committee

Prof. Dr. Ricardo Mata
  Computational Chemistry and Biochemistry Group,
  Institute of Physical Chemistry

Prof. Dr. Kai Tittmann
  Department for Bioanalytics,
  Göttingen Center for Molecular Biosciences

Members of the Examination Board

Reviewer: Prof. Dr. Ricardo Mata
  Computational Chemistry and Biochemistry Group,
  Institute of Physical Chemistry

Second Reviewer: Prof. Dr. Kai Tittmann
  Department for Bioanalytics,
  Göttingen Center for Molecular Biosciences

Further members of the Examination Board:

Prof. Dr. Jörg Behler
  Department for Theoretical Chemistry,
  Institute of Physical Chemistry

apl. Prof. Dr. Burkhard Geil
  Department for Biophysical Chemistry,
  Institute of Physical Chemistry

Prof. Dr. Inke Siewert
  Department for Inorganic Chemistry IV,
  Institute of Inorganic Chemistry

Prof. Dr. Martin Suhm
  Department for Physical Chemistry II,
  Institute of Physical Chemistry

Date of the oral examination: 18.12.2018

Acknowledgements

First of all, I want to express my special thanks to Prof. Dr. Ricardo Mata for the supervision of this PhD project. He was always available for all my questions, which resulted in helpful remarks and stimulating discussions crucial for this thesis. That even goes back to my undergraduate studies, where he introduced me to the interesting field of Theoretical Chemistry with its practical applications. These basics were essential to prepare this thesis.

Next, I would like to thank Prof. Dr. Kai Tittmann for the willingness to be my second reviewer, but in particular for the collaboration in the field of ThDP-dependent enzymes. That includes stimulating discussions, but also the supply of high-quality experimental data on which a major part of this work is built on. In this context, I want also to thank my colleague Dr. Fabienne Libuda for the cooperation in the phosphoketolase project. I believe that both the experimental and theoretical sides gained mutual benefit for this project, which is a key ingredient in current research.

Furthermore, I wish to express my gratitude to my colleagues Axel Wuttke, Thomas Auth and Dr. Jon Uranga for carefully proof-reading parts of this thesis. Their comments and the subsequent discussions have been helpful for the completion of this thesis. Moreover, I would like to thank all former and current group members of the Computational Chemistry and Biochemistry group in Göttingen, who spent time with me in the course of this PhD project. That includes the useful discussions on my projects as well as the insights into their research topics extending my scope in the field of Theoretical Chemistry, and not to forget the entertaining group activities.

Last but not least, I would like to thank my whole family for any help in the course of this PhD project. Above all, I want to express my special thanks to my parents for their unrestricted support throughout my life, which laid the foundation for my personal development until the completion of this thesis. Even if my father cannot read these words anymore, given his untimely decease at the beginning of my PhD project, I would like to address the last words of my gratitude to him: You have always been a support to me. Your strength, even in the darkest times, will have shaped me for the rest of my life. Rest in peace.

Abstract

ThDP-dependent enzymes catalyze a variety of biochemical reactions in all domains of life. The mechanistic studies on these complex systems require a wide range of complementary bioanalytical tools. In this context, electronic absorption spectroscopy has proven itself as a valuable method for the identification of intermediates in ThDP-catalyzed reactions, thereby rationalizing proposed mechanisms. However, the unequivocal assignment of the absorption bands to the correct states of ThDP and intermediates is required for appropriate conclusions. This task is far from trivial considering the transient character of some intermediates as well as the complexity of the surrounding enzyme environment. Controlled experiments on model compounds, frequently employed for the assignments, can be unreliable because of the absence of critical conditions in enzymes. For this purpose, theoretical methods have been applied in this thesis to obtain a detailed view of the involved excitations including the analyses of the determining factors behind the different spectra. These studies provide novel insights into the spectral signatures of ThDP-dependent enzymes and an important contribution to their mechanistic understanding.

In a first step, the spectral signatures of the pre-equilibrium states of ThDP are investigated through model compound calculations up to a full QM/MM description of the cofactors in the *Zm*PDC enzyme. The observed CD bands in the near UV region could all be associated with charge-transfer excitations between the pyrimidine and the thiazolium rings of ThDP. In particular, the fundamental role of the protonation state of the canonical glutamate is highlighted for the location of the ThDP bands. These studies result in the simultaneous assignment of the chemical states of both the cofactor and the activating glutamate to the two previously proposed spectral fingerprints of ThDP, but also to a third hitherto unassigned band.

Investigations have also been carried out on spectra of enzymatic on-pathway intermediates. Calculations for the tautomers of 2-acetyl-ThDP have helped to unravel the experimental observations in the *B. breve* PK enzyme. Both the keto and enolate states of this intermediate are assigned to a common band in the UV-vis spectra. The latter plays a central role for the mechanism of PK through kinetic stabilization in absence of the co-substrate, preventing the system from off-pathway hydrolysis. The second intermediate of particular interest has been 1,2-dihydroxyethyl-ThDP, which was previously identified with a key role in the TK enzyme mechanism. UV-vis measurements

had revealed an uncommon broad absorption band with a range of about 1 eV. Spectra calculations in the active site of *E. coli* TK have helped to identify the source of this observation. A proton transfer equilibrium from the intermediate to a histidine residue is proposed which connects the two limits of the observed band.

The new information gathered from the studies of ThDP and the aforementioned intermediates is a key contribution to understanding their reactivity. The low-lying excited states of the proposed intermediates have been analyzed with respect to their excitation characters. This has lead to a classification of the spectral fingerprints. Subsequent analyses have identified a structure-excitation energy relationship for ThDP intermediates. These results provide a roadmap for the interpretation of absorption spectra of ThDP-dependent enzymes.

The unequivocal assignment of electronic absorption bands through calculated excitation energies requires efficient and robust quantum chemical approaches. Coupled-cluster methods have proven to provide accurate excitation energies, but suffer from the steep scaling of the computational cost regarding the molecular system size. This has motivated the development of low-scaling alternatives. For example, incremental approaches have been previosly applied to reduce the effort with successful performance on ground state calculations. This concept was extended to excited states in an incremental EOM-CCSD approach by Mata and Stoll. Here, a novel scheme with an error of -0.01 ± 0.12 eV for a selected test set is proposed. The approach employs the density overlap between the reference excitation space and the orbitals for a robust increment construction. Additionally, this criterion has allowed the identification of fragments, which can be excluded from the incremental approach with virtually no loss of accuracy. It is expected that the developed method becomes a powerful tool for the study of electronic excitation spectra of large molecular systems by introducing a local treatment of the virtual orbital space.

Contents

1

Introduction

Thiamin diphosphate (ThDP) is the metabolically active form of vitamin B_1. The latter is an essential nutrient, sythesized only in bacteria, funghi and plants. Humans and animals have to ingest it through their diet.[1] The thiamin diphosphokinase enzyme catalyzes the conversion of the vitamin to ThDP,[2] which is employed as a cofactor in enzyme catalysis in all these domains of life.[3] It can be chemically subdivided into three structural units. A 4'-aminopyrimidine and a thiazolium ring are conjoined over a methylene group. The diphosphate anchor is linked to the thiazolium ring by an ethyl group. The structural formula of ThDP together with the atom nomenclature is shown in Figure 1.1.

Each of the structural units is important for the activity. The diphosphate group combined with a second cofactor, a divalent metal ion, is required for binding to the enzyme environment. The actual reactive center is, however, placed at the thiazolium ring.[4;5] Deprotonation of its C2 atom leads to the formation of the reactive carbene. This enables a nucleophilic attack on carbonyl groups of substrates, followed by the
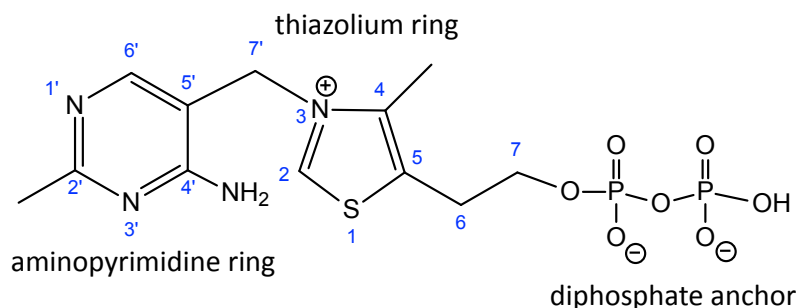

Figure 1.1.: Structural formula and nomenclature of thiamin diphosphate (ThDP).

processing of the substrate-ThDP conjugates (see Figure 1.2a). The latter involves the real strength of ThDP-dependent enzyme catalysis: the reversible cleavage and formation of covalent bonds, including C–S, C–N, C–O and, the most prominent ones, C–C bonds. It is therefore not surprising that ThDP-dependent enzymes are involved in a variety of biochemical processes.[3] For example, the pyruvate decarboxylase enzyme catalyzes the non-oxidative decarboxylation of pyruvate required in the pathway of fermentation processes.[6] Furthermore, the phosphoketolase and transketolase enzymes are essential in the non-oxidative branch of the pentose phosphate pathway in carbohydrate metabolism.[7–9] The latter is also involved in the Calvin cycle of photosynthesis.[10] In closing this short illustration of the importance of ThDP-dependent enzymes, their role for the citric acid cycle is mentioned, being part of the pyruvate dehydrogenase complex as well as the α-ketogluatarate dehydrogenase complex.[11;12]

Although the reactive center of ThDP is placed at the thiazolium ring, the 4'-aminopyrimidine ring is found to have a cocatalytic function in the enzymatic reactions. Two structural motifs are observed for ThDP enzymes which enable this functionality. Firstly, all hitherto characterized enzymes revealed a canonical conformation of the enzyme-bound cofactor,[13;14] the so-called *V* conformation (see Figure 1.2b). Therein, the exocyclic amino group points directly towards the reaction center allowing the influence on the mechanism. Secondly, with one notable exception,[15] a canonical glutamate is observed for all ThDP-dependent enzymes placed directly next to the N1' atom of the pyrimidine ring. This residue is expected to facilitate the tautomerization of the 4'-aminopyrimidine (AP) state of ThDP to its 1',4'-iminopyrimidine (IP) form, including the 4'-aminopyrimidinium (APH$^+$) as an intermediate state.[16] The flexibility of the protonation states of the pyrimidine moiety combined with the favoured *V* conformation provide the possibility for intramolecular Brønsted acid-base catalysis. This plays a critical role for cofactor activation, substrate binding and product release.[16–18]

In order to study the mechanism of ThDP-dependent enzymes, three experimental methods have been predominantly employed. Firstly, structural analysis by protein crystallography and mutagenesis studies are combined with steady-state kinetic measurements.[5;19–21] Beside the structural information, this methodology can identify enzyme residues with a critical role for the reaction mechanism. Secondly, acid-quench isolation of reaction intermediates are used in combination with ^{1}H NMR spectroscopy.[5;22] This analytical tool leads to the identification of the chemical state of the thiazolium moiety including covalently bonded substituents. Finally, electronic absorption spectroscopy is

employed to characterize ThDP-catalyzed reactions.[23–27] This may be either in form of conventional ultraviolet-visible (UV-vis) absorption spectroscopy or in circular dichroism (CD) experiments. The requirement for their application in enzyme catalysis is the absorbance of the studied chromophore at wavelengths larger than approximately 280 nm,[28] as illustrated in Figure 1.2. At shorter wavelengths the large amount of protein residues start to absorb which make the unequivocal identification of non-protein compounds impossible. Fortunately, enzyme-bound ThDP and intermediates of its cat-
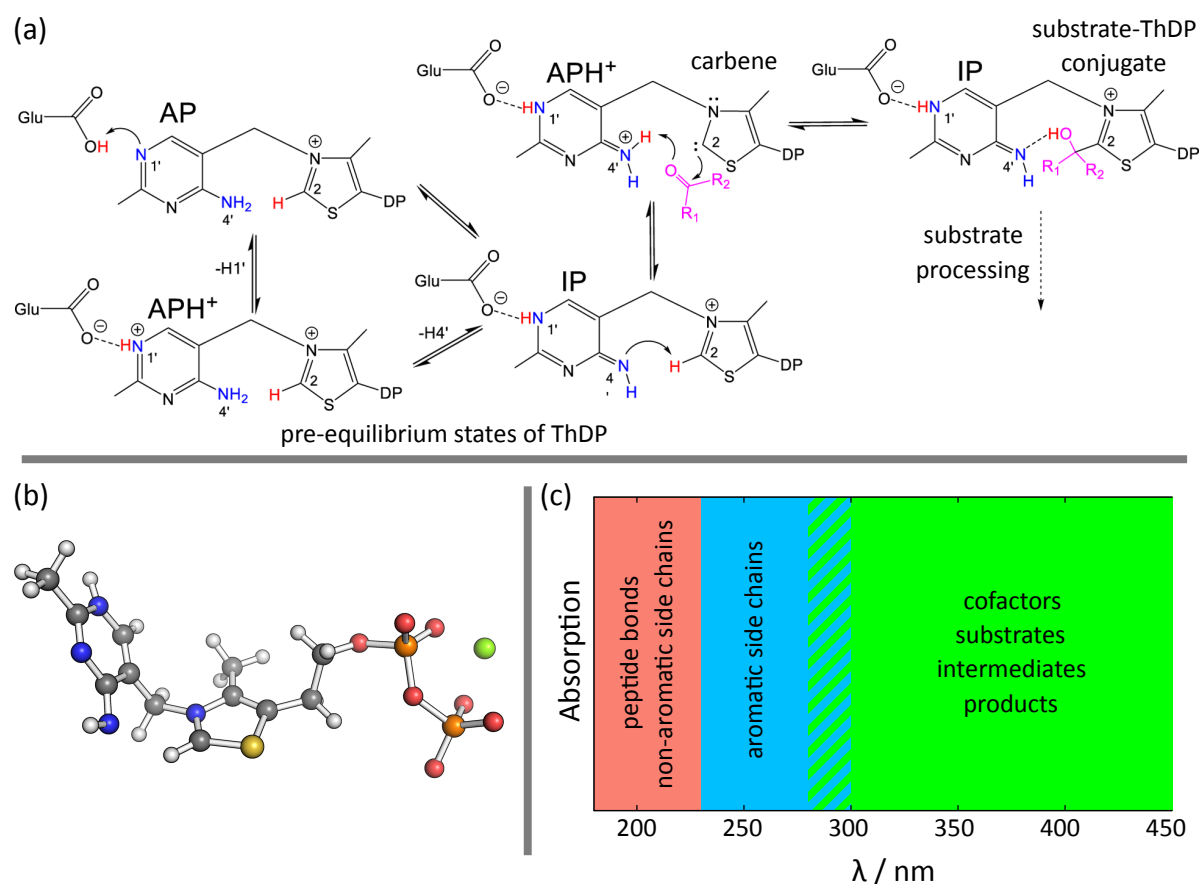
Figure 1.2.: (a) Proposed mechanism of cofactor activation and substrate binding in ThDP-dependent enzymes including the canonical glutamate. The latter is supposed to mediate the tautomerization process of the pre-equilibrium states of ThDP. Abbreviations: AP – 4'-aminopyrimidine, IP – 1',4'-iminopyrimidine, APH+ – 4'-aminopyrimidinium, DP – diphosphate moiety of ThDP. (b) Ball-and-sticks representation of IP-ThDP in its enzyme-bound *V* conformation together with the second cofactor, a divalent metal ion, shown as green sphere. The exocyclic imino group is juxtaposed to the C2 atom of ThDP, being the reactive center of the cofactor. (c) Schematic representation of common spectral ranges of electronic absorptions in proteins.

alyzed reactions are frequently observed in the higher wavelength regime (>280 nm), so that the application of these techniques provide valuable information for mechanistic studies. However, the unequivocal assignment of the measured bands to the correct states of ThDP and intermediates is required for appropriate conclusions on the reaction mechanisms. This task is far from trivial given the transient nature of some states and the complexity of the surrounding enzyme environment. In the past, controlled experiments with model compounds have oftentimes been used to characterize unique spectral fingerprints of ThDP-dependent enzymes,[29;30] but even these can sometimes be unreliable as is later demonstrated within this work. Therefore, the main focus of this work was on the investigation of the spectral signatures of ThDP-dependent enzymes through theoretical methods. These studies supply a detailed view of the electronic excitations responsible for the measured absorption bands. The latter not only consist of calculated excitation energies for the assignment of ThDP and its intermediates to such bands, but also include information on their excitation character, the impact of the cofactor conformation and even allow the assignment of protonation states of titratable residues within the active site.

The above mentioned tasks require robust and efficient quantum chemical methods in order to provide reliable data for the interpretation of electronic absorption spectra. Up to this date, the most robust methods represent the equation-of-motion coupled-cluster (EOM-CC) approaches for the calculation of singly excited states in organic chromophores.[31;32] However, these methods suffer from their steep scaling of the computational cost regarding the molecular system size. In order to reduce the effort, incremental approaches can be employed.[33] Thereby, the system is split into orbital groups based on a reference calculation. For ground state calculations with single reference character Hartree-Fock is the method of choice. The total correlation energy can then be obtained through a many-body expansion, computing the fragments with higher level methods. Considering the short-range nature of electron correlation, a local treatment of this approach proved successful in the calculation of ground states for molecules, ionic crystals and semiconductors.[33–37] An extension of this methodology to excited states was proposed by Mata and Stoll.[38] Therein, a mixed occupied orbital space is employed. This basis contained the dominating natural transition orbitals from a reference configuration interaction with singles calculation in order to describe the excited states of interest. The remaining occupied orbitals were localized to improve the convergence of short-range correlation effects. After the construction of the fragments based on the

distances of orbital charge centroids, vertical excitation energies could then be calculated in a many-body expansion at the EOM-CCSD level of theory.[39] The application to low-lying excited states of organic molecules showed the capability of this approach to recover the excitation energies of the conventional method. However, the benchmark calculations also revealed some issues of the implemented version. The largest deviations in the two-body approach were attributed to deficiences in the fragment criteria, arising from delocalization effects in combination with the orbital charge centroid criterion or the missing information of the virtual natural transition orbitals. In this work, the promising idea of the incremental EOM-CCSD method is picked up and refined. The development focused on the improvement of the group construction algorithm in order to obtain a robust approximation to the conventional method. This includes new fragment criteria, identifying the most important orbitals for the excited state under study. Test calculations on an extended benchmark set and their careful analyses led to insights on how to further improve the performance of the incremental approach.

The structure of this thesis is organized as follows. In Chapter 2 an overview on the applied methods and included approximations is presented. These include the theoretical background of wave function based electronic correlation methods as well as their extensions for the calculation of excited states. Furthermore, the alternative quantum chemical concept, density functional theory, is described. Its time-dependent formulation provides a method to calculate excitation spectra. Afterwards, the classical counterpart for the energy calculations is presented in form of molecular mechanics. This is required in this work for the treatment and simulation of the large protein environment of ThDP-dependent enzymes containing thousands of atoms. The chapter is closed by the discussion of the quantum mechanics – molecular mechanics approach. This allows the study of ThDP at a high-level quantum chemical method within its enzyme environments treated at the efficient molecular mechanics level of theory.

Chapter 3 deals with the refinement of the incremental EOM-CCSD method. New fragment criteria are introduced for the identification of the most important orbitals for the excited state under study. These are benchmarked with different group construction algorithms to a set of small to medium sized organic chromophores. Careful analyses of these calculations provide information on necessary developments (partially already implemented in the course of this work) for the improvement of the incremental approach.

The investigations of the spectral signatures of the pre-equilibrium states of ThDP are

featured in Chapter 4. In a first step, model compounds are used to gain first insights into their absorption spectra. These studies also include the impact of the cofactor conformation as well as the revision of model compounds employed in experiment for the assignment of the ThDP bands. The target is then moved to the influence of the enzyme environment on the cofactor spectra. A newly designed procedure, a perturbative residue-by-residue analysis, is applied to the excited states of the ThDP tautomers in the active site of *Zymomonas mobilis* pyruvate decarboxylase (*Zm*PDC), which allows the identification of enzyme residues with significant impact in their spectra. Considering these information, MD-QM/MM calculations of ThDP in a functional dimer of *Zm*PDC are used for the assignment of the ThDP tautomers in combination with the chemical state of the canonical glutamate to the near UV CD bands.

Chapter 5 contains the studies of a key intermediate, 2-acetyl-ThDP, in the enzymatic reaction of phosphoketolase (PK). This intermediate was detected through UV-vis measurements of the Tittmann group.[40] Model compound calculations of its different tautomers are applied to explain the observed spectral features. Furthermore, a perturbative residue-by-residue analysis in the active site of *B. breve* PK identifies enzyme residues with different impact on the spectra of the tautomers. This information amounts to the calculation of excitation energy shifts of proposed mutant candidates for targeted modulation of the spectral properties. The chapter is closed with the discussion of reaction pathway calculations which were the focus of the Master's project of S. Wille.[41] Altogether, spotlights in the reaction mechanism of PK are worked out.

Chapter 6 features two distinct aspects of the transketolase (TK) enzyme. On the one hand, the spectral signatures of a key intermediate, 1,2-dihydroxyethyl-ThDP, are studied in the active site of *E. coli* TK as absorption measurements in the Tittmann group assigned this intermediate to a broad band in the near UV to visible light range.[40] Different protonation states of the intermediate and enzyme residues are considered in order to explain the spectral observations. On the other hand, a hydrogen bonded network is investigated in the human TK enzyme, which is proposed to allow the communication between two active sites. This study involves molecular dynamics simulations, verifying the stability of the network at ambient conditions. Furthermore, conformational changes of specific enzyme residues are analyzed in order to validate a proposed mechanism for the communication.

Chapter 7 deals with the characterization of the low-lying excited states of a set of proposed intermediates in ThDP-catalyzed reactions. Model compound calculations fol-

lowed by an excited state analysis are used to classify their spectral signatures. These are then investigated according to a structure-excitation energy relationship. The chapter is closed by the application of the incremental EOM-CCSD to selected ThDP intermediates in order to obtain more details on its current performance.

2

Theoretical Background

The fundamental task of computational chemistry is the calculation of the energy of a polyatomic system. In non-relativistic quantum mechanics (QM), the time-dependent Schrödinger equation is the starting point for the treatment of this problem.[42] By using atomic units, this equation corresponds to

$$i\frac{\partial}{\partial t}\left|\Psi(\boldsymbol{\tau},t)\right\rangle = \hat{H}\left|\Psi(\boldsymbol{\tau},t)\right\rangle \ . \tag{2.1}$$

Here, $\hat{H}$ denotes the total energy operator, the Hamilton operator, and Ψ stands for the wave function including all information of the system, which depends on all spatial and spin coordinates of the particles $\boldsymbol{\tau}$ as well as the time t as parameter. If the potential is time-independent, then the equation can be separated into a dynamic and a static part. The latter corresponds to the time-independent Schrödinger equation

$$\hat{H}\left|\Psi(\boldsymbol{\tau})\right\rangle = E\left|\Psi(\boldsymbol{\tau})\right\rangle \ , \tag{2.2}$$

where E is identified as the scalar total energy of the system associated with the static wave function $\Psi(\boldsymbol{\tau})$. Under the Born-Oppenheimer approximation,[42;43] the motions of the nuclei and electrons are decoupled that result in a further separation of Equation (2.2) into a nuclear and a electronic part. The latter is given by

$$\hat{H}^{\text{el}}\left|\Psi_n^{\text{el}}(\mathbf{r},\boldsymbol{\sigma};\mathbf{R})\right\rangle = E_n^{\text{el}}\left|\Psi_n^{\text{el}}(\mathbf{r},\boldsymbol{\sigma};\mathbf{R})\right\rangle \ , \tag{2.3}$$

where the index n stands for different electronic states with the associated electronic wave functions Ψ_n and electronic energies E_n^{el}. The electronic wave functions depend explicitly on the spatial ($\mathbf{r}$) and spin ($\boldsymbol{\sigma}$) coordinates of the electrons, but only parametrically on the nuclear configuration $\mathbf{R}$. Thus, the electronic energy depends in the same manner on the nuclear positions and constitutes the potential for the movement of the latter. In Equation (2.3), the electronic Hamiltonian $\hat{H}^{\mathrm{el}}$ is defined as

$$\hat{H} = \underbrace{-\frac{1}{2}\sum_i \nabla_i^2}_{\hat{T}_{\mathrm{e}}} \underbrace{-\sum_i \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}}_{\hat{V}_{\mathrm{en}}} + \underbrace{\sum_i \sum_{j>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{\mathrm{ee}}} + \underbrace{\sum_I \sum_{J>I} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}}_{\hat{V}_{\mathrm{nn}}} , \quad (2.4)$$

where the summations run over all electrons i and nuclei I of the system. Here, ∇_i stands for the nabla operator and Z_I is the nuclear charge. This Hamiltonian comprises the kinetic energy of the electrons ($\hat{T}_{\mathrm{e}}$), the electron-nuclear interactions ($\hat{V}_{\mathrm{en}}$), the electron-electron repulsions ($\hat{V}_{\mathrm{ee}}$) and the nuclear-nuclear interactions ($\hat{V}_{\mathrm{nn}}$).

In line with the approximations presented above, the electronic energy is the key quantity for the study of molecular systems and its determination together with related properties was a focal point of this work. This chapter provides an overview for the applied methods and the additional approximations therein. As the following sections only consider the calculation of electronic energies, the superscript (el) is henceforth omitted for all quantities.

2.1. Wave Function Theory

In wave function theory, the electronic energy is obtained by applying models for the wave function and subsequently solving the time-independent electronic Schrödinger equation (2.3). In a first step, the wave function of a molecular system is commonly expressed as a single Slater determinant (SD), built from spin orbitals. The latter are comprised of spin functions and spatial functions, the molecular orbitals (MOs). The orbitals are minimized variationally in order to obtain an estimate for the ground state energy. This procedure leads to a self-consistent field method, also called Hartree-Fock (HF) theory.[44;45] Here, the electrons are moving in their own averaged potential and thus, instantaneous electron-electron interactions (dynamic electron correlation) are not considered. Furthermore, the application of the variational principle only gives access

to the ground state of a molecular system, which is why electronically excited states are not directly available. In the following, wave function correlation methods are presented for the calculation of ground states as well as their extensions to describe excited states.

2.1.1. Configuration-Interaction Theory

In configuration-interaction (CI) theory the electronic wave function is expanded in a linear combination of SDs:[44;45]

$$|\Psi_{\mathrm{CI}}\rangle = \sum_i C_i |\Phi_i\rangle \ , \tag{2.5}$$

where C_i and $|\Phi_i\rangle$ denote the expansion coefficients and the SDs, respectively. In ground state calculations the HF configuration usually dominates the linear combination and is chosen as the reference state. The required excited SDs are generated by an excitation operator acting on the HF wave function:

$$|\Psi_{\mathrm{CI}}\rangle = \left(\hat{1} + \hat{T}\right)|\Phi_{\mathrm{HF}}\rangle = \left(\hat{1} + \hat{T}_1 + \hat{T}_2 + \cdots\right)|\Phi_{\mathrm{HF}}\rangle \ , \tag{2.6}$$

$$\text{where} \quad \hat{T}_1 = \sum_{i,a} C_i^a \hat{\tau}_i^a \tag{2.7}$$

$$\hat{T}_2 = \sum_{i>j,a>b} C_{ij}^{ab} \hat{\tau}_{ij}^{ab} \ . \tag{2.8}$$

Here, the $\hat{\tau}$ operators are shorthand notations for spin orbital excitation operators of different rank:

$$\hat{\tau}_i^a = a_a^\dagger a_i \ , \tag{2.9}$$

$$\hat{\tau}_{ij}^{ab} = a_a^\dagger a_i a_b^\dagger a_j \ , \tag{2.10}$$

with a_i and $a_a^\dagger$ being the spin orbital annihilation and creation operators, respectively. The excitation operator $\hat{T}$ can be divided into classes of operators which each produce all excited SDs of a certain rank, e.g. $\hat{T}_1$ for singly excited SDs, $\hat{T}_2$ for doubly excited SDs, and so forth.

The expansion coefficients are determined by a variational optimization of the energy expectation value. This procedure leads to a general eigenvalue problem, which in matrix

form is given by

$$\mathbf{HC} = \mathbf{EC} \quad \text{with } H_{IJ} = \langle \Phi_I | \hat{H} | \Phi_J \rangle \ . \tag{2.11}$$

The diagonal elements of the Hamiltonian matrix $\mathbf{H}$ contain the energies of the SDs, whereas the off-diagonal entries represent the coupling terms between the configurations. The coefficient matrix $\mathbf{C}$ consists of vectors of the linear coefficients for the electronic states and the diagonal matrix $\mathbf{E}$ comprises their corresponding energies.

An expression for the electron ground state energy can be obtained through the projection of the HF wave function onto the CI equation. According to the Slater-Condon rules and Brillouin's theorem, [44;46;47] only the doubly excited SDs couple directly to the HF state in the case of a canonical molecular orbital (CMO) basis. Therefore, the ground state energy is given by

$$E_0 = E_{\text{HF}} + E_{\text{corr}} \ , \tag{2.12}$$

$$E_0 = E_{\text{HF}} + \sum_{i>j,a>b} C_{ij}^{ab} \langle \Phi_{\text{HF}} | \hat{H} | \Phi_{ij}^{ab} \rangle \ , \tag{2.13}$$

$$E_0 = E_{\text{HF}} + \sum_{i>j,a>b} C_{ij}^{ab} (\phi_i \phi_a || \phi_j \phi_b) \ , \tag{2.14}$$

where E_{corr} stands for the correlation energy and $(\phi_i \phi_a || \phi_j \phi_b)$ being the antisymmetrized two-electron MO integral in the usual Mulliken notation. [44] Inclusion of all possible excitations in Equation (2.6) results in the full CI wave function describing the exact solution of the nonrelativistic Schrödinger equation. However, only the smallest systems with a few electrons are computable with this method due to the enormous number of required SDs. Consequently, truncations in the excitation operator $\hat{T}$ are necessary. These truncations lead to a neglect of direct and indirect couplings between different orders of SDs. As a consequence, truncated CI methods are not size extensive. Furthermore, the hierarchy of truncated CI methods tends to have a slow convergence to the full CI limit, for what reason other wave function methods regularly outperform the CI approximations.

The CI with Singles (CIS) approximations includes only the singly excited SDs and the reference state. [48] Referring to Equation (2.13), there will be no improvement over the HF result for the ground state in a CMO basis, but higher solutions of the eigenvalue

problem lead to approximated excited states. The CIS matrix elements have the form

$$\langle \Phi_i^a | \hat{H} | \Phi_i^a \rangle = (E_{\text{HF}} + \epsilon_a - \epsilon_i) + (\phi_i \phi_a | \phi_i \phi_a) - (\phi_i \phi_i | \phi_a \phi_a) \; , \qquad (2.15)$$

$$\langle \Phi_j^b | \hat{H} | \Phi_i^a \rangle = (\phi_i \phi_a | \phi_j \phi_b) - (\phi_i \phi_j | \phi_a \phi_b) \; , \qquad (2.16)$$

where ϵ_i and ϵ_a denote the occupied and virtual orbital energies, respectively. The diagonalization of this CIS Hamiltonian matrix results in the HF ground state and excited state energies together with the expansion coefficients matrix **C**. As the applied MOs are optimized for the HF ground state, the resulting CIS excitation energies are generally too high. Typical errors are larger than 1 eV.[48] Additionally, electron correlation and orbital relaxation effects are neglected, which are required for a quantitative description of excited states. However, the qualitative description of single electron excitations is often reasonable, so that the CIS solution can be applied as a reference for high-level calculations.

2.1.2. Coupled-Cluster Theory

Coupled-cluster (CC) theory is one of the most widely used high level correlation methods in theoretical chemistry.[45;49] Besides being size extensive, the hierarchy of truncated CC methods shows a faster convergence to the full CI limit than the truncated CI approximations and this at a comparable computational demand.

In the following, the basic formalism of CC theory for the calculation of electronic ground states as well as the extensions to excited states are presented. The latter can be divided in two formalisms, the equation-of-motion and response methods.

2.1.2.1. Ground State Coupled-Cluster Theory

Instead of the linear combination of SDs in CI theory, the CC wave function is expressed in a product form of spin orbital excitation operators of different rank

$$|\Psi_{\text{CC}}\rangle = \exp(\hat{T}) |\Phi_{\text{HF}}\rangle \; , \qquad (2.17)$$

which is known as the CC exponential ansatz. Equation (2.17) can be rewritten by expanding the exponential and collecting terms of the same order:

$$|\Psi_{\mathrm{CC}}\rangle = \sum_i \hat{C}_i |\Phi_{\mathrm{HF}}\rangle \qquad \text{with} \quad \hat{C}_0 = \hat{1} \tag{2.18}$$

$$\hat{C}_1 = \hat{T}_1 = \sum_{i,a} t_i^a \hat{\tau}_i^a$$

$$\hat{C}_2 = \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 = \sum_{i>j,a>b} t_{ij}^{ab} \hat{\tau}_{ij}^{ab} + \frac{1}{2}\sum_{i,a}\sum_{j,b} t_i^a t_j^b \hat{\tau}_i^a \hat{\tau}_j^b$$

$$\vdots$$

Here, $\{t_i^a, t_{ij}^{ab}, \dots\}$ are the expansion coefficients, called amplitudes in CC theory. For the CC wave function equation, connected and disconnected terms can be distinguished. The former are constructed by a single excitation operator $\hat{T}_k$, whereas the disconnected terms feature a multiplicative structure. This multiplicative form ensures the contribution of all excited SDs in the wave function even if the cluster operator is truncated. As a consequence, CC methods are size extensive and show a faster convergence in the truncation levels to the full CI limit than CI approximations. However, a variational procedure becomes intractable for all but the smallest systems. Therefore, a projection procedure is used to obtain the ground state energy and the CC amplitude equations. Commonly, a non-Hermitian similarity transformed Hamiltonian is applied for this purpose. The ground state energy is then defined by projection of the HF state onto the CC equation:

$$E_{\mathrm{CC}} = \langle\Phi_{\mathrm{HF}}|\exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_{\mathrm{HF}}\rangle \ , \tag{2.19}$$

$$E_{\mathrm{CC}} = E_{\mathrm{HF}} + \sum_{i>j,a>b} \left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right)(\phi_i\phi_a||\phi_j\phi_b) \ . \tag{2.20}$$

Only the single and double amplitudes enter directly in the energy expression. The amplitudes are determined by projection of the excited SDs onto the CC equation

$$\langle\Phi_{\mathrm{HF}}|\hat{\tau}_\mu^\dagger \exp(-\hat{T})\hat{H}\exp(\hat{T})|\Phi_{\mathrm{HF}}\rangle = 0 \ , \tag{2.21}$$

where $\hat{\tau}_\mu^\dagger$ denotes an excitation of arbitrary rank. The procedure results in a set of non-linear equations with respect to the amplitudes which have to be solved iteratively.

Because of the structure of Equation (2.21) also higher order excitations influence the CC energy since all amplitudes are coupled.

Similar to CI theory, the cluster operator $\hat{T}$ has to be truncated to obtain a practicable model. The CC singles and doubles (CCSD) is the common approximation, where only the $\hat{T}_1$ and $\hat{T}_2$ operators are considered.[50] CCSD considers the important single and double amplitudes and scales to the power of six with system size ($\mathcal{O}(N^6)$). An approximate model of CCSD is referred to as CC2.[51] Here, the computational demanding N^6-terms of the double amplitude equations ($\hat{\tau}_\mu$ of Equation (2.21) is a two-electron excitation operator) are neglected, but the singles equations are retained. As a consequence, the CC2 model scales to $\mathcal{O}(N^5)$, but yields less accurate energies in comparison to the CCSD model. In contrast, the additional incorporation of connected triple excitations would lead to a scaling of $\mathcal{O}(N^8)$, making the approximation feasible only for small systems. However, it was ascertained that triple excitations have to be considered for a highly accurate description of the electronic structure. Therefore, various approximations have been developed to include the triple amplitudes. A successful approach has been to treat the triple amplitudes by perturbation theory as in the CCSD(T) method,[52] which often is denoted as the golden standard in quantum chemistry.

2.1.2.2. Equation-of-Motion Coupled-Cluster Theory

So far the discussion of CC theory was restricted to the calculation of the ground state energy. The equation-of-motion coupled-cluster (EOM-CC) theory represents a method for the calculation of excited states.[39;45] The ansatz is to set a CI-type parametrization on top of a CC ground state wave function:

$$|\Psi_{\text{EOM}}) = \hat{R}\,|\Psi_{\text{CC}}\rangle \qquad \text{with} \quad \hat{R} = \sum_\mu C_\mu \hat{\tau}_\mu \ . \tag{2.22}$$

The operator $\hat{R}$ covers all possible excitations, but also includes the identity operator ($\hat{\tau}_0 = \hat{1}$). Therefore, the method can be regarded as CI theory employing correlated excited configurations as basis expansion.[53] This correlated basis has to be defined individually for bra and ket states in order to obtain orthonormality for the excited

states:

$$|\mu) = \exp(\hat{T})\hat{\tau}_\mu |\Phi_{\text{HF}}\rangle \ , \tag{2.23}$$

$$(\nu| = \langle\Phi_{\text{HF}}| \hat{\tau}_\nu^\dagger \exp(-\hat{T}) \ . \tag{2.24}$$

Here, biorthonormality for the bra and ket configurations is achieved from the orthonormality of the SDs $(\mu|\nu) = \delta_{\mu\nu}$. Accordingly, the expansion coefficients C_μ will also differ for the two sets of the same excitation operator $\hat{\tau}_\mu$. The variation principle can now be applied to a pseudo-expectation value of the electronic energy. Minimization of this expectation value with respect to the bra and ket expansion coefficients results in an unsymmetric eigenvalue problem:

$$\mathbf{HC} = E\mathbf{C} \ , \tag{2.25}$$

$$\bar{\mathbf{C}}^{\mathbf{T}}\mathbf{H} = \bar{\mathbf{C}}^{\mathbf{T}}E \ . \tag{2.26}$$

Here, $\bar{\mathbf{C}}$ and $\mathbf{C}$ are column vectors containing the coefficients for the bra and ket state, respectively. The elements of the unsymmetric Hamiltonian matrix $\mathbf{H}$ are given by

$$H_{\mu\nu} = \langle\Phi_{\text{HF}}|\hat{\tau}_\mu^\dagger \exp(-\hat{T})\hat{H} \exp(\hat{T})\hat{\tau}_\nu|\Phi_{\text{HF}}\rangle \ . \tag{2.27}$$

Although the matrix $\mathbf{H}$ is unsymmetric, it features a special structure. The elements of the first column represent the projected CC ground state equations (2.19) and (2.21). The Hamiltonian matrix can then be expressed in the form

$$\mathbf{H} = \begin{pmatrix} 0 & \boldsymbol{\eta}^{\mathbf{T}} \\ \mathbf{0} & \mathbf{A} \end{pmatrix} + E_{\text{CC}}\mathbf{1} \ , \tag{2.28}$$

where $\boldsymbol{\eta}$ is a column vector with the elements

$$\eta_\mu = \langle\Phi_{\text{HF}}| \exp(-\hat{T})\hat{H} \exp(\hat{T})\hat{\tau}_\mu|\Phi_{\text{HF}}\rangle \ , \tag{2.29}$$

and $\mathbf{A}$ is the CC Jacobian matrix with the elements

$$A_{\mu\nu} = \langle\Phi_{\text{HF}}|\hat{\tau}_\mu^\dagger \exp(-\hat{T})\left[\hat{H}, \hat{\tau}_\nu\right] \exp(\hat{T})|\Phi_{\text{HF}}\rangle \ , \tag{2.30}$$

$$A_{\mu\nu} = \langle\Phi_{\text{HF}}|\hat{\tau}_\mu^\dagger \exp(-\hat{T})\hat{H} \exp(\hat{T})\hat{\tau}_\nu|\Phi_{\text{HF}}\rangle - \delta_{\mu\nu}E_{\text{CC}} \ . \tag{2.31}$$

Diagonalization of the unsymmetric Hamiltonian matrix provides the excited state energies and EOM-CC wave functions. If only excitation energies are required, it is sufficient to diagonalize the CC Jacobian matrix as apparent from Equation (2.31).

Analogous to ground state CC theory the excitation space has to be restricted in order to obtain a computationally feasible model. In the EOM-CCSD approximation only the single and double spin orbital excitation operators are considered for the cluster operator $\hat{T}$ as well as for the CI-type operator $\hat{R}$. This approach gives reasonable results for single electron excitations. Various applications, most notably on small organic molecules, show that EOM-CCSD holds an accuracy of about 0.2 eV.[31] Furthermore, this method scales with the same dependence on the system size as its related ground state method ($\mathcal{O}(N^6)$). Similarly, the computationally less demanding CC2 model scales within the EOM formalism also with $\mathcal{O}(N^5)$. As previously mentioned, the approximated double amplitude equations lead to a loss of accuracy in terms of the energy for the ground state calculations. However, the excitation energies of singly excited states are correct to the same order for both EOM-CC2 and EOM-CCSD,[32] since the important singles are treated in the same way. This results in excitation energies of comparable accuracy.[31]

2.1.2.3. Linear Response Coupled-Cluster Theory

Response theory represents a different approach to access electronically excited states on the basis of a ground state calculation.[54] Here, the response of the wave function to a time-dependent perturbation is analyzed. In the Fourier-expanded form, the perturbation operator is given by:

$$\hat{V}^{\text{PT}} = \sum_{\omega} \exp(-i\omega t)\epsilon(\omega)\hat{B} \;, \tag{2.32}$$

with $\hat{B}$ being a Hermitian operator describing the perturbation, ω the frequency and $\epsilon(\omega)$ a strength parameter. In spectroscopy the perturbation arises from the exposure of the molecular system to light, so that the operator $\hat{B}$ is associated with the electric or magnetic field of the latter. According to this perturbation, the expectation value of an observable A can be expanded in a series of orders of the strength parameter:

$$\langle\Psi_0(t)|\hat{A}|\Psi_0(t)\rangle = \langle\Psi_0|\hat{A}|\Psi_0\rangle + \sum_{\omega} \epsilon(\omega)\langle\langle\hat{A},\hat{B}\rangle\rangle_{\omega} \exp(-i\omega t) \;. \tag{2.33}$$

Here, the expansion is terminated after the linear term incorporating the linear response (LR) function $\langle\langle \hat{A}, \hat{B} \rangle\rangle_\omega$. Higher order terms are negligible for weak interactions between the system and the radiative perturbation, which is usually the case for standard UV/vis absorption experiments relevant to this thesis. The LR function then contains all information how the property A changes according to the interaction described by operator $\hat{B}$. For exact states, $\langle\langle \hat{A}, \hat{B} \rangle\rangle_\omega$ can be written in terms of the stationary eigenstates of the Hamiltonian:[55]

$$\langle\langle \hat{A}, \hat{B} \rangle\rangle_\omega = \sum_I \left[\frac{\langle \Psi_0 | \hat{A} | \Psi_I \rangle \langle \Psi_I | \hat{B} | \Psi_0 \rangle}{\omega - \omega_I} + \frac{\langle \Psi_0 | \hat{B} | \Psi_I \rangle \langle \Psi_I | \hat{A} | \Psi_0 \rangle}{\omega_I - \omega} \right] , \qquad (2.34)$$

where ω_I denotes the energy difference $E_I - E_0$ between an excited state I and the ground state 0. The first term describes the excitation process at positive frequencies and the second term the corresponding deexcitation. Obviously, this LR function has poles at frequencies which match the transition energies of the molecular system. Furthermore, the transition strength is determined by the residuals at the poles.

The connection between the LR equation for exact states (2.34) and CC theory was elaborated by Koch and Jørgensen in 1990.[56] The poles are determined by diagonalization of the CC Jacobian matrix:

$$\mathbf{A}\tilde{\mathbf{C}} = \omega\tilde{\mathbf{C}} , \qquad (2.35)$$

where $\mathbf{A}$ stands for the same as in the EOM-CC fomalism (Equation (2.30)) and $\tilde{\mathbf{C}}$ contains the expansion coefficients while excluding the CC ground state. Therefore, both formalisms result in the same excitation energies for a molecular system and are in this respect equivalent. Similarly, the same truncations are applicable for the CC model leading to the same computational effort and accuracies.

2.2. Density Functional Theory

Wave function based methods provide the possibility for a systematical improvement by extending their truncation levels. However, the electronic wave function depends on three spatial coordinates and one spin coordinate for each electron. For large molecules, the number of required parameters becomes immense for an accurate description. Den-

sity functional theory (DFT) represents an alternative concept to determine the physical properties of molecular systems without explicitly solving the Schrödinger equation.[57] In principle, the only required quantity is the ground state density, which depends merely on three spatial coordinates.

In the first part of this section, the foundations of static DFT is presented as well as its practical realization, which leads to an introduction of molecular orbitals. As a consequence, the conventional DFT scheme is again a full dimensional problem. The second subsection covers the dynamical counterpart, which provides a method to calculate excitation spectra within DFT.

2.2.1. Ground State Density Functional Theory

The Hohenberg-Kohn theorems build the foundations of static DFT.[58] The existence theorem states that it is impossible to produce the identical ground state density ρ_0 from two different external potentials v_{ext}. Consequently, there is an one-to-one correspondence between these two quantities:

$$\rho_0(\mathbf{r}) \leftrightarrow v_{\text{ext}}(\mathbf{r}) \ . \tag{2.36}$$

Thus, the external potential can be reproduced by the knowledge of the ground state density. Likewise, this means that also the Hamiltonian and its eigenstates become functionals of the ground state density. The logical chain can be represented as

$$\rho_0 \rightarrow v_{\text{ext}} \rightarrow \hat{H} \rightarrow \{\Psi_I\} \ . \tag{2.37}$$

Applying the Rayleigh-Ritz principle to the total energy leads to the variational theorem of Hohenberg and Kohn.[58] This second theorem states that the exact ground state density yields an upper bound for the total energy:

$$E[\rho] > E_0 \qquad \text{for } \rho \neq \rho_0 \tag{2.38}$$

$$E[\rho] = E_0 \qquad \text{for } \rho = \rho_0 \ . \tag{2.39}$$

Therefore, a minimization of the energy with respect to the ground state density would avoid the problem of explicitly solving the Schrödinger equation. However, complications arise in the calculations of the kinetic energy T_{e} and the electron-electron interaction V_{ee}.

The functionals are in fact universal for any *N*-electron system, but their exact form is unknown and any direct approximations are still unsatisfying.[59] The Kohn-Sham (KS) formalism constitutes an elegant way to bypass this problem.[60] The starting point is the definition of a fictive system of non-interacting (ni) particles, which reproduces the exact ground state density of the real system. Such a system can be represented with a single SD, for which the single-particle orbitals ϕ_i satisfy the equation:

$$\left(-\frac{\nabla^2}{2} + v_{\text{ext(ni)}}\right)\phi_i = \epsilon_i\phi_i \ . \tag{2.40}$$

Here, ∇ is the nabla operator and ϵ_i denotes the molecular orbital energy. Access to the real system is achieved by rewriting the total energy functional as

$$E[\rho] = T_{\text{e}}[\rho] + V_{\text{ext}}[\rho] + V_{\text{ee}}[\rho] \ , \tag{2.41}$$

$$E[\rho] = T_{\text{e(ni)}}[\rho] + V_{\text{ext}}[\rho] + V_{\text{Coul}}[\rho] + E_{\text{xc}}[\rho] \ , \tag{2.42}$$

where V_{Coul} stands for the classical Coulomb energy and E_{xc} is the so called exchange-correlation (xc) functional:

$$E_{\text{xc}}[\rho] = T_{\text{e}}[\rho] - T_{\text{e(ni)}}[\rho] + V_{\text{ee}}[\rho] - V_{\text{Coul}}[\rho] \ . \tag{2.43}$$

Besides the exchange interaction and correlation effects, the xc functional also includes the difference between the kinetic energy of the real system and the non-interacting KS system. The xc functional is the only unknown in this formalism, which, however, makes the smallest contribution to the total energy. By minimizing the total energy functional, it is straightforward to show that the following relationship between the KS system and the real system holds:

$$v_{\text{ext(ni)}}[\rho] = v_{\text{ext}}[\rho] + v_{\text{Coul}}[\rho] + v_{\text{xc}}[\rho] \ . \tag{2.44}$$

Equation (2.40) combined with Equation (2.44) constitute the KS equations. The solution is quite similar to HF theory, but instead of the Fock operator the KS operator is introduced.

As mentioned before, the xc functional is the only unknown of the KS formalism and consequently approximations have to be introduced. One approach is to calculate the

xc energy density of an uniform electron gas for the local density on each point **r** of a grid (local density approximation − LDA). Integration then yields the total xc energy. As the density can become a strongly varying quantity in molecular systems, correction terms including the gradient of the density were added to LDA (generalized gradient approximation − GGA). A further approach are hybrid functionals, which split the exchange energy in two parts. A fraction of the exact HF exchange E_x^{HF} is used and the remainder is taken from LDA/GGA-DFT exchange energy. The B3LYP functional is a popular example of this class of xc functionals and has the form[61;62]

$$E_{xc}^{B3LYP} = (1-a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B88} + (1-c)E_c^{LDA} + cE_c^{LYP} \;, \qquad (2.45)$$

where ΔE_x^{B88} stands for the exchange energy correction term of the B88-GGA functional and E_c^{LYP} is a GGA correlation functional.[63;64] The empirical parameters a, b and c were fitted to experimental data and have values of 0.2, 0.72 and 0.81, respectively.

All conventional DFT functionals show a wrong asymptotic behaviour for the exchange interaction. This has spurred the development of range-separated hybrid functionals, which tackle this problem by splitting the r_{12}^{-1}-term in the following form:

$$\frac{1}{r_{12}} = \frac{1 - [\alpha + \beta\mathrm{erf}(\mu r_{12})]}{r_{12}} + \frac{\alpha + \beta\mathrm{erf}(\mu r_{12})}{r_{12}} \;, \qquad (2.46)$$

where α, β and μ are fitted parameters, which have to satisfy the conditions

$$0 \leq \alpha + \beta \leq 1 \;, \qquad (2.47)$$

$$0 \leq \alpha \leq 1 \;, \qquad (2.48)$$

$$0 \leq \beta \leq 1 \;. \qquad (2.49)$$

The short-range part is then dominated by the DFT exchange, whereas the long-range part is controlled by the exact HF exchange. This approach is commonly referred to as Coulomb-attenuating method (CAM). A popular functional of this type is denoted as CAM-B3LYP,[65] using the same correlation expression as the B3LYP funcional. The exchange part is built from the HF exchange and the B88-GGA functional, which are modified with Equation (2.46). The corresponding parameters are chosen as $\alpha = 0.19$, $\beta = 0.65$ and $\mu = 0.33$.

The application of range-separated functionals is required within this thesis, because the wrong asymptotic behaviour of conventional DFT functionals causes a systematic

underestimation of Rydberg and charge-transfer (CT) excitation energies.[66;67] These energies can be calculated by the dynamical counterpart of static DFT, described in the following section.

2.2.2. Time-Dependent Density Functional Theory

In principle, the ground state density also includes the information about the excited states of a system. This can be derived from the logical chain in Equation (2.37). The ground state density determines the Hamiltonian through the external potential that in turn also comprises all its eigenstates. However, there has been no efficient method to extract the excitation energies from static DFT until now. Time-dependent density functional theory (TD-DFT) is the dynamical counterpart of stationary DFT, which allows to evaluate time-dependent properties of a system, including the calculation of excitation spectra in an efficient manner.[68]

The Runge-Gross (RG) theorem can be regarded as the dynamical analogue to the existence theorem of Hohenberg and Kohn in static DFT.[69] The former states that when starting from the same initial state $\Psi(t_0)$, two different time-dependent potentials always cause different time-dependent densities. Therefore, an one-to-one correspondence between these two quantities is given:

$$\rho(\mathbf{r}, t) \leftrightarrow v_{\text{ext}}(\mathbf{r}, t) \ . \tag{2.50}$$

The RG theorem only holds for time-dependent potentials that are Taylor-expandable around the initial time. This limitation is mostly satisfied in practical applications, especially for the calculation of excitation spectra, which deals with potentials arising from monochromatic light.

The van Leeuwen theorem provides the formal foundation to apply the KS formalism in TD-DFT.[70] This theorem states that the same time-dependent density of a system in an initial state $\Psi(t_0)$ with the electron-electron interaction $\hat{V}_{\text{ee}}$ and external potential v_{ext} can be completely reproduced by another system in an initial state $\Psi'(t_0)$ with a different electron-electron interaction $\hat{V}'_{\text{ee}}$ and a unique external potential v'_{ext}. Thus, a fictive system with absent electron-electron interaction (KS system) is appropriate to describe the exact time-dependent density.

In the following, the system is assumed to be in its electronic ground state before

a time-dependent potential is switched on at a certain time t_0. The KS orbitals then satisfy the time-dependent Schrödinger equation

$$\left[-\frac{\nabla^2}{2} + v_{\text{ext(ni)}}[\rho](\mathbf{r}, t) \right] \phi_i(\mathbf{r}, t) = i\frac{\partial}{\partial t}\phi_i(\mathbf{r}, t) \ , \tag{2.51}$$

where the effective potential is related to the real system by

$$v_{\text{ext(ni)}}[\rho](\mathbf{r}, t) = v_{\text{ext}}[\rho](\mathbf{r}, t) + v_{\text{Coul}}[\rho](\mathbf{r}, t) + v_{\text{xc}}[\rho](\mathbf{r}, t) \ . \tag{2.52}$$

Here, $v_{\text{Coul}}(\mathbf{r}, t)$ denotes the classical Coulomb potential, which only depends on the instantaneous density. $v_{\text{xc}}(\mathbf{r}, t)$ stands for the time-dependent xc potential and is the only unknown in this formalism. An ubiquitous approximation is to evaluate the time-dependent xc potential at the instantaneous densities within the ground state xc potential

$$v_{\text{xc}}^{\text{A}}(\mathbf{r}, t) \approx v_{\text{xc}}^0[\rho_0](\mathbf{r})\big|_{\rho_0(\mathbf{r}) \to \rho(\mathbf{r},t)} \ , \tag{2.53}$$

and thus, ignores memory effects during the time propagation. The concept is called adiabatic approximation and used throughout this section.

For the calculation of excitation spectra, Equation (2.51) is rarely solved explicitly. The essential task is to consider the response of the system to a weak perturbation. Therefore, linear response theory is applied,[68] as already introduced in Section 2.1.2.3. In the context of TD-DFT, the formalism results in the frequency-dependent density response function for the KS system:

$$\rho_1(\mathbf{r}, \omega) = \int \mathrm{d}^3 r' \chi_{\text{ni}}(\mathbf{r}, \mathbf{r}', \omega) v_{\text{ext(ni)}}^{(1)}(\mathbf{r}', \omega) \ , \tag{2.54}$$

with $\chi_{\text{ni}}(\mathbf{r}, \mathbf{r}', \omega)$ being the KS density-density response function

$$\chi_{\text{ni}}(\mathbf{r}, \mathbf{r}', \omega) = \lim_{\eta \to 0^+} \sum_{i,j} (n_j - n_i) \frac{\phi_i(\mathbf{r})\phi_j^*(\mathbf{r})\phi_i^*(\mathbf{r}')\phi_j(\mathbf{r}')}{\omega - \omega_{ij} + i\eta} \ . \tag{2.55}$$

Here, n_i and n_j stand for the occupation numbers of the KS orbitals i and j in the ground state, respectively and ω_{ij} denotes the difference between their corresponding orbital energies. The summation runs over all occupied and virtual KS orbitals. The

first-order effective perturbation is given by

$$v^{(1)}_{\mathrm{ext(ni)}}(\mathbf{r}', \omega) = v^{(1)}_{\mathrm{ext}}(\mathbf{r}', \omega) + \int \mathrm{d}^3 r'' \left[\frac{1}{|\mathbf{r}' - \mathbf{r}''|} + f^{\mathrm{A}}_{\mathrm{xc}}(\mathbf{r}', \mathbf{r}'') \right] \rho_1(\mathbf{r}'', \omega) \; . \qquad (2.56)$$

In Equation (2.56) the xc kernel $f^{\mathrm{A}}_{\mathrm{xc}}(\mathbf{r}', \mathbf{r}'')$ was introduced in the adiabatic approximation, which is the functional derivative of the xc potential with respect to the density

$$f^{\mathrm{A}}_{\mathrm{xc}}(\mathbf{r}', \mathbf{r}'') = \frac{\delta v^0_{\mathrm{xc}}[\rho_0](\mathbf{r}')}{\delta \rho_0(\mathbf{r}'')} = \frac{\delta^2 E_{\mathrm{xc}[\rho_0]}}{\delta \rho_0(\mathbf{r}') \delta \rho_0(\mathbf{r}'')} \; . \qquad (2.57)$$

Electronic excitations can be regarded as characteristic eigenmodes of an interacting system, which are induced by an external perturbation. Nevertheless, those eigenmodes exist already in the absence of such a perturbation ($v_{\mathrm{ext}}[\rho](\mathbf{r}, t) = 0$) and there are non-trivial solutions of Equation (2.54) at the exact excitation energies Ω_n. Casida has recast this problem,[71] resulting in the expression

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \Omega \begin{pmatrix} -\mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \; . \qquad (2.58)$$

The column vectors $\mathbf{X}$ and $\mathbf{Y}$ contain the coefficients to describe the transition. The elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$A_{ia\sigma,jb\tau} = \delta_{ij}\delta_{ab}\delta_{\sigma\tau}\omega_{ia\sigma} + K_{ia\sigma,jb\tau} \; , \qquad (2.59)$$

$$B_{ia\sigma,jb\tau} = K_{ia\sigma,jb\tau} \; , \qquad (2.60)$$

$$K_{ia\sigma,jb\tau} = \int \mathrm{d}^3 r \int \mathrm{d}^3 r' \phi^*_{i\sigma}(\mathbf{r})\phi_{a\sigma}(\mathbf{r}) \left[\frac{1}{|\mathbf{r} - \mathbf{r}'|} + f^{\mathrm{A}}_{\mathrm{xc},\sigma\tau}(\mathbf{r}, \mathbf{r}') \right] \phi^*_{j\tau}(\mathbf{r}')\phi_{b\tau}(\mathbf{r}') \; , \qquad (2.61)$$

where the usual notation for occupied (i, j) and virtual (a, b) orbitals is used. σ and τ denote the orbital spin index.

The Casida equations result in pairs of excitation energies with positive and negative values corresponding to excitation and deexcitation processes, respectively. These two types of transitions are decoupled within the Tamm-Dancoff approximation (TDA) by setting the $\mathbf{B}$ matrix to zero.[72] In this case, the Casida formalism results in a simple

eigenvalue problem of the form

$$\mathbf{AX} = \Omega\mathbf{X} \ . \tag{2.62}$$

The Casida equations are closely related to the time-dependent framework of HF theory, also called random-phase approximation (RPA).[48] The latter is obtained by interchanging the xc kernel with the exact exchange operator. If the TDA approximation is additionally applied, then the CIS equations are obtained. However, the major difference constitutes the application of already correlated KS orbitals from DFT ground state calculations instead of uncorrelated HF orbitals. The calculated excitation energies are dominated by the diagonal elements featuring the MO energy differences, which has a significant effect on the accuracy. TD-DFT benchmark calculations for valence excited states of organic molecules have shown an accuracy of about 0.3 eV for global hybrid functionals such as B3LYP.[73;74] This is a large improvement compared to RPA or CIS excitation energies, for which the errors often exceed 1 eV.[48]

2.3. Molecular Mechanics

The treatment and simulation of large biomolecules, for example proteins containing thousands of atoms, exceed the current computational resources for the use of quantum chemical methods. An alternative approach constitutes molecular mechanics (MM), which allows the efficient energy calculation of such systems. Here, the explicit motion of the electrons is neglected and the total energy is classically expressed as a function of the nuclear positions, the so-called force fields.[75] Generally, the total energy can be divided into two components. The bonded interactions incorporate bond stretching, angle bending as well as torsional motions. The non-bonded part includes electrostatic interactions and van der Waals interactions. As the force fields are empirically defined, their functional form and parametrization differ from each other. Moreover, additional terms can occur depending on the purpose the force field is designed for. For example,

the functional form of the AMBER force field is given by[76]

$$V^{\text{AMBER}} = \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \sum_n \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] +$$

$$\sum_{I<J} \frac{Q_I Q_J}{4\pi\epsilon_0 R_{IJ}} + \sum_{I<J} \left(\frac{A_{IJ}}{R_{IJ}^{12}} - \frac{B_{IJ}}{R_{IJ}^6} \right) + \sum_{\text{H-bonds}} \left(\frac{C_{IJ}}{R_{IJ}^{12}} - \frac{D_{IJ}}{R_{IJ}^{10}} \right) . \tag{2.63}$$

The first two terms are the harmonic potentials for bond stretching and angle bending with the force constants k_r and k_θ, respectively. The reference values of the bond lengths and angles are given by the parameters r_0 and θ_0. The third term in Equation (2.63) accounts for the rotation around bonds. The parameter n denotes the multiplicity of a torsional motion and γ is a phase factor. V_n determines the form of the potential, but it cannot be interpreted directly as barrier height of a torsional motion, because there are also 1,4-non-bonded interactions, which influence the height of the barriers. The last three terms account for the non-bonded interactions. The distance between two nuclei I and J are denoted with R_{IJ}. The electrostatic energy (fourth term) is determined by the Coulomb energy between atomic partial charges Q_I and Q_J and the van der Waals interactions (fifth term) are incorporated using a Lennard-Jones potential. A_{IJ} and B_{IJ} are the corresponding Lennard-Jones parameter. As the AMBER force field is designed for the simulation of proteins and nucleic acids, special attention has been given to hydrogen bonds. The sixth summation runs over all non-bonded atom pairs, for which a hydrogen atom acts as an acceptor and a heteroatom being the donor. The functional form was chosen to be a Mie-(12,10)-potential with parameters C_{IJ} and D_{IJ}.

Computational resources limit the system size that can be treated with MM. Periodic boundary conditions can be applied to reduce the system size and simultaneously exclude boundary effects.[75] In such an approach, the system is placed in a box and copied in all dimensions. Assuming a rectangular box for simplicity, the atomic positions of a real atom and all its images can be expressed as

$$\mathbf{R}_I = \begin{pmatrix} (x_I + n_x)a \\ (y_I + n_y)b \\ (z_I + n_z)c \end{pmatrix} \quad \text{with} \quad \{n_x, n_y, n_z\} \in \mathbb{Z} . \tag{2.64}$$

Here, x_I, y_I and z_I stand for the fractional coordinates of an atom I and the parameters a, b and c are the cell constants of the box. The real system corresponds to the tuple

$\{0, 0, 0\}$, whereas all other combinations describe the images. If an atom leaves the real system during the simulation, then the procedure ensures that one of its images enters the box.

The handling of the potential energy calculations for such an extended system is easily achieved for the bonded and non-bonded van der Waals interactions, since these are of local and short-range character, respectively. The latter can be treated by inclusion of a cutoff for atom pairs to be considered. The situation differs for the electrostatic interactions, which are decaying only with a R^{-1}-dependence. This leads to a slowly converging summation which is in fact conditionally convergent. The standard procedure to circumvent this problem is applying an Ewald summation,[75;77] for which the R^{-1}-term is split up into a short-range (sr) and a long-range (lr) part by the identity:

$$\frac{1}{R} = \frac{\mathrm{erfc}(\alpha|\mathbf{R}_{IJ} + \mathbf{n}|)}{R} + \frac{1 - \mathrm{erfc}(\alpha|\mathbf{R}_{IJ} + \mathbf{n}|)}{R} \qquad (2.65)$$
$$\text{with} \quad \mathrm{erfc}(x) = \frac{2}{\pi} \int_x^\infty \exp(-t^2)\mathrm{d}t \ .$$

The complementary error function (erfc) ensures that a neutralizing Gaussian charge distribution surrounds each atomic partial charge for the first term and a counteracting charge distribution is introduced for the second term. The parameter α controls the shape of the Gaussian distribution. The electrostatic interactions are then calculated by a summation of the short-range part in the real space and the long-range part in the reciprocal space. Assuming a cubic box with cell length L, this results in the following summations:

$$V_{\mathrm{sr}}^{\mathrm{real}} = \frac{1}{2} \sum_I \sum_{J'} \sum_{|\mathbf{n}|} \frac{Q_I Q_J}{4\pi\epsilon_0} \frac{\mathrm{erfc}(\alpha|\mathbf{R}_{IJ} + \mathbf{n}|)}{|\mathbf{R}_{IJ} + \mathbf{n}|} \ , \qquad (2.66)$$

$$V_{\mathrm{lr}}^{\mathrm{reciprocal}} = \frac{1}{2} \sum_{k \neq 0} \sum_I \sum_J \frac{1}{\pi L^3} \frac{Q_I Q_J}{4\pi\epsilon_0} \exp\left(-\frac{k^2}{4\alpha^2}\right) \cos(\mathbf{k} \cdot \mathbf{R}_{IJ}) \ , \qquad (2.67)$$

$$V^{\mathrm{self}} = -\frac{\alpha}{\sqrt{\pi}} \sum_I \frac{Q_I^2}{4\pi\epsilon_0} \ . \qquad (2.68)$$

The prime in the real sum indicates that for $|\mathbf{n}| = 0$ the terms with $I = J$ are excluded and $\mathbf{k}$ in Equation (2.67) stands for the reciprocal cell vectors. The term V^{self} just corrects for the self-interactions of the Gaussian charge distributions in the real sum. Within this procedure both summations are rapidly converging in comparison to the

original Coulomb summation, making the computation feasible for large systems.

2.4. Quantum Mechanics – Molecular Mechanics

The treatment of large systems such as biomolecules is intractable within quantum mechanical (QM) approaches due to the large number of electrons. Usually, the region of interest, where for example a reaction takes place, is much smaller. In the quantum mechanics − molecular mechanics (QM/MM) approach the system is divided into two regions. High level QM methods are applied to the region of interest and the remainder is treated at the MM level.[78;79] The total energy for an additive QM/MM scheme is given by

$$E_{\mathrm{QM/MM}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM-MM}} \ , \tag{2.69}$$

where E_{QM} and E_{MM} are the energies of the QM and MM part, respectively. $E_{\mathrm{QM-MM}}$ denotes the coupling between the two regions. The coupling energy contains bonded, van der Waals as well as electrostatic interactions

$$E_{\mathrm{QM-MM}} = E_{\mathrm{QM-MM}}^{\mathrm{bonded}} + E_{\mathrm{QM-MM}}^{\mathrm{vdW}} + E_{\mathrm{QM-MM}}^{\mathrm{el}} \ . \tag{2.70}$$

While the bonded and van der Waals terms are always treated at the MM level, there are several schemes for the electrostatic interactions.[79] In the mechanical embedding scheme the simple charge model of the MM method is also applied to the QM region. Such an embedding scheme suffers from several limitations. For example, the QM density is not directly polarized by the electrostatic environment, leading to a insufficient description of the important QM part. Moreover, the derivation of adequate parameters for the QM region can become cumbersome.

The most popular approach is the electrostatic embedding scheme. Here, the QM calculation is performed in the presence of the MM charge model. Therefore, the MM region polarizes the QM charge distribution. In the case of MM point charges, one-electron terms are incorporated into the QM Hamiltonian

$$\hat{H}_{\mathrm{QM-MM}}^{\mathrm{el}} = -\sum_{i,J} \frac{Q_J}{|\mathbf{r}_i - \mathbf{R}_J|} + \sum_{I,J} \frac{Q_I Q_J}{|\mathbf{R}_I - \mathbf{R}_J|} \ , \tag{2.71}$$

where the first summation runs over all electrons i of the QM region and atomic charges J of the MM region. The electrostatic interaction between all nuclei I of the QM region and the atomic charges of the MM region J (second term) can be treated classically. For the calculation of excitation energies within a QM/MM scheme, the coupling Hamiltonian is simplified. As the MM parameters are commonly derived by disregarding the electron state of a chromophore in the QM region, the pure MM coupling terms vanish. Here, the electrostatic interaction between the electrons and the MM atoms is the only term to be considered.

In the polarized embedding scheme, the QM charge distribution is also allowed to polarize the MM charges. Consequently, this embedding scheme has to be done self-consistently and a polarizable force field is required. Thus, the computational demand can become immense.

Special care is required if covalent bonds are cut in the QM−MM boundary region. Various schemes have been developed to deal with this situation. The link-atom method in combination with a charge-shift scheme is one of them.[80] Within this approach, a link-atom, usually a hydrogen atom, is placed along the cut covalent bond and included in the QM region to saturate the free valency. Moreover, as the charge at the MM atom involved in the covalent bond would lead to an overpolarisation of the QM density, the atomic charge is equally redistributed to its next neighbour MM atoms. Additional charges are then introduced close to them and along their bonds to compensate the created dipole moment. By this means, the charge distribution of the MM region is preserved, but the charge close to the link-atom is shifted away from the QM region.

Incremental EOM-CCSD

A considerable disadvantage for the application of correlated wave function methods is the steep scaling of the computation time with the system size, which only allows the treatment of small to medium sized molecules. Fragmentation techniques provide a method to circumvent this drawback.[81] Here, the system is subdivided into fragments, whose energies can be computed individually and combined afterwards to obtain the total energy. Incremental approaches for the calculation of the correlation energy of a system represent a specific variant of fragmentation techniques.[33] In such a procedure, the system is split into orbital groups based on a reference calculation and the energy is then expressed in a many-body expansion

$$E = \sum_I E_I + \sum_{I>J} \Delta E_{IJ} + \sum_{I>J>K} \Delta E_{IJK} + \cdots \ , \tag{3.1}$$

where

$$\Delta E_{IJ} = E_{IJ} - E_I - E_J \ , \tag{3.2}$$

$$\Delta E_{IJK} = E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK} - E_I - E_J - E_K \ , \tag{3.3}$$

E_I being the energy of fragment I, E_{IJ} the energy of the union of fragment I and J and so forth. The method has been applied successfully to estimate the ground state correlation energy for molecules, ionic crystals and semiconductors.[33–37] After a HF calculation followed by an orbital localization, any single reference correlation method could be applied. Here, the use of localized molecular orbitals (LMOs) is crucial for a fast convergence of the many-body expansion. However, the extension of this approximation

to excited states is not straightforward, since an appropriate reference is missing for the basis of the expansion in contrast to the ground state calculation with a localized HF basis.

In the following, an incremental approach for excitation energies is detailed, which was originally introduced by Mata and Stoll and extended within the scope of this thesis.[38] Afterwards, the benchmark set is presented and the results of different algorithms for setting up the fragments are discussed. The chapter concludes with the current state of the method and an outlook of future improvements is given.

3.1. Method

The following describes the methodology for the calculation of excitation energies within an incremental approach based on the procedure of Mata and Stoll.[38] Therein, a suitable basis for the many-body expansion is obtained by the calculation of CIS excited states from the HF ground state. After this, a natural transition orbital (NTO) transformation follows for the excited state of interest,[82] which results in hole-particle orbital pairs for the valence and virtual orbital space. Their contribution to the excitation is weighted by the singular values λ_i, arising from the decomposition of the CIS transition density matrix. For single electron excitations, there is typically one dominating orbital pair with $\lambda_i > 0.9$, so that the excitation can be qualitatively represented within one or at least only a few of such pairs. These are selected according to a threshold of $\lambda_{\text{thresh}}^{\text{NTO}} > 0.05$ for the singular values of the NTOs. Afterwards, the remaining valence orbitals are localized with the Pipek-Mezey method.[83] The resulting mixed occupied orbital space builds the basis for the expansion and eventually, the orbitals are grouped into fragments. Notably, there will be a main fragment M containing the NTOs and thus describing the dominant excitation space. Following this procedure, the excitation energy within a two-body (2B) approach can be expressed as[38]

$$\omega^{2B} = \omega^{1B} + \sum_{I \neq M} (E_{MI}^* - E_{MI}^0 - \omega^{1B}) \tag{3.4}$$

with

$$\omega^{1B} = E_M^* - E_M^0 \ . \tag{3.5}$$

The asterisk denotes the excited state and 0 stands for the electronic ground state. For the calculation of the energy terms the EOM-CCSD method is applied.[39] Here, the CIS transition vector is followed in order to converge to the desired excited state. However, the energy terms could be computed with any single reference method capable of providing excitation energies and states. The described scheme is depicted in Figure 3.1.

After the basis for the expansion has been established, the challenge for grouping the orbitals into fragments remains. Certainly, the selected NTOs should be part of the main fragment, because they describe the excitation space. However, adding localized molecular orbitals (LMOs) to the main fragment and constructing the other groups requires a reasonable criterion. In the original procedure, the distances between orbital charge centroids were employed.[38] This approach leads to fragments whereby the LMOs are in close proximity and account for short-range correlation effects. A problem with this criterion arises in the definition of the main fragment. As the NTOs feature a delocalized character, their charge centroids are not appropriate as fragmentation criterion, which is illustrated for the lowest $\pi \to \pi^*$ transition of naphthalene (see Figure 3.2).

The NTOs of the lowest $\pi \to \pi^*$ excitation of naphthalene are symmetrically delocalized over the whole molecule. This finding results in orbital charge centroids that
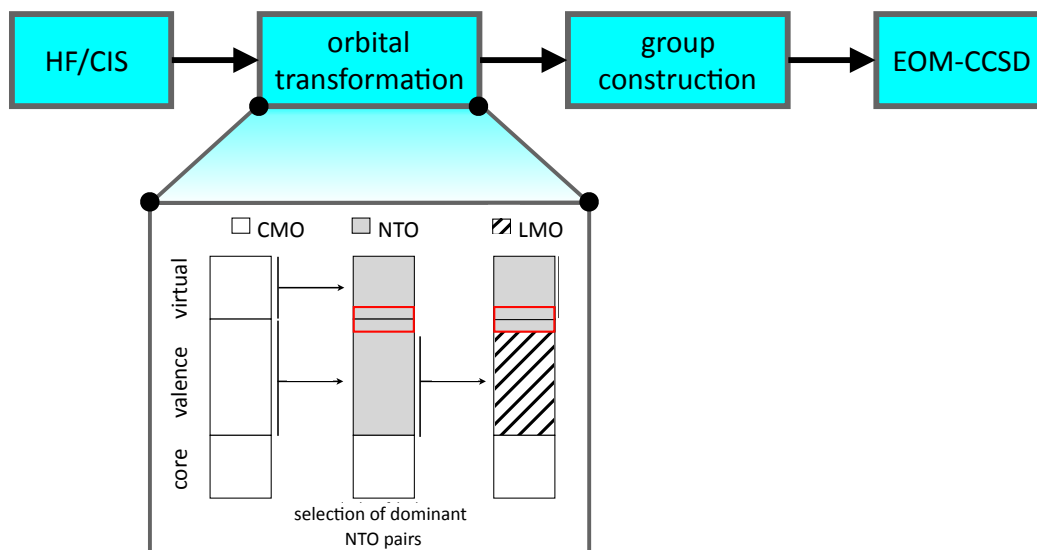
Figure 3.1.: Scheme of the incremental correlation approach for the calculation of excitation energies. CMO: canonical molecular orbitals, NTO: natural transition orbitals, LMO: localized molecular orbitals.
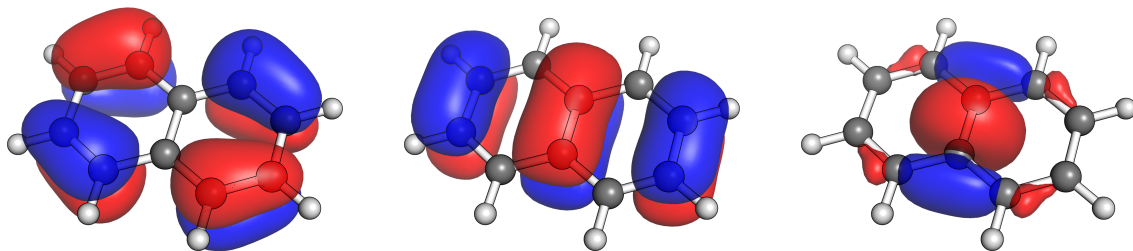
Figure 3.2.: Selected MOs for the first excited state of naphthalene after the orbital transformatioms illustrated in Figure 3.1. The orbital charge centroids nearly coincide for the three shown MOs. Left: NTO with a singular value of $\lambda = 0.49$. Middle: NTO with a singular value of $\lambda = 0.46$. Right: LMO of the central C$-$C bond.

are located in the center of the molecule. As a consequence, the central $\sigma_{\text{C}-\text{C}}$ orbital would be assigned to the main fragment. This LMO is not expected to be of major importance for the excitation, but rather the remaining π-type LMOs are required. These orbitals are strongly involved in the excitation due to their overlap with the hole/particle densities, resulting in significant correlation effects.

The described shortcoming required new criteria, in order to identify the LMOs with a significant influence on the excitation. The overlap between the hole (particle) and LMO densities is proposed to be such a criterion. For calculating this parameter, the orbital densities are first fitted to an auxilary basis set

$$\rho_i \approx \tilde{\rho}_i = \sum_\mu \tilde{C}_{\mu i} \tilde{\chi}_\mu \ , \tag{3.6}$$

where $\tilde{C}_{\mu i}$ and $\tilde{\chi}_\mu$ are the fitting coefficients and auxiliary basis functions, respectively. The hole (particle) densities are then approximated by the sum of the selected occupied (virtual) NTOs weighted with their singular values:

$$\rho_{\text{exc}}^{\text{h/p}} \approx \tilde{\rho}_{\text{exc}}^{\text{h/p}} = \sum_i^{\text{NTOs}} \lambda_i \tilde{\rho}_i \ . \tag{3.7}$$

In a second step, the density overlap is calculated by

$$S_{\text{exc},i}^{\text{h/p}} = \frac{\langle \tilde{\rho}_{\text{exc}}^{\text{h/p}} | \tilde{\rho}_i \rangle}{\sqrt{\langle \tilde{\rho}_{\text{exc}}^{\text{h/p}} | \tilde{\rho}_{\text{exc}}^{\text{h/p}} \rangle \langle \tilde{\rho}_i | \tilde{\rho}_i \rangle}} \ , \tag{3.8}$$

where the normalization factor ensures that the parameter is within the interval $S_{\text{exc},i}^{\text{h/p}} \in [0,1]$. Since the density overlap can be separately determined for the hole and particle of the excitation, both data can be considered for the construction of the groups. Here, it is sufficient to take the larger value $S_{\text{exc},i}^{\text{h}}$ or $S_{\text{exc},i}^{\text{p}}$. A threshold can be used to identify LMOs with a significant overlap to the excitation space. For example, in the case of the $\pi \to \pi^*$ transition of naphthalene shown in Figure 3.2, the remaining π-type LMOs are selected by $S_{\text{exc},i}^{\text{h/p}} > 0.25$, while all σ-LMOs feature values below this threshold.

In order to group the remaining LMOs in spatially contracted fragments, a less demanding procedure can be applied. Here, the orbital exchange interaction is chosen as the criterion:

$$\tilde{K}_{ij} = \frac{\langle \phi_i \phi_j | \phi_j \phi_i \rangle}{\sqrt{\langle \phi_i \phi_i | \phi_i \phi_i \rangle \langle \phi_j \phi_j | \phi_j \phi_j \rangle}} \ , \tag{3.9}$$

where the normalization factor again ensures that the value is in the range between 0 and 1. The exchange interaction decays exponentially for well separated particles.[84] Therefore, distant LMO pairs have a vanishing magnitude, whereas close lying LMOs can be identified and grouped together. A threshold of $\tilde{K}_{ij} > 0.015$ turned out to be reasonable to define those spatially close LMO pairs.

3.2. Benchmark Calculations

3.2.1. Benchmark Set

A set of 16 small to medium sized organic molecules was chosen to test the incremental approach and the different criteria. The corresponding structural formula in combination with their retained names are shown in Figure 3.3. This selection is a subset of Thiel's benchmark set for excitation energies,[31] which was extended by four molecules: acrolein, anthracene, guanine as well as a histidine with the carboxyl and amino groups capped by *N*-methyl and acetyl units, respectively. These additional compounds complement the chromophore diversity for organic and biological relevant systems, so that a larger extended π-system (anthracene), an aliphatic conjugated π-system containing heteroatoms (acrolein) and a amino acid within a protein-like structure (capped histidine) are included. Thus, the benchmark set represents various structural properties commonly found in organic and biological chromophores, which correspond to the ap-

plication field of the presented method.

The geometries of the benchmark molecules were first optimized at the B3LYP/6-31G* level of theory.[61;62;85] Full EOM-CCSD calculations were then carried out to obtain vertical excitation energies which served as reference values for the evaluation of incremental approaches. Within these calculations, a modified aug-cc-pVDZ basis set was employed, for which the augmented functions were removed from the hydrogen atoms.[86;87] This basis will be denoted as A'VDZ. In order to characterize these excitations, CIS/A'VDZ excited states were calculated and assigned according to their singles amplitudes. The CIS transition density matrices were then employed to construct the NTOs by which the excitation character could be determined. For the benchmark calculations, only low-lying valence excited states were considered, including 44 excitations in total. These comprise of 27 $\pi \rightarrow \pi^*$ and 17 $n \rightarrow \pi^*$ transitions. Rydberg states were excluded from the benchmark set, because the calculations with augmented basis functions can be contaminated with ghost states of this type. In principle, these states are of minor importance for UV-vis spectroscopy as analytic tool, which is the main application of the presented approach. The reference data are collected in Table 3.1. All calculations were carried out with a development version of the MOLPRO 2015.1 program package.[88;89]
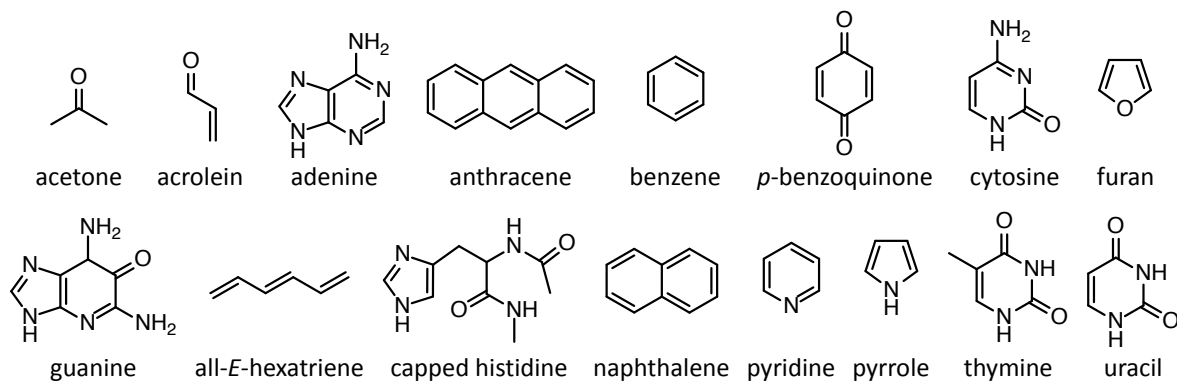
Figure 3.3.: Structural formulas and retained names of the molecules of the benchmark set for the incremental EOM-CCSD method.

Table 3.1.: Excitation characters and vertical excitation energies of low-lying valence excited states for the benchmark molecules shown in Figure 3.3, employing the A'VDZ basis set. All energies are given in electronvolts (eV). The EOM-CCSD excitation energies served as reference values for the incremental approach. The CIS states were used as reference states for the latter.

molecule	character	EOM-CCSD	CIS
acetone	$n \to \pi^*$	4.512	5.197
acrolein	$n \to \pi^*$	3.904	4.697
	$\pi \to \pi^*$	6.907	7.036
	$n \to \pi^*$	7.229	8.749
adenine	$\pi \to \pi^*$	5.314	6.280
	$\pi \to \pi^*$	5.456	6.097
	$n \to \pi^*$	5.494	7.006
anthracene	$\pi \to \pi^*$	3.860	4.560
	$\pi \to \pi^*$	3.970	4.033
	$\pi \to \pi^*$	5.505	5.760
benzene	$\pi \to \pi^*$	5.222	6.041
	$\pi \to \pi^*$	6.567	6.211
	$\pi \to \pi^*$	7.382	7.809
benzoquinone	$n \to \pi^*$	3.124	4.059
	$n \to \pi^*$	3.296	4.267
	$\pi \to \pi^*$	4.902	5.179
cytosine	$\pi \to \pi^*$	4.867	5.921
	$n \to \pi^*$	5.353	6.790
	$\pi \to \pi^*$	5.895	7.342
	$n \to \pi^*$	5.990	7.276
furan	$\pi \to \pi^*$	6.510	6.352
	$\pi \to \pi^*$	6.761	7.936
guanine	$\pi \to \pi^*$	5.125	5.917
	$n \to \pi^*$	5.650	6.741
	$\pi \to \pi^*$	5.719	6.414
hextriene	$\pi \to \pi^*$	5.488	5.344
	$\pi \to \pi^*$	6.529	8.035
histidine	$n \to \pi^*$	5.757	6.729
	$n \to \pi^*$	5.929	6.957
	$\pi \to \pi^*$	6.260	6.481
naphthalene	$\pi \to \pi^*$	4.393	5.164
	$\pi \to \pi^*$	5.059	5.023
	$\pi \to \pi^*$	6.107	6.343
pyridine	$n \to \pi^*$	5.185	6.153
	$\pi \to \pi^*$	5.271	6.091
	$n \to \pi^*$	5.592	7.339
pyrrole	$\pi \to \pi^*$	6.362	6.365
	$\pi \to \pi^*$	7.110	7.145
thymine	$n \to \pi^*$	5.188	6.308
	$\pi \to \pi^*$	5.428	6.159
	$n \to \pi^*$	6.584	7.742
uracil	$n \to \pi^*$	5.163	6.313
	$\pi \to \pi^*$	5.526	6.344
	$n \to \pi^*$	6.504	7.679

3.2.2. Group Construction Algorithms

With the scheme for the incremental approach described in Section 3.1 various algorithms can be designed for the group construction. In the following, the basic algorithm is presented. Afterwards, the differences of four applied algorithms are explained in detail.

The group construction is carried out right after the selection of the NTOs (threshold for their singular value $\lambda_{\text{thresh}}^{\text{NTO}} > 0.05$) and the localization of the remaining valence orbitals. Depending on the applied algorithm, LMOs considerably far away from the excitation space were excluded from the incremental approach before building the groups and effectively treated as core orbitals. A threshold for the density overlap to the excitation space $S_{\text{exc},i}^{\text{h/p}}$ was defined in order to identify these LMOs. The selected NTOs then constituted the starting point the main fragment. Subsequently, LMOs were added to this group according to $S_{\text{exc},i}^{\text{h/p}}$ in descending order. This was done until the parameter $S_{\text{exc},i}^{\text{h/p}}$ came below a given threshold or a pre-defined group size limit was reached. The corresponding next LMO then constituted the starting point of a new group and served as reference orbital. LMOs were added to this group according to the orbital exchange interaction $\tilde{K}_{ij}$ in descending order until this parameter fell below a threshold or the group size limit was reached. The corresponding next LMO then constituted the starting point of a new group and the protocol was repeated until all selected LMOs were assigned to a group. Eventually, EOM-CCSD/A'VDZ calculations completed the incremental approach.

Within the basic algorithm, four different settings were used in order to benchmark the effect of different group sizes, selection of orbital spaces or extension of the many-body expansion on the incremental approach. The specifications of the four different algorithms A–D are listed in Table 3.2. In summary, the main characteristics of these group construction algorithms are:

(A) This algorithm was adapted from the original method of Mata and Stoll by considering the new criteria for the group construction.[38]

(B) The group size limit of the main fragment was moderately raised to six orbitals. This modification is intended to result in a better description of the excitation within the main fragment.

(C) LMOs which are expected to have a minor impact on the excitation because of a low overlap to the excitation space were excluded *a priori* from the incremental

Table 3.2.: Settings for the four different group construction algorithms A–D applied to the base algorithm described in the text.

| algorithm | exclusion of LMOs | | group size limit | | thresholds | |
	choice	$S_{\mathrm{exc},i}^{\mathrm{h/p}}$	main	other	main ($S_{\mathrm{exc},i}^{\mathrm{h/p}}$)	other ($\tilde{K}_{ij}$)
A	no	—	4	4	> 0.25	> 0.015
B	no	—	6	4	> 0.25	> 0.015
C	yes	< 0.1	4	4	> 0.25	> 0.015
D	yes	< 0.1	∞	—	> 0.00	—

approach. These were identified according to a density overlap to the excitation space with $S_{\mathrm{exc},i}^{\mathrm{h/p}} < 0.1$.

(D) LMOs were excluded *a priori* from the incremental approach as in algorithm C. The remaining LMOs were assigned to the main fragment resulting in a full EOM-CCSD calculation with a reduced orbital space.

3.2.3. Results and Discussion

The incremental approach with the different group construction algorithms was implemented as its own module and carried out within a development version of MOLPRO 2015.1.[88;89] The deviations from the full EOM-CCSD/A'VDZ excitation energies are depicted together with the fraction of included valence orbitals for algorithms C and D in Figure 3.4. A statistical analysis for the benchmark calculations is summarized in Table 3.3. This analysis includes the mean errors (MEs) and corresponding standard deviations (STDs) as well as the mean absolute deviations (MADs) for each algorithm and excitation type.

The results for algorithm A, which corresponds to the original procedure in combination with the new criteria, already look promising with an error of -0.03 ± 0.15 eV. A systematic error is not observed for this scheme. Furthermore, the MAD of about 0.1 eV is within the intrinsic error of the conventional method (≈ 0.2 eV).[31] However, some important issues concerning this method can be identified. With respect to the MADs, a discrepancy between $n \to \pi^*$ and $\pi \to \pi^*$ is observed (0.15 and 0.08 eV, respectively). This stems most likely from the more complex nature of $n \to \pi^*$ transitions, for which the

Table 3.3.: Statistical descriptors for the deviations of the incremental EOM-CCSD method for the different group construction algorithms to the full calculations. All values are given in eV. The analysis was carried out individually for $n \to \pi^*$ and $\pi \to \pi^*$ transitions as well as for the complete benchmark set (all). ME: mean error, STD: standard deviation, MAD: mean absolute deviation.

	algorithm A			algorithm B			algorithm C			algorithm D		
	$n \to \pi^*$	$\pi \to \pi^*$	all	$n \to \pi^*$	$\pi \to \pi^*$	all	$n \to \pi^*$	$\pi \to \pi^*$	all	$n \to \pi^*$	$\pi \to \pi^*$	all
ME	−0.07	0.00	−0.03	−0.01	−0.01	−0.01	0.18	0.09	0.13	0.21	0.08	0.13
STD	0.20	0.11	0.15	0.16	0.08	0.12	0.17	0.11	0.14	0.10	0.09	0.11
MAD	0.15	0.08	0.11	0.09	0.06	0.07	0.21	0.10	0.14	0.21	0.08	0.13
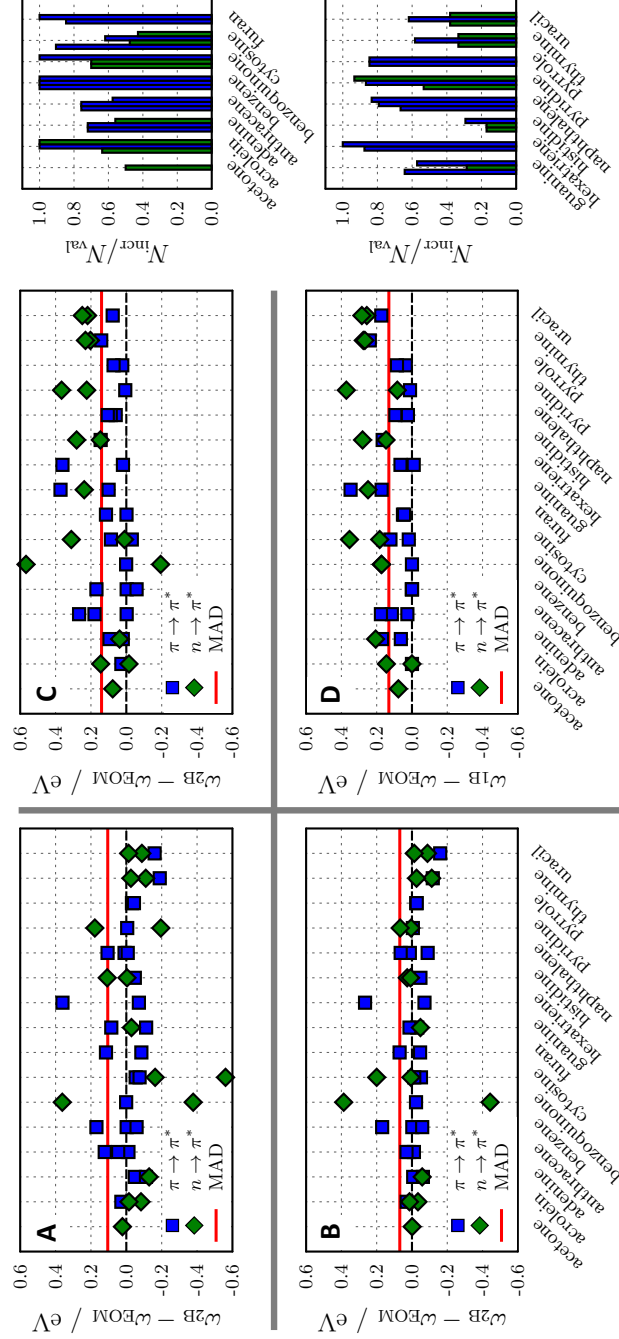
Figure 3.4.: Deviations of the incremental EOM-CCSD excitation energies for the different group construction algorithms A–D from the full calculations (ω_{EOM}). Values for the $n \to \pi^*$ and $\pi \to \pi^*$ transitions are represented as green diamonds and blue squares, respectively. Additionally, the mean absolute deviations (MAD) for the complete benchmark set are given as red lines. On the right hand side, the fraction of included valence orbitals for the algorithms C and D are depicted, for which the order of the reference EOM-CCSD calculations are taken into account as well as the color-code for the excited state characters.

the hole density is often spatially contracted to a specific part of the molecule and the particle density is delocalized over the π-system. Thus, the construction of the main fragment is more sensitive to the incremental approach, where more LMOs can be important for the excitations. A representative example is described further below ($n \rightarrow \pi^*$ transition of cytosine). In contrast, $\pi \rightarrow \pi^*$ excitations constitute a similar hole and particle space. Here, the most important LMOs are of π-type, which are easily identified and included within the algorithm.

Some outliers are present for algorithm A. For example, both $n \rightarrow \pi^*$ transitions of *p*-benzoquinone exhibit large deviations (-0.38 and 0.36 eV for the first and second excited state, respectively). The selected occupied-virtual NTO pairs for these states are shown in Figure 3.5. For both states the hole densities are symmetrically delocalized over the two carbonyl oxygens atoms, but additionally have significant contributions from the whole σ-bond framework. The algorithm is designed to group spatially close LMOs together. This results in a loss of symmetry in the incremental EOM-CCSD calculations, which most likely causes the large deviations to the full calculation. In line with this explanation, the errors reduce to 0.02 and -0.01 eV for the first and second excited state, respectively, if symmetry equivalent orbitals are grouped together. Consequently, such systems have to be treated with care by the incremental approach. However, the focus for the development of this method is the application to large biomolecules, which have seldom such a high symmetry. Therefore, the analysis and development should not be concerned because of these particular outliers.
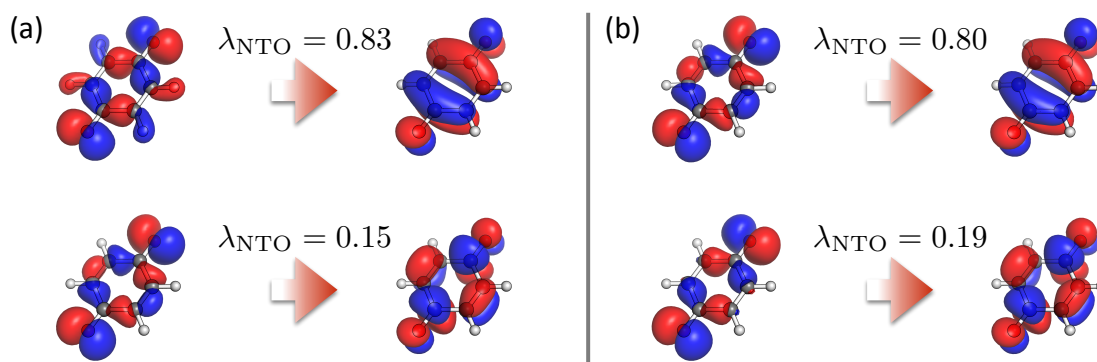
Figure 3.5.: Occupied-virtual NTO pairs of *p*-benzoquinone selected within the incremental EOM-CCSD method according to their singular values λ_{NTO} of the CIS transition density matrix for (a) the first and (b) the second $n \rightarrow \pi^*$ transition.

Another apparent deviation is observed for the second $\pi \to \pi^*$ transition of hexatriene showing a difference of 0.36 eV in comparison to the full calculation. The origin of this error differs from the one observed for the $n \to \pi^*$ transitions of p-benzoquinone. In contrast to the latter, the full EOM-CCSD result featured a significant doubles contribution of 19 % to the excitation. The CIS reference is not able to capture this effect properly, so that more correction terms seem to be required than only 2B terms. The first $\pi \to \pi^*$ excitation of hexatriene suffers not from this problem (only 7 % doubles contribution) and has a reasonable error of -0.07 eV. Accordingly, excitations with significant doubles contribution have to be taken with care, when employing a CIS reference state in the incremental scheme. However, single electron excitations, which should be primarily addressed by the incremental EOM-CCSD method, commonly dominate low-energy UV-vis spectra of organic chromophores.

The last prominent outlier corresponding to the second $n \to \pi^*$ transition of cytosine has a deviation of -0.56 eV. This seems to be a special case of the difficulty in describing the main fragment for $n \to \pi^*$ excitations as discussed above. The orbitals of the main fragment, the first LMO of the second group as well as the virtual NTOs are shown in Figure 3.6. The σ_{C-O} LMO cannot be assigned to the main fragment due to the group size limit of four orbitals for algorithm A, although the density overlap to the hole density is quite large ($S_{exc}^h = 0.26$). Regarding the dominating occupied NTO, this LMO is expected to be important for the description of the excitation, which is also indicated by a large 2B correction term of -0.77 eV for its associated group. A better description of the main fragment seems to be required. This is exactly what was examined with algorithm B, for which six orbitals are allowed for the main fragment. Here, only the σ_{C-O} LMO was added, but the latter already reduced the error to 0.20 eV. The largest 2B contribution of -0.34 eV and a reduced 1B term (by -0.3 eV) indicate that this orbital has to be included for an appropriate description of the excitation.

The moderate enlargement of the main fragment in algorithm B significantly improves the overall performance of the approach yielding an error of -0.01 ± 0.12 eV. The MAD is also considerably reduced to 0.07 eV. Furthermore, the results for both excitation classes are improved in comparison to algorithm A. The outliers for p-benzoquinone and hexatriene are still observed, but as mentioned before, the developed method has not been designed for the corresponding excitations. If these cases are excluded from the statistical analysis, the error and MAD decrease to -0.01 ± 0.06 eV and 0.05 eV, respectively. This is striking. The deviations of the incremental approach (algorithm B)

from the full EOM-CCSD calculations are only a fraction of the accuray of the EOM-CCSD method itself.

Within the algorithms C and D, LMOs were determined, which are far away from the excitation space, based on the density overlap. These identified LMOs were excluded beforehand of the group construction and thus, treated effectively as core orbitals within the incremental approach. For the 2B and 1B approaches the results show a systematic overestimation for nearly all transitions with errors of 0.13 ± 0.14 and 0.13 ± 0.11 eV, respectively. The reason for this trend is most likely due to an undercorrelated CCSD ground state. This in turn raises the excitation energy in the EOM formalism, for which a CI-type parametrization is set on top. Probably, the benchmark set is not well suited to test this methodology, because it only consists of small to medium sized organic molecules, which still allow for full EOM-CCSD calculations as reference values. Therefore, most LMOs are close to the excitations space and have to be included. The latter is also evident from the fraction of included valence orbitals as depicted in Figure 3.4. For some transitions even all valence MOs were selected. Thus, the selection of the valence orbital space would be more interesting for large molecules, for which the chromophore is only a small part.

Another problem in this scheme might arise from the exclusion of only specific valence MOs. The latter can have an indirect correlation effect to spatially close orbitals, which have to be included to describe the excitation properly. A solution to this problem would be to change the exclusion criterion according to the following procedure: Whole groups are rejected from the incremental approach after their construction, if every LMO has no significant overlap to the hole and particle densities. For this purpose, the sum of

(a) $\lambda_{\mathrm{NTO}}^{\mathrm{occ}} = 0.94$ $\lambda_{\mathrm{NTO}}^{\mathrm{occ}} = 0.05$ $S_{\mathrm{exc}}^{\mathrm{p}} = 0.68$ $S_{\mathrm{exc}}^{\mathrm{h}} = 0.48$ (b) $S_{\mathrm{exc}}^{\mathrm{h}} = 0.27$ (c) $\lambda_{\mathrm{NTO}}^{\mathrm{virt}} = 0.94$ $\lambda_{\mathrm{NTO}}^{\mathrm{virt}} = 0.05$
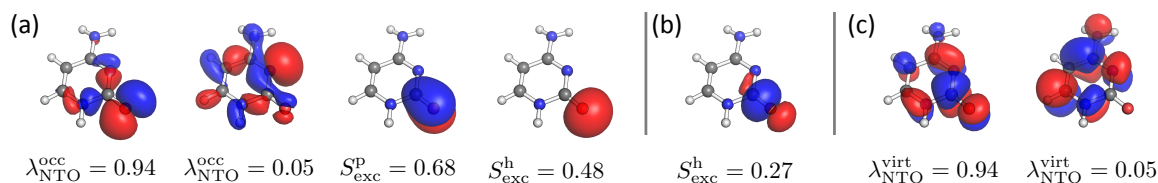
Figure 3.6.: Molecular orbitals of cytosine after the transformations within the incremental EOM-CCSD scheme for the second $n \rightarrow \pi^*$ transition. Additionally, the singular values of the CIS transition density matrix for the selected NTOs and the overlap between hole/particle and LMO densities are given. (a) NTOs and LMOs of the main fragment according to group construction algorithm A. (b) $\sigma_{\mathrm{C-O}}$ LMO which starts a new fragment in algorithm A, but is included in the main fragment for algorithm B. (c) Associated virtual NTOs in order to display the particle space after the excitation.

the density overlaps of all LMOs in a group G could be employed as criterion to identify those fragments:

$$S_{\text{exc},G} = \sum_{i \in G} S_{\text{exc},i}^{\text{h/p}} \ . \tag{3.10}$$

In order to test this suggestion, *a posteriori* analysis of the incremental EOM-CCSD calculations resulting from algorithm B was carried out for the excitations of the capped histidine. This molecule is the only one from the benchmark set that can be considered in this regard. It is large enough and consists of three well-separated chromophore moieties (two amide groups and the imidazole ring), where the transitions are localized. Figure 3.7 illustrates the new concept for the first $n \to \pi^*$ transition of the capped histidine. The group shown in Figure 3.7b is far away from the excitation space which is well represented by the new criterion ($S_{\text{exc},G} = 0.002$). As the 2B correction term associated with this group is -0.001 eV, its incorporation in the incremental scheme is of minor importance and safely can be neglected.In contrast, the group shown in Figure 3.7c is closer to the excitation space ($S_{\text{exc},G} = 0.08$) and thus has a significant

(a)   (b)   (c)

$\lambda_{\text{NTO}} = 0.98$

$S_{\text{exc},G} = 0.002$
$\Delta\omega_{\text{2B}} = -0.001$ eV

$S_{\text{exc},G} = 0.080$
$\Delta\omega_{\text{2B}} = -0.066$ eV
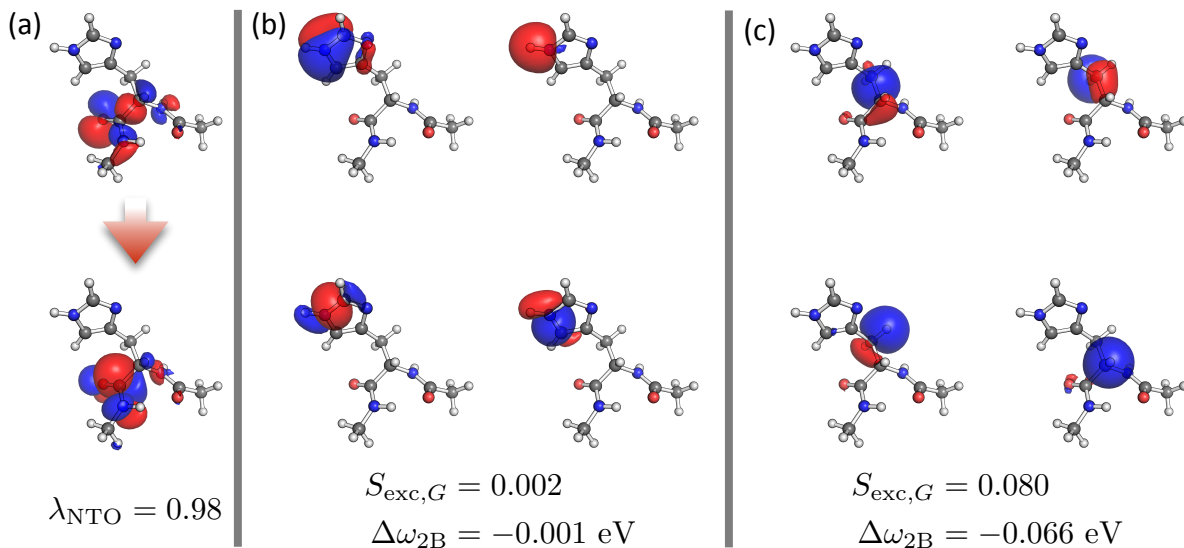
Figure 3.7.: Illustration of the new exclusion criterion for whole groups for the first $n \to \pi^*$ transition of the capped histidine molecule after the group construction according to algorithm B. (a) Dominant occupied-virtual NTO pair describing the excitation space. (b) Orbital group being far apart from the excitation space with minor impact on the excitation energy. This is reflected by the parameter $S_{\text{exc},G}$. (c) Orbital group with significant overlap to the excitation space and a relevant 2B correction term.

impact on the calculated excitation energy ($\Delta\omega_{2B} = -0.066$ eV).

With a tight threshold of $S_{\text{exc},G} < 0.01$, the analysis was then carried out for the three excited states of the capped histidine. In case of the first $n \rightarrow \pi^*$ transition, five groups with 16 LMOs in total are excluded, corresponding to 39 % of all valence orbitals. Thereby, the effect on the excitation energy is negligibly small ($\Delta\omega = 0.007$ eV). For the second $n \rightarrow \pi^*$ and the $\pi \rightarrow \pi^*$ transition seven and five groups, containing 20 and 18 orbitals, are rejected with virtually no loss of accuracy ($\Delta\omega_{n\rightarrow\pi^*} = 0.008$ eV and $\Delta\omega_{\pi\rightarrow\pi^*} = 0.02$ eV). The new exclusion criterion outperforms the one used in algorithm C and D, for which deviations larger than 0.1 eV are observed. Despite these encouraging results, further test systems should be considered to confirm this method for the exclusion of groups within the incremental calculations. Unfortunately, the benchmark set lacks of further molecules for such a treatment, but it should be kept in mind for the application to other larger molecules.

3.3. Conclusions and Outlook

The benchmark calculations have shown the ability of the incremental approach to accurately recover the full EOM-CCSD excitation energies. The group construction algorithm B outperforms the other ones because of a better description of the main fragment by allowing more orbitals to be considered. The MAD of 0.05 eV after excluding the discussed outliers is rather low and constitute only a fraction of the intrinsic accuracy of the conventional EOM-CCSD method (≈ 0.2 eV).[31] A smaller group size of the main fragment (algorithm A) raises the corresponding MAD to 0.09 eV, which is still satisfying. The benchmark calculations also helped to identify some of the issues in this incremental procedure. They can occur for highly symmetric molecules, for which the fragment calculations break the symmetry, leading to inbalanced descriptions. Moreover, excitations with large doubles contributions are poorly described because of the CIS reference state. However, the incremental EOM-CCSD method is intended for the application on single electron excitations for large biomolecules, having seldom a high symmetry. Indeed, the transitions of the benchmark set, which are more representative for the purpose of the method, were excellently described by the implemented version.

The exclusion of LMOs preceding the group construction (algorithms C and D) systematically overestimated the excitation energies obtained from the incremental approach.

In contrast, the exclusion of whole groups based on the sum of the density overlaps of the LMOs showed a better performance, which was demonstrated for the three transitions of the capped histidine molecule. The combination of algorithm B and the prior exclusion of fragments seems promising for the application to large molecules, in particular if the chromophore moieties are restricted to a small region. In line with this, the method should be well suited for the application to the ThDP cofactor and derivatives featured in this work. The latter will provide further information, for example, on the performance to charge-transfer states.

The developed method also performs well in comparison to other fragmentation techniques available for excited state calculations. The divide-and-conquer (DC) approach represents such a method.[81] Therein, the system is spatially divided into subsystems, which are extended by a buffer region. The individual calculations for these subsytems are then combined to recover the full calculation. For the application to excited states, one of these fragments has to be the pigment, where the excitation takes place. The fragment MOs are then employed as basis in the excited state calculations. This has been implemented for CIS, TD-DFT and SACCI methods and successfully applied to formaldehyde in water, the $n \rightarrow \pi^*$ transition of a conjugated aldehyde ($C_{16}H_{17}CHO$) and the active center of the photoactive yellow protein.[90] The DC-based methods showed deviations of about 0.1 eV in comparison to the conventional calculations. This error is within the intrinsic accuracy of the latter, but does not outperform the presented incremental approach. Furthermore, the manual definition of a excited state subsystem for the DC calculations can be problematic, as the results vary with the size of this region. This drawback is circumvented in the incremental approach by employing a low-level CIS reference calculation in order to determine the important excitation space. Additionally, large conjugated π-systems could not be divided into several subsystems by the DC approach, requiring the full calculation. In contrast, the incremental approach employs orbital fragments in combination with a many-body expansion. Therefore, its application to such excitations is straightforward.

Another important feature used in all fragmentation techniques is the capability of parallel computing. Each fragment can be calculated individually on different processors or nodes of a computing cluster and afterwards combined to get the full result. This procedure could also be applied for the incremental approach. After the group construction the calculations of the increment EOM-CCSD excitation energies are independent from each other. In future development, parallelization should be implemented within

a standard master-worker model, which will raise the efficiency of the method.

Local correlations methods represent another methodology to speed up quantum chemical calculations. Nowadays, these are well established for ground state calculations[91–93] and have been extended to the EOM-CCSD method for excited states by Korona and Werner.[94] Therein, a CIS reference calculation is carried out in order to determine the excitation space and hence, important LMOs, which significantly overlap to this region, are identified. Excitation domains for each important LMO are then constructed consisting of projected atomic orbitals. Truncation of the excitation operater $\hat{T}$ to these domains (and pair domains for double excitations) significantly reduces the computational demand. Test calculations showed an average agreement of 0.06 eV in comparison to the full EOM-CCSD excitation energies.[94] Most of the described procedure is already featured in the incremental EOM-CCSD method, for which the use of the CIS reference state identifies the excitation space through the NTOs. Important LMOs are also determined by their density overlap to the excitation space. Therefore, the two methodologies should be coupled in the future development. The application of local EOM-CCSD calculations for each increment would speed up the approach enormously.

Pair natural orbital (PNO) based methods are somewhat different approaches for exploiting the short-range character of correlation effects.[95–98] PNOs are built as linear combinations of virtual orbitals, which diagonalize LMO pair density matrices of correlated approximate wave functions. The number of PNOs for pair excitations is then truncated according to their natural orbitals occupation numbers leading to a tremendous reduction of the virtual space. This concept has been applied to excited states by Hättig and coworkers.[99–101] Initially, it was implemented for the perturbative doubles correction (D) to CIS excitation energies. Afterwards, the extension to the iterative linear response CC2 and CCSD methods were made using the CIS(D) model as initial guess.[102] Therein, state-spcific PNOs for excited states were employed besides a different ground state PNO basis. At reasonable thresholds for the truncation of the virtual space, this approach leads to a significant speed up as well as high accuracy (average error of about 0.01 eV in comparison to the conventional calculations). Thus, a combination of the PNO with the incremental method would be highly desirable. For this combination, the mixed NTO-LMO occupied orbital space could become a bottleneck. This was already investigated by Hättig and coworkers in their initial publication.[99] In comparison to a full localized basis, the stronger delocalized nature of the corresponding occupied space requires more PNOs to achieve the same accuracy. However, the

reduced occupied orbital space and the possibility of massive parallization of the incremental approach might overcompensate this issue. Therefore, the development of a PNO increment scheme is well worth an attempt.

<div style="text-align: right">*4*</div>

Pyruvate Decarboxylase

Parts of this chapter are featured in

"Theoretical Studies of the Electronic Absorption Spectra of Thiamin Diphosphate in Pyruvate Decarboxylase"

<u>M. Paulikat</u>, C. Wechlser, K. Tittmann, R. A. Mata, *Biochemistry* **2017**, *56*, 1854–1864. doi: 10.1021/acs.biochem.6b00984

Pyruvate decarboxylase (PDC) is a ThDP-dependent enzyme, catalyzing the conversion of pyruvate in acetaldehyde and carbon dioxide.[13] This reaction is a fundamental step in the pathway of fermentation processes under anaerobic conditions, where carbohydrates are metabolized into alcoholic species.[6] The overall reaction catalyzed by the PDC enzyme is shown in Figure 4.1.

PDC from *Zymomonas mobilis* (*Zm*PDC) was employed as a model system to study the absorption signatures of the pre-equilibrium states of ThDP.[16;23] The crystal structure of *Zm*PDC is shown in Figure 4.2 together with the structural formula of the 4'-aminopyrimidine (AP) and 1',4'-iminopyrimidine (IP) state of ThDP.[103] Addition-
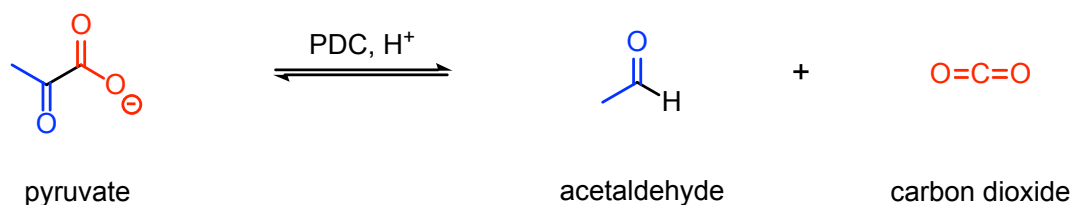
Figure 4.1.: Overall reaction of the conversion of pyruvate into acetaldehyde and carbon dioxide catalyzed by PDC.

ally, CD spectra of *Zm*PDC wild-type as well as its mutants Glu473Gln and Glu50Gln have been revisited by the Tittmann group (Figure 4.2c).[104]

*Zm*PDC is a homotetrameric enzyme which can be divided into two subunits, each consisting of two monomers. Within a subunit the monomers are tightly bound and related by a two-fold symmetry axis. The cofactors are placed in the boundary of these monomers and proposed to be in equilibrium between the AP and IP states of ThDP. The tautomerization is mediated by a canonical glutamate which is placed next to the N1' atom of the pyrimidine ring.[16] After protonation of this atom in AP-ThDP,
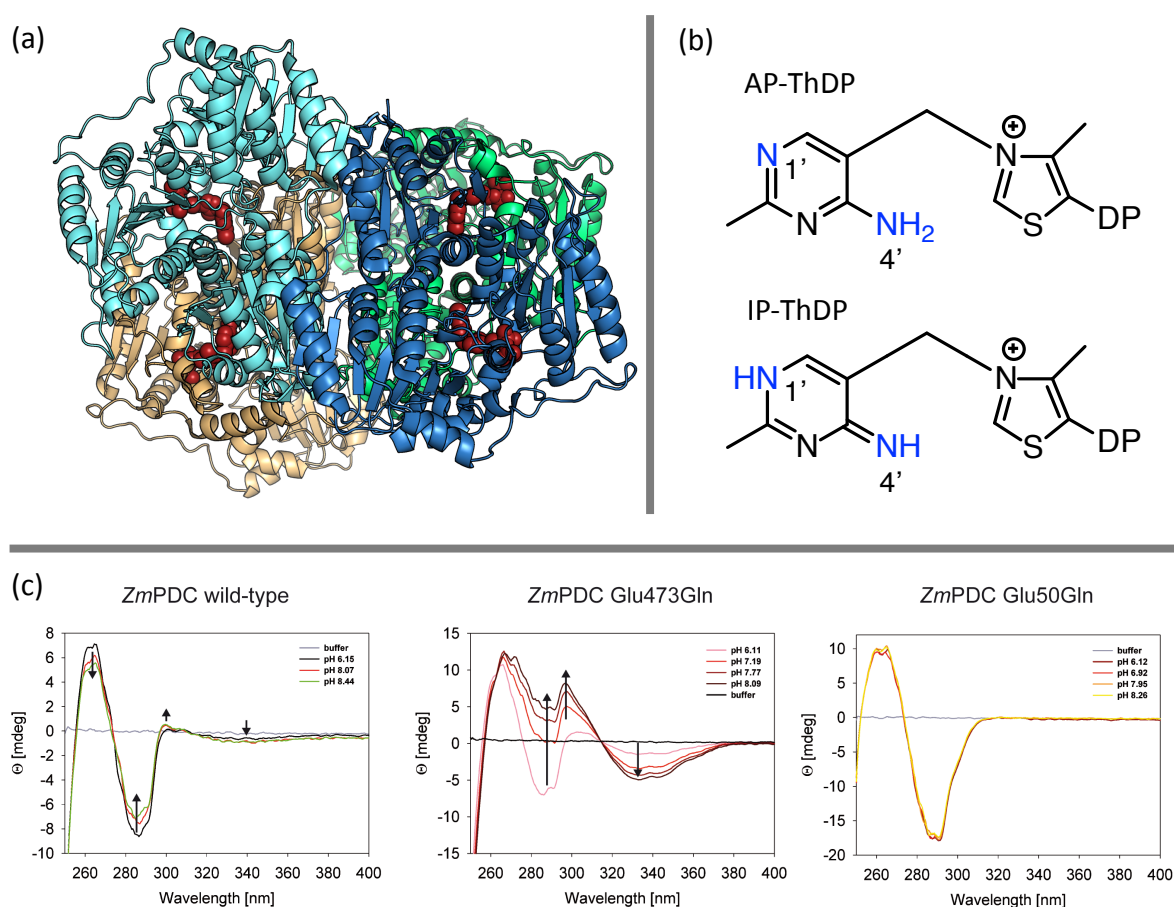
Figure 4.2.: (a) Cartoon plot of the crystal structure of *Zm*PDC (PDB entry 2WVG).[103] The enzyme monomers are highlighted in different colors. The cofactor analogues are shown as red spheres. (b) Structural formulas of the 4'-aminopyrimidine (AP; top) and 1',4'-iminopyrimidine (IP; bottom) tautomers of ThDP. DP abbreviates the diphosphate moiety of ThDP. (c) Revisited near-UV spectra of *Zm*PDC wild-type and its variants Glu473Gln and Glu50Gln for different pH values, recorded at the Tittmann group in Göttingen. The figure is reprinted with permission from K. Tittmann.

its conjugated acid, 4'-aminopyrimidinium (APH$^+$) ThDP could be formed, which in turn would lead to IP-ThDP through proton abstraction at the exocyclic amino group. However, the APH$^+$ state of ThDP was not detected in the resting state of *Zm*PDC until now. A coupled proton transfer would also be conceivable, converting directly between AP and IP-ThDP. In this resting state of *Zm*PDC, two CD bands with opposite phase were observed in the spectral region of 300–340 nm, also a common feature for other ThDP-dependent enzymes.[23] In the revisited spectra of Tittmann *et al.*, a third band could be detected at slightly lower wavelengths. These spectral signatures can be summarized as follows:

(1) Negative CD band at 320–340 nm (3.65–3.87 eV): The band was assigned to the 4'-aminopyrimidine (AP) state of ThDP, arising from a charge-transfer (CT) excitation between the AP and thiazolium ring. The assignment was based on model experiments, where a band at 340 nm (3.65 eV) was observed for a binary solution of 4-aminopyrimidine and 3,4,5-trimethylthiazolium.[29]

(2) Positive CD band at 300–310 nm (4.00–4.13 eV): This band was referred to the 1',4'-iminopyrimidine (IP) tautomer of ThDP. Structural analysis of ThDP in the transketolase enzyme at high resolution ($\leq$1.0 Å) lead to this conclusion, where the state of ThDP could be identified by the calculated electron density maps.[105] The excitation character has been deduced from chemical model experiments of N^1,N^4-dimethyl-1,4-iminopyrimidine and derivatives. The absorption in the corresponding spectral region together with an observed hypsochromic shift pointed toward a $n \rightarrow \pi^*$ transition localized in the IP ring.[30]

(3) Negative CD band at 280–290 nm (4.28–4.43 eV). This is a hitherto unassigned band, observed in the revisited spectra of *Zm*PDC by Tittmann *et al.* In particular, this is the only band remaining where the canonical glutamate is isosterically mutated to a glutamine (see Figure 4.2c).

Up to date the spectral signatures of the pre-equilibrium states of ThDP have not been fully understood. Therefore, theoretical methods were employed to provide a clear view. This includes excited state calculations for model compounds of the cofactors, but also a full QM/MM description of ThDP in *Zm*PDC. Additionally, a perturbative residue-by-residue analysis was carried out in the active site to identify those with a significant effect on the cofactor spectra.

4.1. Computational Methods

4.1.1. Perturbative Analysis of Enzyme Residues on Cofactor Spectra

The efficient calculation of excitation energies in complex environments is highly desirable. A way to implement that is the computation of easily available reference systems *in vacuo* and adding the complex environment by means of perturbation theory (PT).[106] Renger and coworkers proposed such a method for the calculation of site energy shifts of pigments embedded in a protein.[107–110] The ground and excited states of these pigments were coupled to the electrostatic field by the Coulomb interaction as perturbation. The protein atoms were hereby represented as atomic point charges. This has been successfully applied to chlorophyll pigments in light-harvesting proteins, essential for photosynthesis. However, a related approach can be employed for the analysis of the influence of the protein environment on the electronic absorption spectra of the cofactor ThDP. In the following, first-order PT is briefly reviewed for the calculation of excitation energy shifts. Afterward, the procedure is explained to apply this as an analytic tool.

The first-order PT energy correction is given by[106]

$$E_n^{(1)} = \langle \Psi_n^{\mathrm{ref}} | \hat{H}^{\mathrm{PT}} | \Psi_n^{\mathrm{ref}} \rangle \;, \tag{4.1}$$

where a perturbation described by the operator $\hat{H}^{\mathrm{PT}}$ acts on a reference state Ψ_n^{ref} in electronic state n. This equation can be rewritten for a multiplicative, one-electron perturbation operator $\hat{h}^{\mathrm{PT}}$ in the following way:

$$E_n^{(1)} = \int \mathrm{d}\mathbf{r}\, \hat{h}^{\mathrm{PT}} \rho_n^{\mathrm{ref}}(\mathbf{r}) \;. \tag{4.2}$$

Here, ρ_n^{ref} denotes the one-electron density of the reference system in the corresponding electronic state. Employing molecular orbital theory, the electronic density can be expressed by means of the one-electron density matrix contracting the MOs:

$$\rho_n^{\mathrm{ref}}(\mathbf{r}) = \sum_{pq} D_{pq}^{n(\mathrm{MO})} \phi_p^*(\mathbf{r}) \phi_q(\mathbf{r}) \;, \tag{4.3}$$

with $D_{pq}^{n(\mathrm{MO})}$ and ϕ_p being the elements of this density matrix and the MOs, respec-

tively. This representation of the density can now be substituted into Equation (4.2) and simplified as

$$E_n^{(1)} = \sum_{pq} D_{pq}^{n(\mathrm{MO})} h_{pq}^{\mathrm{PT(MO)}} \; , \tag{4.4}$$

$$E_n^{(1)} = \mathrm{Tr} \left\{ \mathbf{D}^n \mathbf{h}^{\mathrm{PT}} \right\} \; , \tag{4.5}$$

where the matrix representation of the perturbation operator in the MO basis was employed. The extension to vertical excitation energy shifts is now straightforward, since the energies of the different states are just additive. For an excitation from the ground state 0 to an excited state a this can be expressed as

$$\omega_{0a}^{\mathrm{PT}} = E_a^{(0)} + E_a^{(1)} - \left(E_0^{(0)} + E_0^{(1)} \right) \; , \tag{4.6}$$

$$\omega_{0a}^{\mathrm{PT}} = \omega_{0a}^{(0)} + \Delta\omega_{0a}^{\mathrm{PT}} \; , \tag{4.7}$$

where the excitation energy of the reference system $\omega_{0a}^{(0)} = E_a^{(0)} - E_0^{(0)}$ is shifted due to the perturbation by $\Delta\omega_{0a}^{\mathrm{PT}} = E_a^{(1)} - E_0^{(1)}$. For the first-order PT energy corrections of both states the expression from Equation (4.4) can be applied and summarized as

$$\Delta\omega_{0a}^{\mathrm{PT}} = \sum_{pq} \left(D_{pq}^{a(\mathrm{MO})} - D_{pq}^{0(\mathrm{MO})} \right) h_{pq}^{\mathrm{PT(MO)}} \; , \tag{4.8}$$

$$\Delta\omega_{0a}^{\mathrm{PT}} = \mathrm{Tr} \left\{ \mathbf{\Delta D}^{0a} \mathbf{h}^{\mathrm{PT}} \right\} \; , \tag{4.9}$$

where the difference density matrix $\mathbf{\Delta D}^{0a}$ between the ground and excited state is introduced. Up to this point only the restriction that the perturbation operator is a multiplicative, one-electron operator has been made, but neither the reference system nor the nature of the perturbation was specified. In order to analyze the effect of enzyme residues on the cofactor spectra, the reference was chosen as the excitation of ThDP fully embedded in the protein environment represented as atomic point charges. This deviates from the approach of Renger and coworkers who used vacuum calculations of the pigment as reference and afterwards added the environment as the perturbation.[107–110] However, in the present case it is most likely the better choice to employ the fully embedded cofactors as references, since this already includes the polarization due to the enzyme. The perturbation is then defined that the point charges Q_J of specific residues

R are deleted to measure its effect. The perturbation operator is therefore given by

$$\hat{h}^{(R)} = \sum_{J \in R} \frac{Q_J}{|\mathbf{r} - \mathbf{R}_J|} \ , \tag{4.10}$$

with $\mathbf{r}$ and $\mathbf{R}_J$ being the position vectors of an electron and a MM atom, respectively. The shifts could now be calculated with Equation (4.9). In order to interpret these shifts, Equation (4.7) is slightly rearranged to

$$\omega_{0a}^{(0)} = \omega_{0a}^{(R)} - \Delta\omega_{0a}^{(R)} \ . \tag{4.11}$$

This means that the complete excitation energy is decomposed to a part $\omega_{0a}^{(R)}$ being the excitation energy without the residue R which is shifted by $-\Delta\omega_{0a}^{(R)}$ due to its presence. Therefore, the sign was included for the calculation of the descriptor, resulting in

$$\Delta\omega_{0a}^{(R)} = -\mathrm{Tr}\left\{ \mathbf{\Delta D}^{0a}\mathbf{h}^{(R)} \right\} \ . \tag{4.12}$$

A negative sign for the descriptor $\Delta\omega_{0a}^{(R)}$ then determines an assistant role for the excitation by the residue under study, whereas $\Delta\omega_{0a}^{(R)} > 0$ describes the opposite trend.

A last comment should be made about the nature of the perturbation. This can be easily extended to estimate the effect of different protonation states of the residues or even upon mutation of them. For the latter a similar conformation between the wild type and mutated species would be required. However, this can give valuable information for virtually no computational cost.

4.1.2. Computational Details

Model compounds were optimized at the B3LYP-D3(BJ)/def2-TZVP level of theory[61;62;111–113] and confirmed as minima through a frequency analysis. For this purpose and all following DFT computations, the resolution of the identity (RI) approximation with the corresponding auxilary basis set as well as the chain-of-spheres algorithm were employed to speed up the calculation of the Coulomb and exchange interactions, respectively.[114–116] Vertical excitation energies were then determined by TD-DFT in the TDA approximation (TD-DFT/TDA).[72] In order to include environmental effects, the COSMO solvation model was applied.[117] Hereby, the predefined settings were used from the ORCA program package for each solvent. Wave function based methods were consulted

for selected systems serving as reference calculations. Therefore, LR-CC2 and EOM-CCSD methods were applied with a modified aug-cc-pVDZ (A'VDZ) basis set, where the augmented functions were removed from the hydrogen atoms.[39;51;86;87]

The effect of residues on the cofactor spectra in the active site of *Zm*PDC was studied by the perturbation analysis described in the previous section. Here, the unchanged crystal structure of the active site (PDB entry 2WVG, 1.75 Å resolution) was taken and the cofactor analogue were replaced by authentic ThDP in both tautomeric states.[103] A residue shell of about 5 Å around the cofactors was considered, including 32 amino acids in total. Free residues were then capped by acetyl or *N*-methyl groups. The enzyme environment was parametrized with the Amber ff10 force field.[118] The assignment of the protonation states of titratable residues are later discussed in the text. The hydrogen atoms were optimized at the MM level, restraining all other atoms in their Cartesian coordinates by a harmonic potential with a force constant of $k = 5$ kcal mol^{-1} Å^{-2}. A QM/MM optimization followed for the cofactors at the B3LYP-D3(BJ)/def2-SVP level of theory.[61;62;111–113] Afterward, CAM-B3LYP/def2-TZVP excited states were calculated for each cofactor serving as reference states,[65;113] followed by the perturbative scheme for the excitation energy shifts.

The pre-equilibrium states of ThDP were also studied in the full enzyme environment. It was modeled on the basis of the crystal structure of *Zm*PDC (PDB entry 2WVG).[103] From the original tetrameric enzyme, a dimer was built. The cofactor analogues were replaced by authentic ThDP in both tautomeric states, reflecting the chemical asymmetry in the active sites as experimentally observed.[119] The protonation states of titratable residues were assigned after inspection of their local environment. For residues in the active site, the perturbative analysis was also considered. Different protonation patterns were tested, but followed the same computational protocol. The more questionable protonation states of important residues are also discussed in the text. Hydrogen atoms were then added to the protein environment and the total system charge was neutralized by sodium ions. The model system was then placed into a TIP3P water box with a distance of 12 Å between the protein and the boundary of the box. The protein was parametrized with the Amber ff10 force field,[118] whereas the GAFF force field was employed for AP and IP-ThDP.[120] Hereby, atomic partial charges were determined through an electrostatic potential fit at the HF/STO-3G level of theory. Two new atom types had to be introduced for the C4' and N4' atoms of IP-ThDP in order to obtain reasonable bond lengths within the IP moiety. For this reason stretch potentials with

single bond character were applied to N3'–C4' and C4'–C5' bonds, whereas double bond parameters were used for the C4'–N4' bond. The parameters of semi-aromatic character were applied to the angle and torsion potentials including these atoms. A MM optimization followed for the hydrogen atoms of the protein, water molecules and sodium ions, while all other atoms were restrained in their Cartesian coordinates by a harmonic potential with a force constant of $k = 3$ kcal mol^{-1} Å^{-2}. The obtained structures then served as reference for the MD simulations.

The MD simulations were initiated by a heating phase to a final temperature of 300 K within a NVT simulation in six stages over a total time of 60 ps. For this purpose and all further simulations, periodic boundary conditions were applied. Furthermore, the SHAKE algorithm was used to constrain bond lengths involving hydrogen atoms.[121;122] This allowed the application of a 2 fs time step. The temperature was controlled using Langevin dynamics with a collision frequency of 2 ps^{-1}. The system was then equilibrated by a NPT simulation, employing isotropic position scaling to achieve a pressure of 1 bar. The root-mean-square deviation (RMSD) of the backbone protein atoms was consulted as a convergence criterion for this phase. The NPT simulation was then extended for the final production phase.

Electronic structure calculations on the active site were then carried out in a QM/MM approach. For this purpose, snapshots were taken in 20 ps intervals from the MD simulation, including the whole enzyme as well as solvation shells of about 45 Å centered around the Mg^{2+} ions next to AP and IP-ThDP. The QM and MM regions were coupled through the electrostatic embedding scheme. In case of cutting covalent bond along the QM/MM boundary, hydrogen atoms were added as link atoms to the QM region, while the charge redistribution scheme of Sherwood and coworkers was applied for the adjacent MM atoms.[80] Spectra were then calculated with TD-DFT/TDA employing the CAM-B3LYP functional and def2-TZVP basis set. Program packages involved in the computations were ORCA versions 3.0.3 and 4.0.1 (DFT),[123] Amber 11 (MD),[124] ChemShell (QM/MM),[80;125] DALTON 2013 (LR-CC2)[126;127] and MOLPRO 2012.1 (EOM-CCSD).[88;89]

4.2. Results and Discussion

4.2.1. ThDP Model Calculations

In a first set of theoretical calculations, the isolated cofactor was considered. The objective was to understand the influence of specific geometries adopted in the active site and free in solution. At the same time, different electronic structure methods for the computation of the electronic excitation spectra were benchmarked. Given that the dominant tautomer of free ThDP in solution is the AP form, this one was considered in greater detail.

The first step was to compare different functionals for the treatment of the charge-transfer excitation of the ThDP cofactor in the AP form. The inclusion of the phosphate group tail with its negative charge can generally lead to ghost states in gas phase calculations, because this is not compensated by the environment (be it in solution or at the active site by protein residues). Therefore, to facilitate the comparison between the environments, the diphosphate moiety was replaced with an OH group. The resulting model molecule was first optimized in the *V* conformation, which is the one adopted in the active sites of enzymes. CC2 calculations were taken as a reference and compared to TD-DFT/TDA excitation energies employing the CAM-B3LYP,[65] M06-2X,[128] BH&HLYP[61] and B3LYP[61;62] functionals. The results are listed in Table 4.1.

The TD-DFT/TDA calculations with the CAM-B3LYP, M06-2X, and BH&HLYP functionals yield consistent excitation energies for the charge-transfer state, where an electron is transferred from the pyrimidine to thiazolium ring (see NTOs in Figure 4.3).

Table 4.1.: Vertical excitation energies for the AP and IP-ThDP model compounds in the *V* conformation, considering different levels of theory. All energies are given in electronvolt (eV). The corresponding wavelengths in nanometers (nm) are listed in parentheses.

ThDP	character	CC2[a]	CAM-B3LYP[b]	M06-2X[b]	BH&HLYP[b]	B3LYP[b]
AP	CT	4.41 (281)	4.26 (291)	4.34 (286)	4.39 (282)	3.10 (400)
	$n \to \pi^*$	4.25 (292)	4.58 (271)	4.66 (266)	5.01 (247)	4.35 (285)
IP	CT	4.36 (284)	4.33 (286)	4.39 (282)	4.47 (277)	3.39 (366)
	$\pi \to \pi^*$	4.80 (258)	4.93 (251)	4.99 (248)	5.15 (241)	4.51 (275)

[a] Basis: A'VDZ

[b] Basis: def2-TZVP
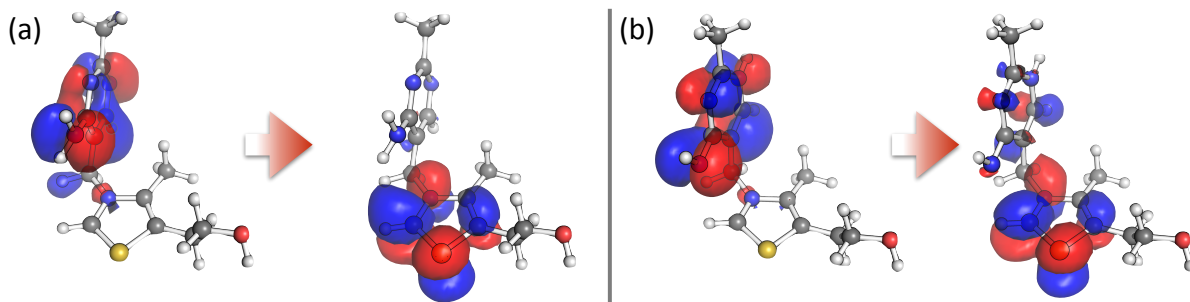
Figure 4.3.: Dominant CAM-B3LYP/def2-TZVP NTOs for the charge-transfer excitation of (a) AP and (b) IP-ThDP model compounds in the *V* conformation.

In contrast, B3LYP excitation energies show the typical underestimation of charge-transfer excitations for common DFT functionals.[66;67] The inclusion of a large fraction of exact exchange (M06-2X and BH&HLYP) or the application of range-separated functionals (CAM-B3LYP) is necessary for the accurate treatment of such excitations.

In the case of AP-ThDP, the reference CC2 calculation yield an $n \to \pi^*$ transition at the pyrimidine ring (4.25 eV, 292 nm) as lowest transition. At slightly larger energies (4.41 eV, 281 nm), the charge-transfer excitation from the pyrimidine to the thiazolium ring was found. The DFT calculations result in a different order of the two states. Here, the charge-transfer excitation comes first, followed by the $n \to \pi^*$ transition. In a first attempt to include environmental effects, CAM-B3LYP calculations were performed using the COSMO model, with toluene as the solvent. The latter is often used to replicate the conditions found in enzymatic active sites.[129] The $n \to \pi^*$ transition is only slightly increased to 4.67 eV (265 nm). Opposed to this, the charge-transfer state is shifted by 0.25 eV to lower energies, resulting in an excitation energy of 4.01 eV (309 nm). When the environmental effect computed at the CAM-B3LYP level is added to the CC2 values as a correction, the charge-transfer excitation corresponds again to the lowest excited state with a transition energy of 4.16 eV (298 nm) compared to a value of 4.34 eV (286 nm) for the $n \to \pi^*$ state. Furthermore, the oscillator strengths differ significantly between the two states. The charge-transfer excitation is 4 times as intense as the $n \to \pi^*$ transition (CC2 oscillator strength of 0.016 compared to a value of 0.004). The absorption spectrum of AP-ThDP in the low-energy region is therefore expected to be dominated by this charge-transfer excitation.

The lowest transition in IP-ThDP corresponds to a similar charge-transfer state as for AP-ThDP with an excitation energy of 4.36 eV (284 nm). The second excited state, a

$\pi \to \pi^*$ transition located at the pyrimidine ring, is well separated with an energy gap of roughly 0.5 eV to the charge-transfer state. All the applied DFT functionals (except B3LYP) perform well in comparison to the CC2 calculations. Altogether, the agreement between the reference values and the CAM-B3LYP results was satisfactory.

Having established a reliable level of theory, the focus was on the impact of the cofactor conformation on the electronic spectra. In contrast to the spectra in enzymes, no absorption was observed for free ThDP in aqueous solution at wavelengths >300 nm.[130] In order to clarify this observation, the first excited states of the AP tautomer were calculated in the gas phase and with the COSMO continuum model in the *V* and *F* conformation, which differ in the orientations between the aminopyrimidine and thiazolium rings.[131] The *F* conformation is the most stable conformation of free ThDP in solution. The structures and gas phase excitation energies are shown in Figure 4.4a and 4.4b for the *V* and *F* conformations, respectively. Spectra in different conformations and environments were simulated as Gaussian bands with a full width at half maximum (FWHM) of 0.3 eV (Figure 4.4c).

The lowest transition is as above mentioned, the charge-transfer excitation. However, the excitation energy is found to be highly dependent on the the medium and conformation. In the gas phase, the energy is lower in the *V* conformer (4.26 eV, 291 nm) than in the *F* conformer (4.41 eV, 281 nm). Moving to the COSMO results in toluene, which should approximately mimic the environment of a hydrophobic enzyme pocket, the energy of the *V* conformation drops significantly to 4.01 eV (309 nm) and shows a prominent band in the spectrum (blue curve in Figure 4.4c). This is well separated from a very strong $\pi \to \pi^*$ band of the thiazolium ring at about 4.96 eV (250 nm). However, moving to an aqueous solution for the *F* conformer of free ThDP, the excitation energy is strongly increased. The first transition is found at 4.66 eV (266 nm) and falls into the region of the much stronger $\pi \to \pi^*$ transition located at the thiazolium ring. Consequently, there was no significant absorbance in the spectra at wavelengths >300 nm, where ThDP-dependent enzymes show their typical CD band structures.

The previous results roughly confirm the nature and general features of the first absorption band for the AP tautomer of ThDP. However, the first results for the IP tautomer hint at a similar excitation as in the AP, of a charge-transfer nature. This finding disagrees with the currently accepted assignment. The model compound studies of Jordan and coworkers pointed toward an $n \to \pi^*$ assignment.[30] The latter authors made use of singly substituted pyrimidine compounds to assert the position of the IP band. The

(a)

$\omega_V^{\text{gas}} = 4.26 \text{ eV}$ $\qquad$ $\lambda_V^{\text{gas}} = 291 \text{ nm}$

(b)

$\omega_F^{\text{gas}} = 4.41 \text{ eV}$ $\qquad$ $\lambda_F^{\text{gas}} = 281 \text{ nm}$

(c)
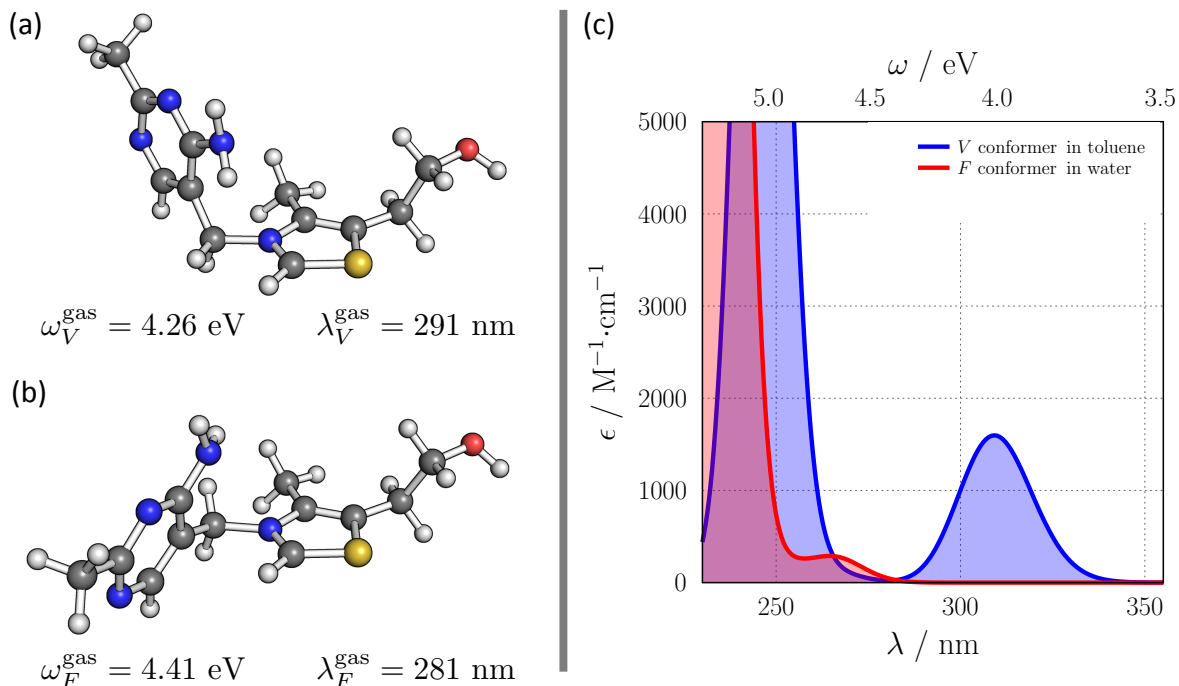
Figure 4.4.: B3LYP-D3(BJ)/def2-TZVP optimized structures for the AP-ThDP model compounds in the (a) *V* and (b) *F* conformation. Additionally, the lowest gas phase excitation energies and the corresponding wavelengths at the CAM-B3LYP/def2-TZVP level of theory are given for both tautomers, associated with a charge-transfer from the pyrimidine to thiazolium rings. (c) Simulated absorption spectra for the AP-ThDP model compounds for the *V* conformer (blue curve) and the *F* conformer in aqueous solution (red curve). The peak positions were represented as Gaussians with a FWHM of 0.3 eV.

nature of the latter was deduced from the measured solvatochromic shifts and the fact that only the pyrimidine ring is in fact represented in the model compounds used. In order to understand this disagreement, these compounds were studied in light of the recent calculations.

TD-DFT/TDA calculations were carried out on one of the IP analogues, N^1,N^4-dimethyl-1,4-iminopyrimidine (dmIP), featured in the study of Jordan and coworkers (see Figure 4.5a).[30] The results are shown in Figure 4.5b, considering the same three solvents as in the latter work and the two possible isomers of the analogue. No agreement between the computed values for both isomers can be found. The largest deviations are found with THF as the solvent. The relative trends for both lowest transitions ($\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$) are also inconsistent with the measured data for varying solvents. The computed values show almost negligible solvatochromic shifts, while experimentally de-

termined absorption bands are shifted as much as 0.24 eV (comparing DMSO and water). Even if the band positions were not in agreement with the measurements, one should at least expect the shifts to be consistent. This raises serious questions about the band assignment. The failure also does not seem to stem from the level of theory used. EOM-CCSD/A'VDZ, PBE0,[132] and CAM-B3LYP results in the gas phase all agree closely in terms of the computed excitation value (282, 302, and 283 nm, respectively).

Another open possibility would be the formation of other species in solution, which would then be responsible for the recorded UV−vis signals. In particular, the formation of dimers in solution was considered as possible candidates, which could result in excitations similar to those observed in the previous section for the IP cofactor. The most promising candidate for the compound under study would be a protonated N^1,N^4-dimethyl-4-aminopyrimidinium (similar to APH$^+$), building a dimer in solution with an unprotonated molecule, effectively sharing the charge (which in turn reduces the pKa in solution). An electron transfer from the unprotonated ring to the protonated species could then occur in solution. The geometry was optimized in a π-stacked conformation at the B3LYP-D3/def2-TZVP level of theory, including the COSMO solvation model with the corresponding solvents. One of these π-stacked dimer structures is shown in
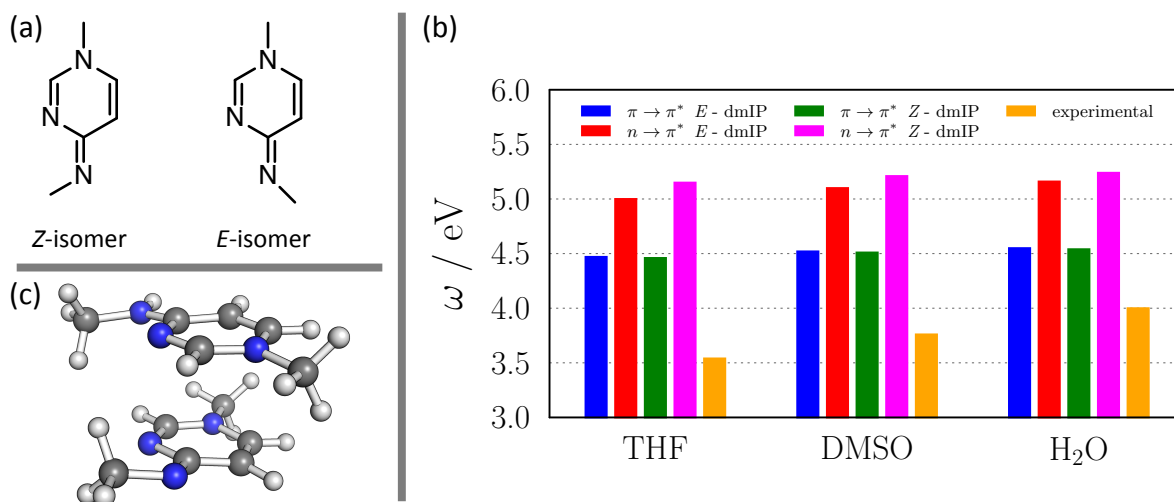
Figure 4.5.: (a) Structural formulas of two isomers of N^1,N^4-dimethyl-1,4-iminopyrimidine (dmIP) featured in the study of Jordan and coworkers.[30] (b) CAM-B3LYP/def2-TZVP vertical excitation energies of the two isomers of dmIP including the COSMO continuum model. The experimental band maxima were taken from Ref. [30]. (c) B3LYP-D3(BJ)/def2-SVP optimized structure of a π-stacked dimer between dmIP and its conjugated acid.

Figure 4.5c. The values for the first electronic transitions are 3.82, 3.93, and 4.04 eV (325, 315, and 307 nm, respectively), in THF, DMSO, and water, respectively. The agreement is striking. An analysis of the NTOs reveals that it is in fact, as expected, an electronic transition from a ring to the other, with the protonated species effectively mimicking the thiazolium ring of ThDP. This could explain the overall coincidence of this study with the spectra recorded in functional enzymes. The trend in solution could also be understood through the charge-transfer character.

4.2.2. Influence of Enzyme Residues on Cofactor Spectra

To determine the impact of neighboring residues in the modulation of the transition energies in the cofactor, a perturbation theory analysis of the excitation was performed. This includes 32 amino acids around the cofactor in *Zm*PDC, corresponding roughly to a 5 Å shell (see Figure 4.6). A few points should be mentioned about the setup, because several ionization sites are present in the pocket. In Figure 4.6a, these residues are shown together with ThDP. Residue Glu473 was considered to be protonated. This decision was based on mutagenesis data (in particular for variant Glu473Gln) and structural analyses of ZmPDC.[103;104] In a comparison of the CD spectra of the wild-type enzyme and its Glu473Gln mutant, no significant shift was observed in the excitation energies, indicating that the charge of the mutated residue has probably not been changed. Residue His414 was considered to be doubly protonated, while His114 was set to be monoprotonated at the Nδ atom. The only difference in the MM shell between the two tautomers was in the protonation of the Glu50 residue. Because the latter is a prime candidate for the transfer of the proton to the aminopyrimidine ring, this residue was initially considered to be protonated (uncharged) in the AP tautomer but ionized (negatively charged) in the IP calculations. As one will later observe, such an assumption can strongly influence the relative placement of the transition energies. The results are shown in Figure 4.6b.

As one can observe, there is a limited list of residues that strongly modulate the excitation energy for both tautomers. His414 lays directly above the pyrimidine ring and was considered hereby as being doubly protonated (positively charged). This assignment was based on its placement relative to Asp82, which also contributes to a shift in the transitions, but with a different sign. The positive charge in the histidine stabilizes the π-system of the pyrimidine, while this is counteracted by the negative charge in the aspartate. However, the magnitude of the effect of His414 is approximately twice that

of the shift resulting from Asp82. The Glu50 residue provides the third most prominent feature in the analysis. As previously mentioned, this residue was considered to be the donor of the proton to N1' for IP formation, leaving it uncharged in the AP case and charged (unprotonated) in the IP tautomer. This choice of setup has a strong impact on
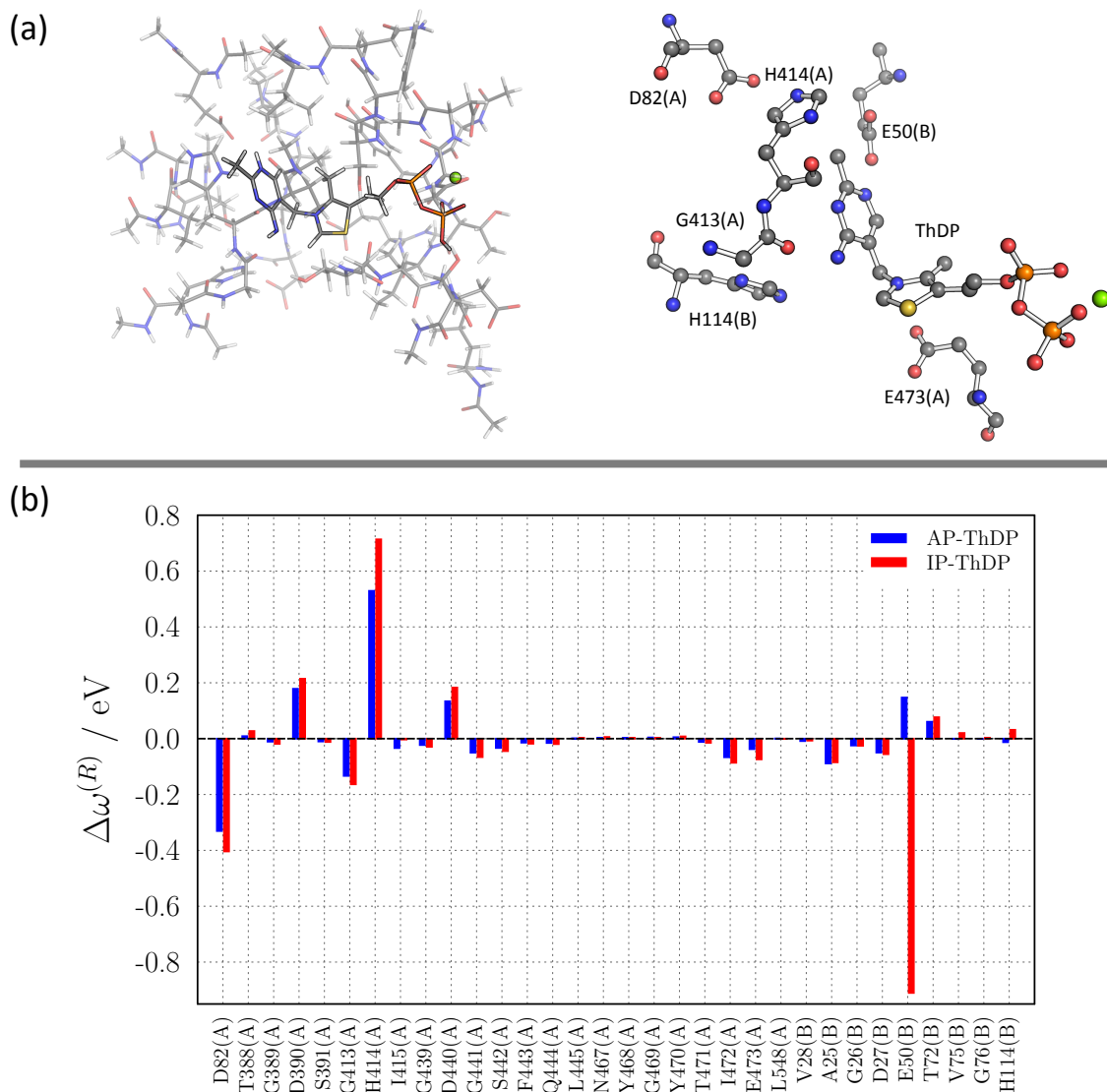
Figure 4.6.: (a) Active site of *Zm*PDC. Left: IP-ThDP in the complete amino acid shell. Right: Ball-and-stick representation of ThDP and selected enzyme residues. (b) Perturbative residue-by-residue analysis of the impact on the charge-transfer excitations of AP and IP-ThDP in *Zm*PDC.

the magnitude of the shifts. In the AP case, it has barely any effect on the transition. In the IP tautomer, it strongly lowers the energy.

The information from this first set of calculations can be summarized as follows. His414 was tentatively taken to be doubly protonated, given the uncertainty in its assignment. The results show that such a protonation would strongly shift the absorption energies to higher values. The surrounding residues do not completely compensate for the positive charge next to the pyrimidine ring, and the latter will necessarily increase the energy penalty for an electron transfer to the thiazolium ring. For the simulation of ThDP in the whole enzyme, His414 was singly protonated at $N\epsilon$, as this shift was deemed unlikely.

The protonation state of Glu50 is critically determining the position of the cofactor absorption bands. The excitation energies taken from the simple model amount to 3.79 and 4.74 eV (327 and 262 nm, respectively) for the IP and AP tautomers, respectively. Given that the character of excitation is similar between the two, it is logical that the second shell of residues should not change the relative placement of the two bands. This, however, disagrees with all experimental observations which place the AP transition slightly below that of IP. One possible explanation would be that an assignment of Glu50 as being uncharged is invalid. This is not at all unlikely, considering that the glutamate could simply serve as a proton relay but not necessarily change its protonation when shifting from IP to the AP form. The X-ray structural analysis of *Zm*PDC revealed that Glu50 directly interacts with two ordered water molecules, and these could be engaged in reversible proton transfers.[103;104] Taking the same shift as that observed for the IP case, one would obtain an excitation energy of 3.68 eV (337 nm), much closer to the measured value of 3.65–3.87 eV (320–340 nm).[23;104] Therefore, in the following calculations both protonation states of the Glu50 residue were considered in the AP case in order to understand its effect in more detail. In the IP tautomer this residue was kept to be deprotonated (charged). All other protonation states of neighboring residues maintained the same for both AP and IP, with the exception of His114. The latter is the main candidate for the role of proton acceptor in the tautomerization process. The reason for this assignment was deduced from the analysis of the molecular dynamics simulations, described in the next section.

4.2.3. Spectra of AP and IP-ThDP in *Zm*PDC

Models of a functional dimer of the *Zymomonas mobilis* PDC (chains A and B) were simulated for total production times of 1 ns, including two active sites. An AP and IP-ThDP were considered in the latter in order to reflect the chemical asymmetry observed in experiments. Two separate simulations were performed differing in the protonation state of the canonical glutamate of the AP-ThDP active site. The only other difference between the AP and IP-ThDP pockets was in the protonation state of His114 (doubly protonated in the IP case, monoprotonated in the AP occupied site) on both simulations. The reason for this was its possible role as a proton acceptor in the tautomerization process. In Figure 4.7a a snapshot of the AP-ThDP pocket is shown. The free protonation site of His114 points towards the amino group, so that a proton transfer, required for the formation of IP-ThDP, should be feasible.

The residue that was a particular focus of the analysis of the simulation was Glu50, which when modeled in its standard ionic state should experience a repulsion from the nearby pyrimidine ring N1' atom in the AP pocket. At the AP site, Glu50 adopts a conformation slightly off-plane relative to the pyrimidine, reducing the level of repulsion, but still in close contact (see Figure 4.7a). Otherwise, the dynamics of the nearest
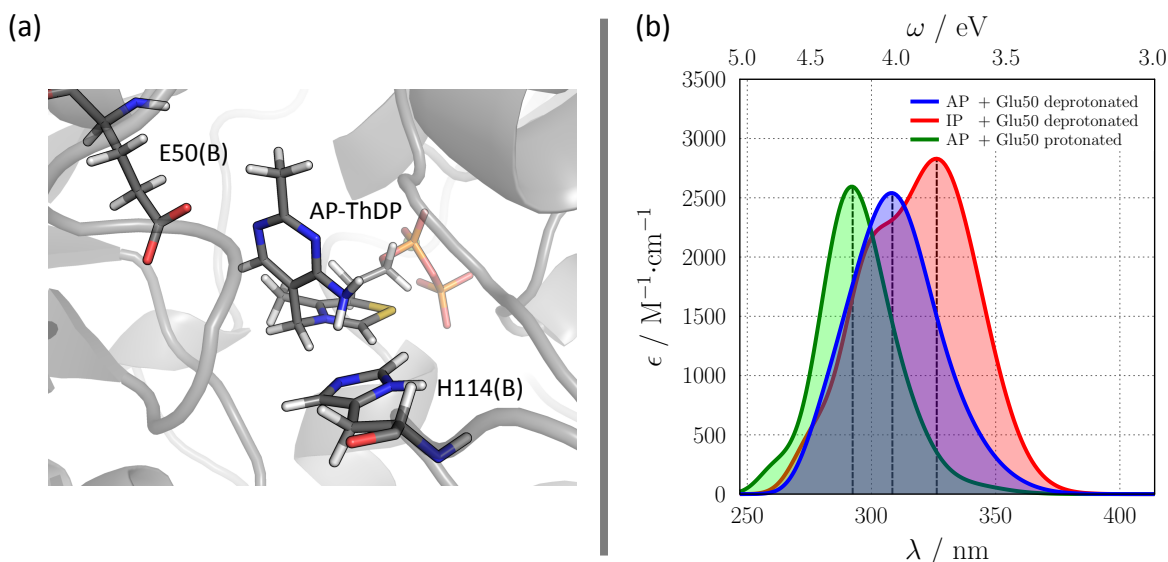
Figure 4.7.: (a) Snapshot of the MD simulation of the AP occupied enzyme pocket with the canonical glutamate E50 (deprotonated) and H114 as possible proton acceptor in the tautomerization process to IP-ThDP. (b) Simulated absorption bands for the charge-transfer excitations of AP and IP-ThDP, considering different protonation states of the canonical glutamate.

residues was comparable. This gives us further confidence in the perturbative analysis of the residue specific shifts performed beforehand.

From the MD run with a deprotonated (charged) Glu50 at the AP occupied site, 50 snapshots were taken in 20 ps intervals, to avoid the use of correlated structures. The latter were taken directly from the simulation, without any further optimization step. In this way, the thermal effect on the structures was included in the final results for the absorption bands. Test optimizations at the QM/MM level, applying B3LYP-D3/def2-SVP, showed only small changes but would, as expected, lead to higher excitation energies ($\leq$0.5 eV). The directly sampled MD structures tend toward elongated bonds that contribute to a destabilization of the ground state relative to the excited state. The first electronic excited states were computed at the CAM-B3LYP/def2-TZVP level, including all MM atoms inside 45 Å large spheres centered on the Mg^{2+} cations at the diphosphate tails. Besides the cofactor under study and its associated Mg^{2+} cation, the corresponding Glu50 side chain was incorporated into the QM region, because the perturbation analysis has shown its crucial importance for the charge-transfer excitations. The peak positions are represented by Gaussian curves centered at the same point, with a full width at half-maximum of 0.3 eV.

For both tautomers, the only excitations with significant oscillator strengths in the higher-wavelength regime were the charge-transfer states already observed in the model calculations. The simulated spectra for these transitions are shown in Figure 4.7b (blue and red curves for AP and IP-ThDP, respectively). The $n \rightarrow \pi^*$ transition of AP-ThDP, discussed for the model calculations, was found to be 0.4–0.7 eV higher in energy, and rather consistently with a 1 order of magnitude smaller oscillator strength. The nature of the excitation, by inspection of the NTOs, is also consistent with the result obtained with the isolated cofactor (charge-transfer). However, the peak positions are strongly influenced by the enzyme environment. The AP cofactor shows a well-defined maximum position at 4.02 eV (308 nm), while the IP peaks at $\approx$3.80 eV (326 nm). In this latter case, a small shoulder is found toward the high-energy regime, at $\approx$4.13 eV (300 nm). A variation in the sample size (increasing the number of configurations to 75) did not lead to a disappearance of this shoulder, and no obvious correlation between the structure and these higher excitation values could be asserted.

In the experimental CD spectra, the AP and IP signatures were assigned to the spectral regions of 3.65–3.87 eV (320–340 nm, AP) and 4.00–4.13 eV (300–310 nm, IP), respectively.[23;104] The simulated absorption spectra agree well with these measurements.

For both tautomers, significant intensities were observed in the expected energy ranges, caused by charge-transfer excitations. Admittedly, the calculated peak maxima do not perfectly match the experimental CD band maxima, but this should have other rational explanations. In particular for enzymatic systems, the oscillator strengths do not necessarily correlate with the rotational strengths, responsible for the CD signatures. Besides the intrinsic contribution from the transition of the cofactors, the excitonic coupling to the excitations of the protein residues can be of major importance for the latter.[133;134] This influences the orientation between the electric and magnetic transition dipole moments, resulting in variations in the rotational strengths up to a change in its sign. However, the oscillator strength, determined only by the magnitude of the electric transition dipole moment, and the calculated excitation energies are affected to a minor extent. The most important fact is that transitions with significant intensity occur at the CD band positions. This was shown by our simulated spectra for both AP and IP-ThDP.

The MD run with a protonated Glu50 in the pocket of the AP cofactor was comparable to the other simulation, except that this residue hydrogen bonds to the N1' atom via its carboxyl. The excited states were calculated using the same protocol as before. Once more, the charge-transfer excitation dominates the higher-wavelength regime. The simulated spectrum is also shown in Figure 4.7b (green curve).

The simulated band holds an intensity and a shape similar to those of AP with deprotonated Glu50 (negatively charged) in its pocket. However, the peak position is significantly shifted to higher energies by 0.22 eV, resulting in a band maximum at 4.24 eV (292 nm). Consequently, much less intensity is observed in the spectral region of 320–340 nm, and the shift to the IP band becomes even larger. This indicates that AP-ThDP together with a deprotonated Glu50 (negatively charged) should be responsible for the measured CD bands found for *Zm*PDC. In fact, the simulated spectrum of AP with uncharged (protonated) Glu50 almost perfectly matches the hitherto unassigned band at 280–290 nm, observed in the revisited spectra of *Zm*PDC (see Figure 4.2 in the introduction of this chapter). The calculated shift of 0.22 eV for the AP-ThDP cofactor with the different ionization states of Glu50 is also in reasonable agreement with the experiment ($\approx$0.4 eV). Further strong evidence for these assignments are gained through the interpretation of the Glu50Gln mutant spectra in light of the calculations. Substitution of the carboxyl with a carboxyamide group should prevent the N1' atom to be protonated, since it hydrogen bonds to the amine group of this residue which pK_a is too

high for a proton transfer. Consequently, the Glu50Gln mutant of *Zm*PDC should only be capable to stabilize the AP tautomer of ThDP where the glutamine approximately represents a protonated (uncharged) Glu50 residue. This is exactly what is observed in the experimental CD spectrum. The only absorption which remains is the negative CD band in the spectral region of 280–290 nm, representing the charge-transfer excitation with a neutral residue at the position 50 as proposed by the calculations.

Summarizing the results of the calculations in light of the experimental observations, the spectral signatures of the pre-equilibrium states of ThDP-dependent enzymes can now be assigned to the different tautomers of ThDP including the protonation state of the canonical glutamate. This is shown for the activation process of ThDP in Figure 4.8. The AP-ThDP cofactor is present for both protonation states of the canonical glutamte, being responsible for two absorption bands in the near-UV spectral region. It tautomerizes then either via its APH$^+$ state or through a coupled proton transfer to the IP form which shows then a positive CD band in the spectral region of 300–310 nm. An intramolecular proton tranfer from this state produces then the activated ThDP in its carbene form.
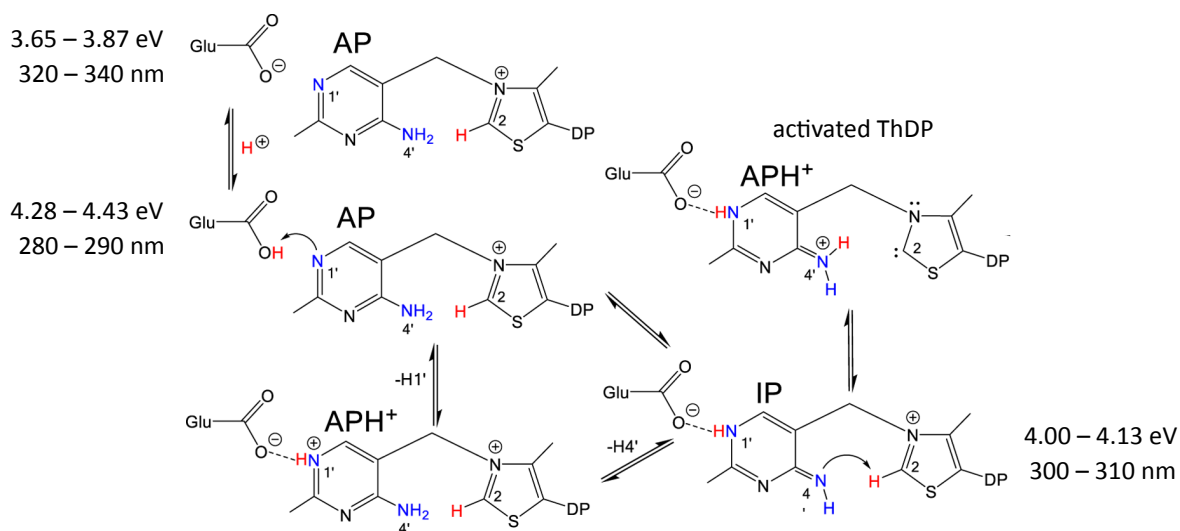
Figure 4.8.: Scheme for the activation of ThDP in enzymes via its pre-equilibrium states together with the spectral assignments of the near-UV CD bands.

4.3. Conclusions

The spectral signatures of the pre-equilibrium states of the ThDP cofactor, both in solution and in the active site of the *Zm*PDC enzyme were studied through the use of a variety of methods. These include DFT, CC calculations and hybrid QM/MM approaches, analyzing the ThDP cofactor in different environments. Additionally, a perturbative analysis of the enzyme environment on the absorption spectra was performed. The main findings can be summarized as follows.

(1) The spectra of free ThDP, particularly its disagreement with the band position in enzymes, can be explained by the change in the conformation as well as in the polarization of the surrounding aqueous environment.

(2) Both AP- and IP-ThDP tautomer signature bands, commonly targeted in biochemical studies, correspond to charge-transfer excitations between the pyrimidine and the thiazolium rings.

(3) An unassigned band found in the spectral region of 280–290 nm has been linked to a change in the protonation state of the canonical glutamate.

(4) The latter residue (Glu50) was highlighted with a fundamental role for the spectra by the perturbative analysis scheme and confirmed by QM/MM calculations on a model dimer of *Zm*PDC. It serves most likely as a proton relay channel but is not necessarily protonated in the presence of AP-ThDP. The exiting proton is most likely further transferred through the hydrogen bond network.

The consistency of the 300 nm IP band observed both in the resting-state enzyme and in tetrahedral substrate−ThDP or product−ThDP conjugates needs not to be attributed to a pyrimidine-localized excitation. The charge-transfer state that was identified will be minimally affected by the substitution of a hydrogen atom at position C2 by a sp^3-hybridized carbon. It could also provide an explanation for the spectra recorded in other substituted intermediates. In the case in which the C2-bound moiety changes the degree of conjugation in the thiazolium ring, the band is expected to be accordingly shifted. Such a pattern can be observed, for example, in 2-acetyl-ThDP, which is subject of study in the next chapter.

5

Phosphoketolase

Phosphoketolase (PK) is a ThDP-dependent enzyme, catalyzing the phosphorolytic cleavage of fructose-6-phosphat (F6P) or xylulose-5-phosphate (X5P) into acetyl phosphate, water and erythrose-4-phosphate (E4P)/glyceraldehyde-3-phosphate (G3P), respectively.[7;8] It holds a fundamental role in the lactic acid fermentation, where glucose is converted into cellular energy in the form of adenosine triphosphate.[135] Based on kinetic measurements a catalytic mechanism of PK was already proposed by Yevenes and Frey (see Figure 5.1a).[136] The activated ylide-ThDP performs a nucleophillic attack at the ketose to build a substrate-ThDP conjugate. After C–C bond cleavage, the aldose product is released and 1,2-dihydroxyethyl-ThDP (DHEThDP) is formed. Dehydration of this intermediate leads to 2-acetyl-ThDP (AcThDP) with three different tautomeric states: keto-AcThDP, enol-AcThDP and enolate-AcThDP. The keto form then reacts with the inorganic phosphate (or water for off-pathway hydrolysis), so that afterwards acetyl phosphate is released.

The Tittmann group investigated this system in further detail. Their main observations can be summarized as follows:[40]

(1) Crystallographic measurements of PK from *Bifidobacterium breve* (*B. breve*) exposed with F6P revealed the AcThDP intermediate, which is trapped by refusing the co-substrate phosphate (see Figure 5.1b). With a resolution of 1.6 Å, the hydrogen positions cannot be determined, so that the state of this intermediate is unclear.

(2) UV-vis measurements showed an absorption band at 2.95 eV (420 nm) by mix-

ing PK with the substrate F6P. This spectrum is shown in Figure 5.1c. If also the second substrate (phosphate) was added, then this absorption band was not observed. The absorption band stems most likely from the AcThDP intermediate.
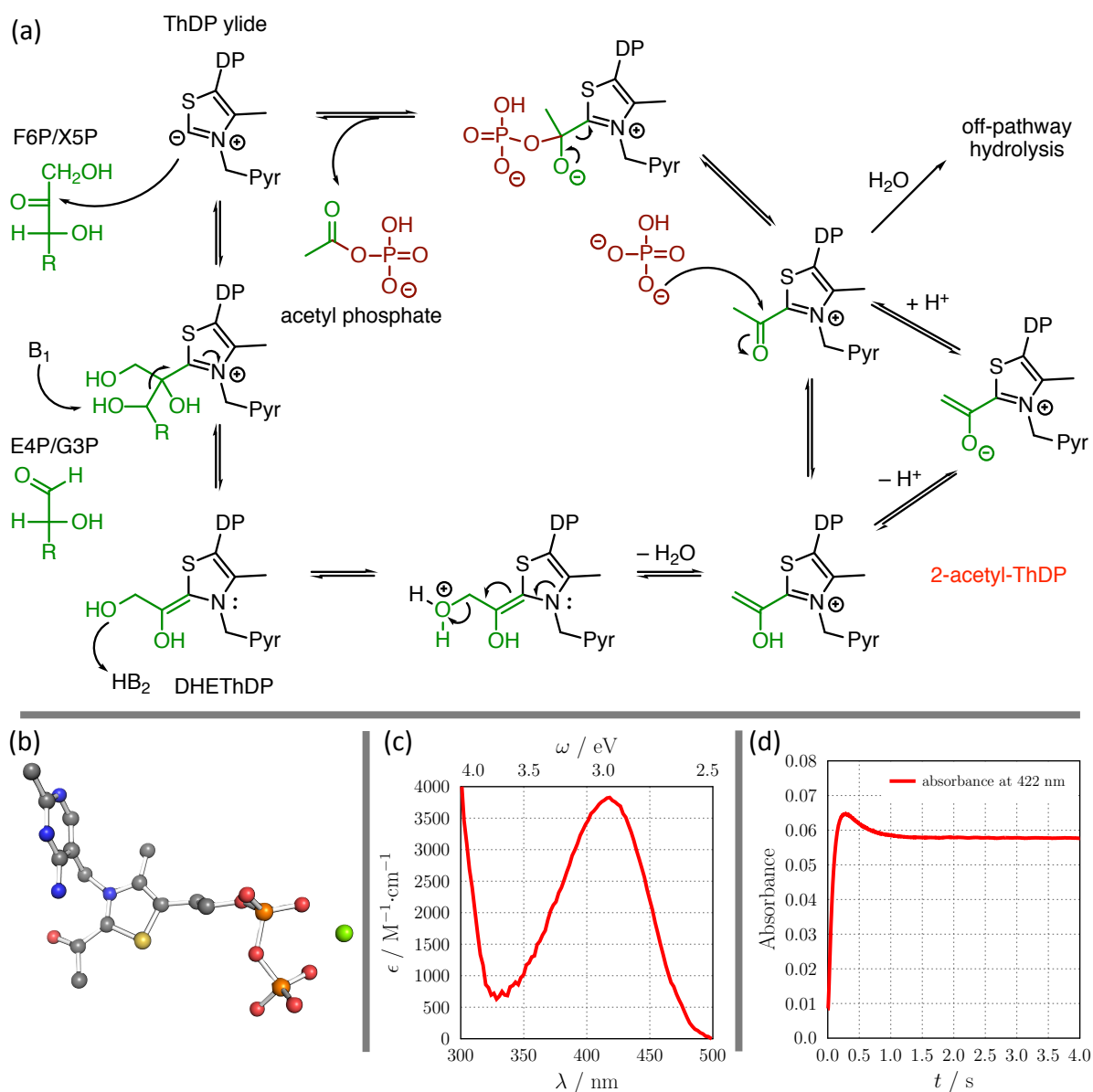
Figure 5.1.: (a) Proposed mechanism in the reaction of PK with F6P/X5P and phosphate based on the experiments of Yevenes and Frey.[136] (b) Crystal structure of the AcThDP intermediate in *B. breve* PK at 1.6 Å resolution. (c) UV-vis spectrum of *B. breve* PK after addition of F6P. (d) Time-resolved absorbance at 422 nm for *B. breve* PK after addition of F6P. Experimental data for the graphics (b)–(d) were provided by the Tittmann group from Göttingen.[40]

(3) Time-resolved spectral measurements at 422 nm showed a kinetic overshooting. This indicates that an intermediate is formed after mixing PK with the substrate, which then converts into another one, resulting in an equilibrium with a kinetic constant of 4 s^{-1}.

(4) The catalytic rate constant in the reaction of PK with both substrates (F6P and phosphate) was determined with a value of 14 s^{-1}. This is in discrepancy to the slower process for the internal conversion (see point 3) by refusing the inorganic phosphate. This substrate should therefore have a significant impact on the reaction.

In this work the focus was on the spectral behavior of AcThDP in order to understand the observed absorption band and kinetic overshooting, resulting in the identification of the state of AcThDP. The role of the inorganic phosphate was investigated in the content of the Master's project of Sebastian Wille.[41] The results are featured in his thesis and here briefly discussed in the context of the spectra calculations.

5.1. Computational Details

Spectra calculations for AcThDP were carried out with model compounds, where the diphosphate moiety was substituted by a hydroxyl group. The pyrimidine ring was set to the AP state, which is the common choice for ThDP conjugates with a coplanar substituent.[23] The three possible states of AcThDP (keto, enol and enolate) were considered. The model compounds were optimized at the B3LYP-D3(BJ)/def2-TZVP level of theory[61;62;111–113] and confirmed as minima through a frequency analysis. For this purpose and all following DFT computations, the resolution of the identity (RI) approximation with the corresponding auxiliary basis set as well as the chain-of-spheres algorithm was employed to speed up the calculation of the Coulomb and exchange interactions, respectively.[114–116] Vertical excitation energies were then determined by TD-DFT, employing the CAM-B3LYP functional.[65] Various basis sets were tested for convergence of the excitation energies as well as the effect of the TDA approximation.[72] In order to include environmental effects, the COSMO solvation model was applied.[117] Hereby, the predefined settings were used from the ORCA program package for each solvent.[123]

The effect of residues on the spectra of AcThDP in the active site of the Glu437Gln mutant *B. breve* phosphoketolase was studied by the perturbative analysis scheme described

in Section 4.1.1. Here, the crystal structure was provided by the Tittmann group,[40] including AcThDP as intermediate by soaking PK with F6P. The intermediate was saturated with hydrogen atoms, considering its three tautomeric states. A residue shell of about 5 Å was taken, including 39 amino acids in total. Free residues were capped by acetyl or *N*-methyl groups. The AMBER ff10 force field was employed for the enzyme environment,[118] whereas the GAFF force field was used for AcThDP.[120] A MM optimization followed for the hydrogen atoms of the enzyme environment, thereby restraining all other atoms in their Cartesian coordinates by a harmonic potential with a force constant of $k = 5 \, \text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$. A QM/MM optimization was then carried out for the intermediates at the B3LYP-D3(BJ)/def2-SVP level of theory. Hereby, AcThDP and the Mg^{2+} ion were included in the QM region, whereas the enzyme environment was treated at the MM level of theory. Reference excitation energies and difference densities were obtained with TD-DFT/TDA calculations (CAM-B3LYP/def2-TZVP) for each state of AcThDP. Afterwards, the perturbative analysis scheme for the excitation energy shifts was applied. Additionally, other mutations of PK were tested with the objective to shift the absorption band at 2.95 eV (420 nm). For this purpose, the corresponding side chain of the residue under study was modified and optimized at the MM level. The excitation energy shifts upon the mutation were then determined by first-order perturbation theory (PT). The perturbation operator now reads

$$\hat{h}^{(R \to R')} = \sum_{J \in R} \frac{Q_J}{|\mathbf{r} - \mathbf{R}_J|} - \sum_{J \in R'} \frac{Q_J}{|\mathbf{r} - \mathbf{R}_J|} \, , \tag{5.1}$$

where a residue R is mutated to a residue R'. The excitation energy shift is then calculated by Equation (4.9), but now employing the matrix representation of the PT operator from above:

$$\Delta\omega_{0a}^{(R \to R')} = \text{Tr} \left\{ \mathbf{\Delta D}^{0a} \mathbf{h}^{(R \to R')} \right\} \, . \tag{5.2}$$

Here $\mathbf{\Delta D}^{0a}$ denotes the difference density matrix for an excitation from the ground state 0 to an excited state a. The results were then compared to full TD-DFT/TDA calculations at the CAM-B3LYP/def2-TZVP level of theory.

Program packages involved in the computations were ORCA versions 3.0.3 and 4.0.1 (DFT),[123] Amber 11 (MM)[124] and ChemShell (QM/MM).[80;125]

5.2. Results and Discussion

5.2.1. Spectra of AcThDP

The observed absorption band at 2.95 eV (420 nm) is most likely linked to the formation of AcThDP. In order to identify the state of this key intermediate in the PK pathway, spectra calculations were carried out for model compounds of AcThDP for its three different tautomeric forms: keto, enol and enolate-AcThDP. For this purpose, the pyrimidine ring was set to the AP state. This is the generally accepted state where the substituent at C2 is conjugated towards the thiazolium ring.[23] The phosphate group tail with its negative charge generally leads to ghost states, since it is not compensated by the environment. Therefore, the diphosphate moiety was replaced by a hydroxyl group just as in the case of unsubstituted ThDP. The resulting model compounds were then optimized in the *V* conformation, being the one adapted in the active site of enzymes.[13;131] In a first set of spectra calculations, TD-DFT calculations with the CAM-B3LYP functional were carried out in the gas phase, thereby benchmarking different basis set sizes and the effect of the TDA approximation. The choice of the functional was based on the success for its application to unsubstituted ThDP featured in Chapter 4. The results for the first excited states are given in Table 5.1.

All TD-DFT/TDA calculations give nearly constant results with increasing basis set sizes for the first excited states of the three tautomeric forms of AcThDP. The excitation energies seem to be converged already at the def2-TZVP basis, so that for further calculations this was the one of choice. The TDA approximation shows nearly no effect for keto- and enol-AcThDP, whereas the first excited state of enolate-AcThDP is blue

Table 5.1.: Vertical excitation energies for the first excited states of the three different tautomeric forms of AcThDP in the *V* conformation, employing TD-DFT with the CAM-B3LYP functional. Different basis sets were benchmarked as well as the effect of the TDA approximation. The energies are given in electronvolts (eV). The corresponding wavelengths in nanometers (nm) are given in parentheses.

TDA	basis set	keto-AcThDP	enol-AcThDP	enolate-AcThDP
no	def2-TZVP	3.09 (401)	3.92 (316)	2.84 (437)
yes	def2-TZVP	3.11 (399)	3.92 (316)	2.99 (415)
yes	aug-cc-pVTZ	3.11 (399)	3.94 (315)	2.99 (415)
yes	def2-QZVPP	3.11 (399)	3.93 (315)	2.98 (416)

shifted by 0.15 eV. The reason for this should be connected to the different excitation character (see NTOs in Figure 5.2). Keto and enol-AcThDP transitions correspond to a charge-transfer from the aminopyrimidine to the thiazolium ring delocalized towards the substituent. Hole and particle densities are therefore well separated and the TDA approximation is known to have a negligible effect for such transitions.[137] For the first excited state of enolate-AcThDP a different kind of charge-transfer is observed. Here, an electron is transferred from the substituent to the thiazolium ring. Hole and particle densities are now spatially closer, resulting in larger energies through the TDA approximation. However, the calculated excitation energies are still in reasonable agreement. Altogether, TD-DFT/TDA with CAM-B3LYP/def2-TZVP is a satisfactory method for the calculation of excitation energies of AcThDP.
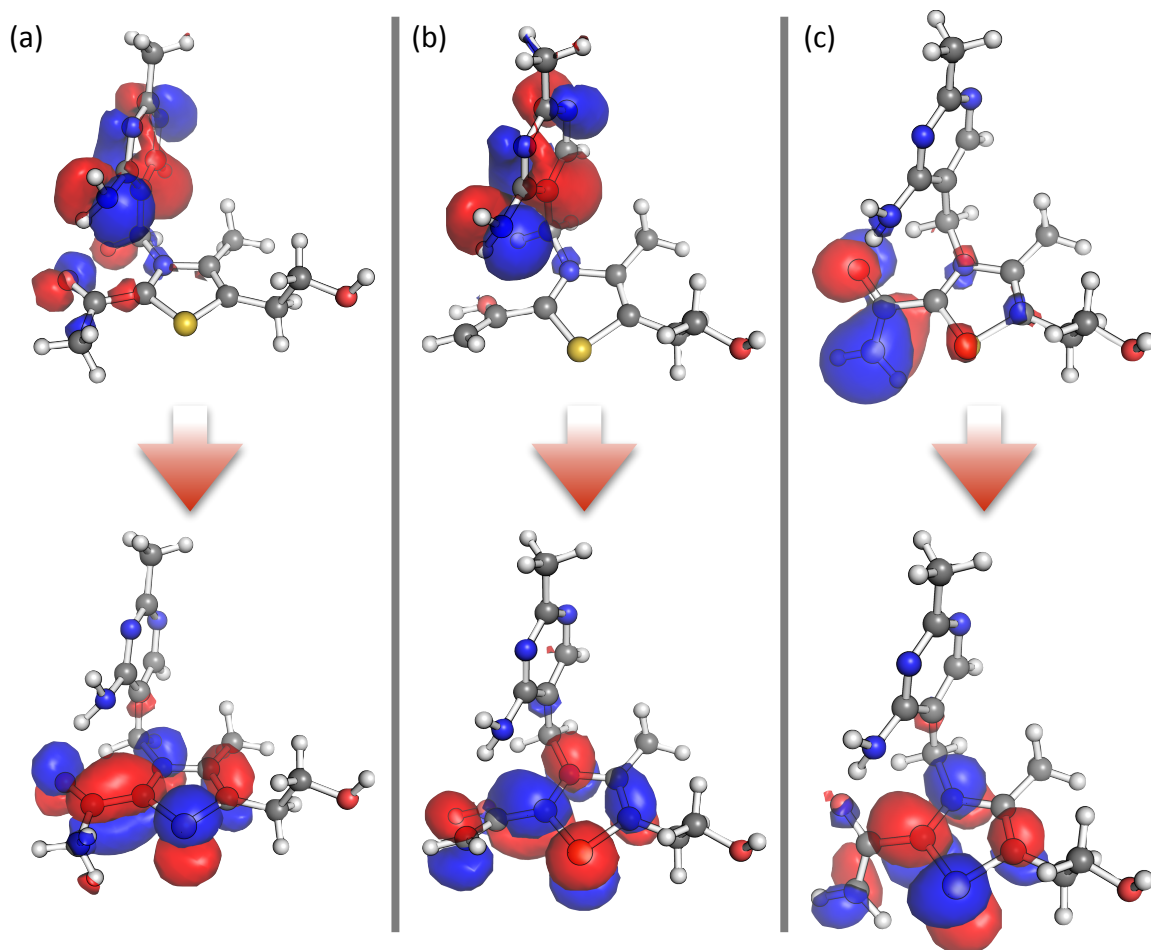
Figure 5.2.: Dominant CAM-B3LYP/def2-TZVP NTO pairs for the first excited states of (a) keto, (b) enol and (c) enolate-AcThDP.

Summarizing the gas phase results in light of the experimental observations, both the keto and enolate-AcThDP intermediates could be responsible for the band at 2.95 eV (420 nm) with calculated excitation energies of 3.11 and 2.99 eV (399 and 415 nm), respectively. Higher excited states are found well separated for both intermediates by at least 0.5 eV, making other contributions to the band unlikely. The lowest excitation energy for enol-AcThDP is significantly larger with 3.92 eV (316 nm) and falls out of the spectral region of interest. This can be rationalized by inspection of its associated NTOs in comparison to keto-AcThDP. Both intermediates share a similar occupied NTO, namely a π-orbital of the aminopyrimidine ring. However, the virtual NTOs differ in the delocalization to the substituent. In enol-AcThDP the carbon atom of the substituent contributes to the π-orbital, whereas the oxygen atom shows a larger role in keto-AcThDP. This causes the orbital energy to be raised for enol-AcThDP. In consequence, a larger excitation energy is observed for the enol state of AcThDP in comparison to keto-AcThDP.

As an attempt to include environmental effects the COSMO solvation model was employed. Toluene was chosen as solvent, since this should approximately mimic the hydrophobic enzyme pocket.[129] With that the charge-transfer states are slightly red shifted by 0.11, 0.27 and 0.11 eV to 3.00, 3.65 and 2.88 eV (413, 340 and 431 nm) for keto, enol and enolate-AcThDP, respectively. These results confirm the conclusions from the gas phase calculations, so that the observed band at 2.95 eV (420 nm) could originate from both keto as well as enolate-AcThDP. However, the computed oscillator strength in toluene is somewhat larger for the enolate state than for keto-AcThDP (0.11 compared to 0.04). These results could give a hint on the observed kinetic overshooting for the absorbance at 420 nm. A reasonable scenario would be a fast formation of enolate-AcThDP with a large absorbance, which then equilibrates to the less absorbing keto-AcThDP with a experimentally determined rate constant of 4 s^{-1} in the absence of phosphate.[40]

5.2.2. Influence of Conformation and Solvent on the AcThDP Spectra

One further point addressed in the spectra calculations was the difference in the observed bands of AcThDP in PK and free in aqueous solution, where a band at 4.00 eV (310 nm) was identified as main feature.[130] This time the effect is even more pronounced than

for unsubstituted ThDP with a shift of about 1 eV. The *S* conformation is known to be
the most stable one for free AcThDP in solution, differing in the orientations of the two
rings compared to the *V* conformation of enzyme-bound AcThDP.[131] Furthermore,
the keto form is expected to be the dominant tautomer in solution.[130] The B3LYP-
D3(BJ)/def2-TZVP optimized structures are shown in Figure 5.3 together with the
lowest gas phase excitation energies at the CAM-B3LYP/def2-TZVP level of theory.
The simulated spectra are also shown in Figure 5.3. Hereby, the COSMO solvation
model was employed, using toluene as solvent mimicking the hydrophobic active site of
PK and water for free AcThDP in its *S* conformation. The absorption bands were then
represented as Gaussian curves with a full width at half maximum of 0.3 eV.

(a)

$\omega_V^{\text{gas}} = 3.11 \text{ eV} \quad \lambda_V^{\text{gas}} = 399 \text{ nm}$

(b)

$\omega_S^{\text{gas}} = 3.39 \text{ eV} \quad \lambda_S^{\text{gas}} = 365 \text{ nm}$

$\omega_S^{\text{gas}} = 3.78 \text{ eV} \quad \lambda_S^{\text{gas}} = 328 \text{ nm}$
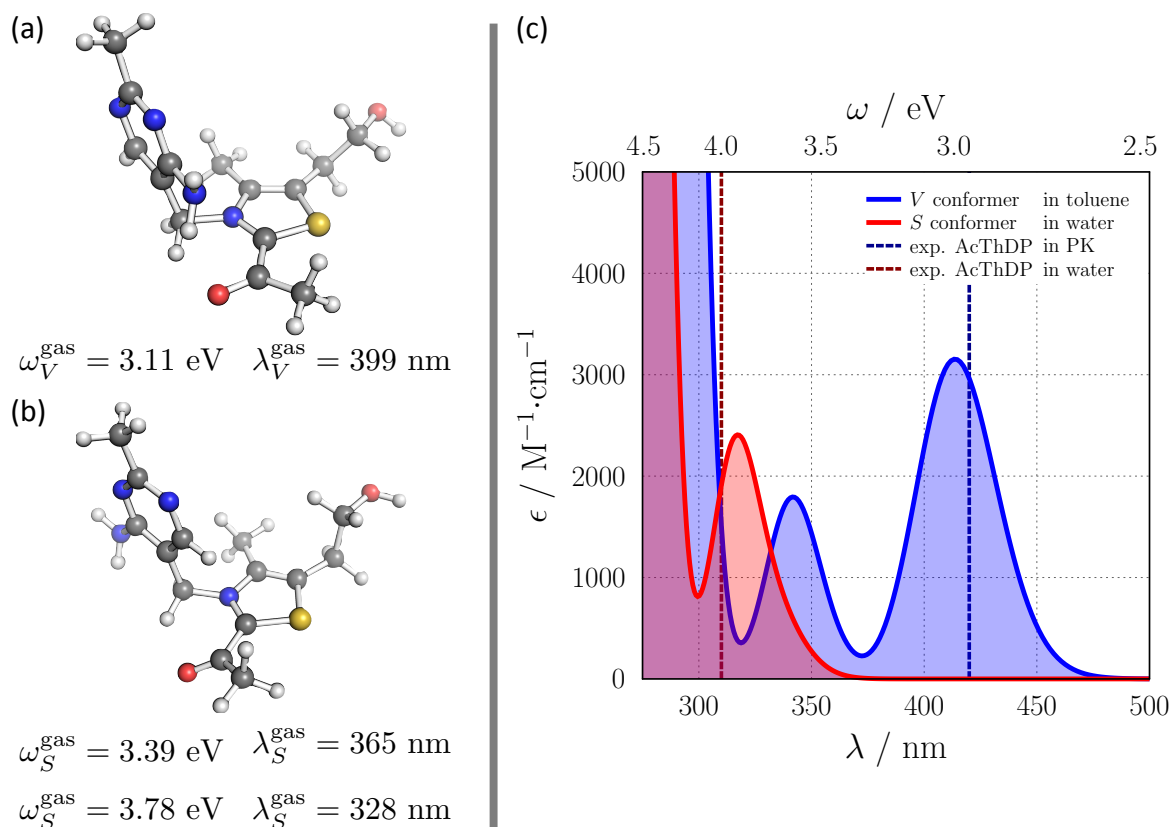
Figure 5.3.: B3LYP-D3(BJ)/def2-TZVP optimized structures of keto-AcThDP in (a) the *V* conforma-
tion observed in enzymes and (b) the *S* conformation for free AcThDP in solution. Addi-
tionally, the lowest gas phase excitation energies are given at the CAM-B3LYP/def-TZVP
level of theory. (c) Simulated absorption spectrum of keto-AcThDP for the *V* conforma-
tion in toluene (blue curve) and for the *S* conformation in aqueous solution (red curve) The
peak positions were represented as Gaussians with a FWHM of 0.3 eV. The positions of the
experimentally observed band maxima are shown as dashed lines.

The first and second excited states of AcThDP in the *S* conformation show similar charge-transfer characters as in the *V* conformation, where an electron is transferred from the pyrimidine ring to the thiazolium ring delocalized toward the acetyl group. Notably, the gas phase excitation energies are significantly larger with 3.39 and 3.78 eV (365 and 328 nm) compared to the conformation adopted in enzymes (3.11 eV, 399 nm). Moving to an aqueous solution, both excited states are further raised by 0.30 and 0.14 eV to 3.69 and 3.92 eV (336 and 317 nm). Since the second excited state has a significantly larger oscillator strength than the first one (0.026 compared to 0.005), this is the dominating excitation in the UV-vis spectrum at large wavelengths (see the red curve in Figure 5.3c). Employing toluene as solvent for keto-AcThDP in the *V* conformation, the band peaks at 3.00 eV (413 nm). The calculated shift of 0.92 eV between enzyme-bound AcThDP and free AcThDP in aqueous solution is in good agreement with the experimental observations. Therefore, the uncommon large shift is connected to a change in the conformation as well as to the different polarization of the environment.

5.2.3. Influence of Enzyme Residues on the AcThDP Spectra

The AcThDP spectra were studied in the enzyme environment of *B. breve* PK with the perturbation scheme also applied for unsubstituted ThDP (see Chapter 4). However, the aim for the application differed this time. Since keto and enolate-AcThDP absorb in the same spectral region with different excitation characters, the perturbative approach was employed to identify residues having an opposite impact on their spectra. This was then used for the search of possible mutants of PK which would be capable of splitting the absorption band and in turn directly lead to an observation of the two species. For this purpose, the unchanged crystal structure of the Glu437Gln mutant of *B. breve* PK with the AcThDP intermediate in the active site was provided by the Tittmann group.[40] This mutant showed the same spectral properties as the wild-type enzyme, having an absorption band at 2.95 eV (420 nm). In total, 39 residues around the intermediate was considered, roughly corresponding to a 5 Å amino acid shell (see Figure 5.4a). Titratable residues were set to their standard ionic states with two notable exceptions. The canonical glutamic acid (Glu479) was set protonated (neutral), forming a hydrogen bond to the N1' atom of the aminopyrimidine ring. His553 had different protonation states for the tautomers of AcThDP, since it most likely acts as acid-base catalyst in the tautomerization process.[138] For this reason, His553 was set monoproto-

nated at its Nδ atom for keto and enol-AcThDP, whereas a doubly protonated (positively charged) state was assumed in case of enolate-AcThDP. TD-DFT/TDA calculations with CAM-B3LYP/def2-TZVP were then carried out for the lowest charge-transfer excitation energies of the three tautomers in the amino acid shells and served as reference values for the subsequent perturbative scheme. The results are shown in Figure 5.4b.

The results of the perturbative analysis of the enzyme environment indenfies only a few
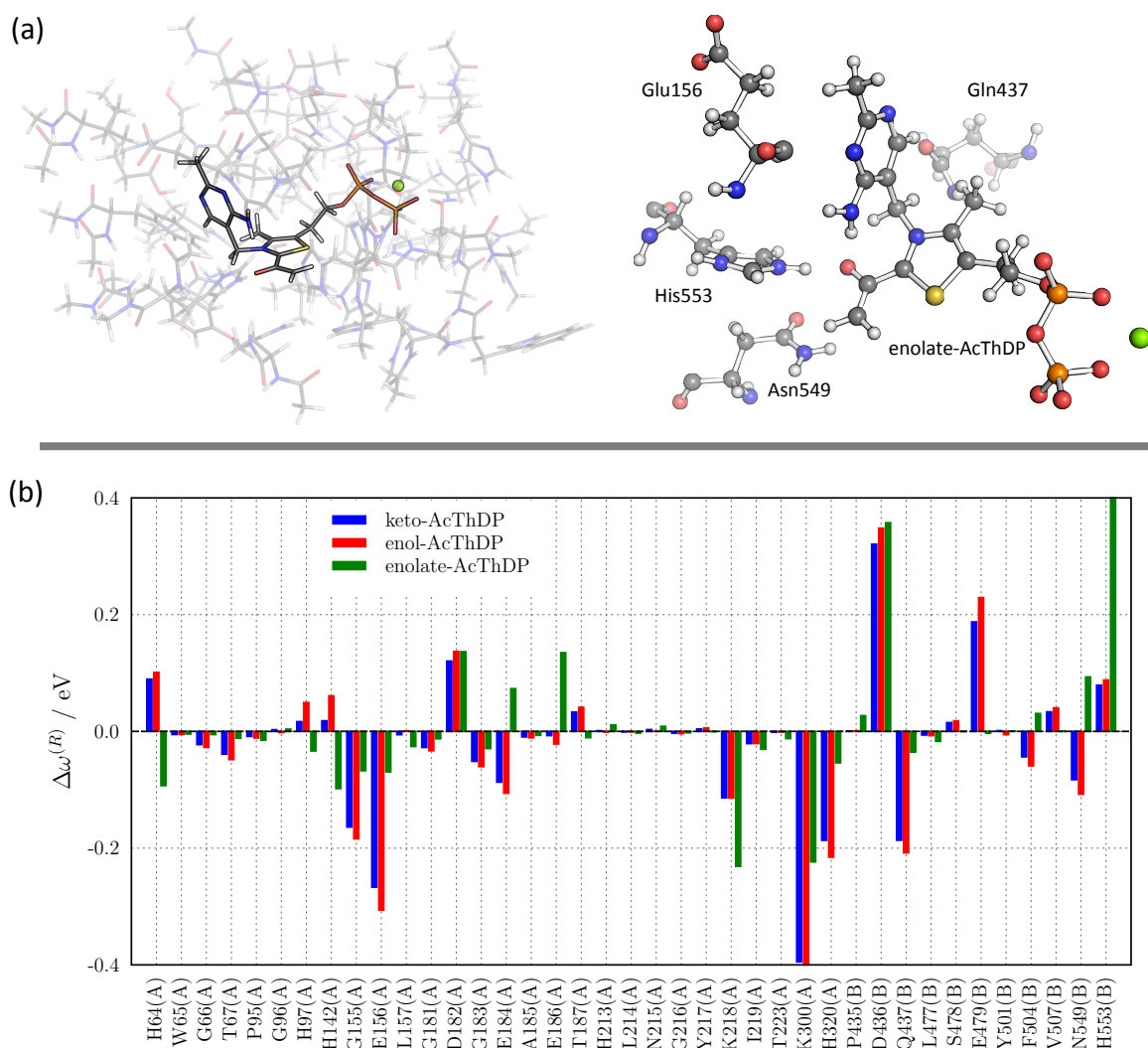
Figure 5.4.: (a) Active site of the Glu437Gln mutant of *B. breve* PK. Left: Enolate-AcThDP in the complete amino acid shell. Right: Ball-and-stick representation of enolate-AcThDP and selected enzyme residues. (b) Perturabative residue-by-residue analysis of the impact on the lowest charge-transfer excitations of keto, enol and enolate-AcThDP in the Glu437Gln mutant of *B. breve* PK.

residues significantly influencing the spectra of the tautomers of AcThDP. Most of them are charged and therefore have a stronger impact. Furthermore, keto and enol-AcThDP show the same trends for the residue analysis due to their similar charge-transfer excitation. Here, the interest is on keto-AcThDP, since it contributes to the UV-vis band at 2.95 eV (420 nm). The discussion is therefore restricted to this species. The impact of the residues for enolate-AcThDP differ in the amount and even the sign in comparison to keto-AcThDP. This arises from the different charge-transfer excitation. For the search of possible mutants which can split the observed absorption band, the residues showing the strongest deviations are required. There are three prominent residues with this feature, which will be discussed in further detail.

The first residue with a large deviation for the contribution to the excitation energies of the two intermediates is His553. It raises strongly the excitation energy for enolate-AcThDP, but has only a minor impact on the excitation of keto-AcThDP. The reason for this is the different protonation state of the histidine residue in the two cases. For enolate-AcThDP, His553 was set doubly protonated (positively charged). Its location directly next to the excitation hole leads to a strongly increased excitation energy. In case of keto-AcThDP, His553 was assumed to be monoprotonated (neutral) and is now close to the particle density of the excitation. Therefore, the effect is much less pronounced. However, His553 is expected to be involved as acid-base catalyst in the tautomerization process.[138] Mutation of this residue could lead to a different reaction mechanism, so that the residue is excluded as possible mutant candidate.

The second residue showing the desired feature is Glu156. This residue is negatively charged and located in the proximity of the aminopyrimidine ring of AcThDP. This decreases the excitation energy for the keto tautomer by about 0.3 eV. For enolate-AcThDP nearly no effect is observed since Glu156 is sufficiently far apart from the excitation space (located at the thiazolium ring and the substituent). Additionally, the Glu156 residue is not known to be directly involved in the reaction of PK and can be considered as mutant candidate.

The last residue in this analysis is Asn549, which is located below the acetyl moiety (see Figure 5.4a). This residue raises the excitation energy for keto-AcThDP by increasing the energy of the virtual NTO. In contrast to this, the excitation energy of the enolate tautomer is decreased since here the occupied NTO is closer to this residue and raised in energy. In order to split the absorption band, this opposite trend has to be increased by introducing a negative charge. The proximity to the substituent can lead

Table 5.2.: Vertical excitation energy shifts for the charge-transfer excitations of keto and enolate-AcThDP upon mutations in the *B. breve* PK enzyme. These are based on the perturbative approach ($\Delta\omega_{\text{PT}}$) and full TD-DFT/TDA calculations ($\Delta\omega_{\text{TD-DFT}}$) at the CAM-B3LYP/def2-TZVP level of theory. The predicted new band maxima of the mutants were obtained for the TD-DFT/TDA calculations with respect to the experimentally measured band maximum in wild-type PK. The energies are given in electronvolts (eV). The corresponding wavelengths in nanometers (nm) are given in parentheses.

mutant	keto-AcThDP			enolate-AcThDP		
	$\Delta\omega_{\text{PT}}$	$\Delta\omega_{\text{TD-DFT}}$	ω_{max}	$\Delta\omega_{\text{PT}}$	$\Delta\omega_{\text{TD-DFT}}$	ω_{max}
Glu156Gln	0.43	0.29	3.24 (382)	0.07	0.07	3.02 (410)
Asn549Asp	0.47	0.54	3.49 (355)	-0.79	-0.34	2.61 (475)

to complications since this can strongly influence the energetics or mechanism of the reaction. However, as possible mutant candidate Asn549 seems to be reasonable.

Having identified residues with large deviations for the impact on the excitation energies of keto and enolate-AcThDP, the next step was to propose possible mutations and predict the excitation energy shifts upon this. In order to achieve reliable results, the conformation of the enzyme should not change significantly by mutating these residues. For this reason, isosterical mutations were employed with the variants Glu156Gln and Asn549Asp. The corresponding side chains were modified in the original crystal structure and optimized at the MM level of theory. Excitation energy shifts were then calculated with the perturbative approach and compared to the shifts from full TD-DFT/TDA calculations. The predicted excitation energies of the mutants were obtained for the TD-DFT/TDA calculations with respect to the experimentally measured band maximum at 2.95 eV (420 nm). The results are listed in Table 5.2.

The calculated excitation energy shifts with the perturbative approach are in reasonable agreement in comparison to the full TD-DFT/TDA calculations. Except one value, the deviations are less than 0.15 eV and all shifts predict the same trends. The largest deviation is observed in case of the Asn549Asp variant for the charge-transfer excitation of enolate-AcThDP. Here, a negative charge is introduced close to the excitation space, leading to a strong perturbation for the system. This could not be captured by simple first order perturbation theory, so that the shift is significantly overestimated by 0.45 eV. Therefore, the discussion of the mutant spectra is restricted to the more accurate TD-DFT/TDA results. However, the perturbative approach is capable to predict the correct trends and is much less computational demanding. In further applications

this would be a promising alternative.

In case of the Glu156Gln variant of PK, a negative charge is removed from the proximity of the aminopyrimidine ring of AcThDP. Therefore, the supporting effect for the charge-transfer excitation of keto-AcThDP is reduced, so that the energy is raised by 0.29 eV. The predicted band maximum is now 3.24 eV (382 nm). For enolate-AcThDP the impact of the Glu156 was already observed to be much lower, since it is sufficiently far apart from the excitation space. Therefore, the excitation is only slightly increased by 0.07 eV, resulting in a band maximum at 3.02 eV (410 nm). This deviation of 0.22 eV in the shifts should be enough to observe the two species separately in the UV-vis spectra. The Tittmann group synthesized this variant of PK and measured its time-dependent UV-vis spectra. Unfortunately, no difference could be detected in comparison to the wild-type enzyme.[40] Both the absorption band at 2.95 eV (420 nm) and the kinetic overshooting were retained. The reason for this could be that the assumption of a deprotonated (negatively charged) Glu156 was not valid. A protonated (neutral) Glu156 approximately mimics the electrostatic potential of a glutamine residue. Regarding this protonation state, a significant shift of the absorption band of both tautomers of AcThDP cannot be expected.

The second mutant, Asn549Asp, shows even more pronounced excitation energy shifts for the intermediates. The excitation energy of keto-AcThDP is blue shifted by 0.54 eV to 3.49 eV (355 nm). The introduced negative charge close to the acetyl unit causes an energetically increased virtual NTO, leading to a larger excitation energy. Opposed to this, the occupied NTO of the charge-transfer excitation of enolate-AcThDP is closer to this residue, so that the excitation energy is red shifted by 0.34 eV to 2.61 eV (475 nm). This variant was also synthesized in the Tittmann group. Unfortunately, no absorption could be detected in the spectral region of 2.48–3.54 eV (350–500 nm).[40] Introducing a negatively charged residue in the reaction center can strongly influence the energetics of the intermediates, so that both tautomers of AcThDP may not be stabilized. Another explanation would be a change in the reaction mechanism with aspartate acting as acid-base catalyst. This could result in off-pathway reactions, making an observation of AcThDP impossible.

Although the experimental realization could not confirm the predicted excitation energy shifts upon mutation of PK, the perturbative approach shows promising features. It predicts the same trends compared to the full TD-DFT/TDA calculation with virtually no computational cost after the reference calculation. Furthermore, the approach is not

restricted to a single (crystal) structure. It could be easily extended to MD-QM/MM calculations in future applications, thereby sampling the configurational space which could have an important effect on the calculated spectra. Employing this methodology, the perturbative approach could then be applied as screening method for a variety of mutants of an enzyme, so that the most promising ones for a desired excitation energy shift are identified. The reduced number of mutant candidates would then be subject of higher-level methods and in turn provide suggestions of enzyme mutants with the best chances of success for experimental studies.

5.3. Conclusions

The AcThDP intermediate was experimentally observed in the reaction pathway of the *B. breve* PK enzyme and connected to an absorption band with maximum at 2.95 eV (420 nm), which additionally showed a kinetic overshooting. In this work, the spectral signatures of the tautomers of AcThDP were studied through the use of TD-DFT, thereby benchmarking the effect of the TDA approximation and different basis set sizes. Additionally, the COSMO solvation model was employed to represent the polarization of the enzyme environment. This was then compared to the spectral signature of free AcThDP in aqueous solution. Furthermore, the effect of enzyme residues of PK on the AcThDP spectra were analyzed with a perturbative approach. This was used to propose mutants of PK, where the excitation energy shifts were calculated. The main findings can be summarized as follows.

(1) The large shift between the band position of enzyme-bound and free AcThDP is connected to a change in the conformation as well as in the polarization of the surrounding environment. This was already observed for unsubstituted ThDP and is likely to be a general trend for ThDP derivatives.

(2) Both keto and enolate-AcThDP absorb in the spectral region of 2.95 eV (420 nm). Enol-AcThDP shows a significant larger excitation energy, so that this intermediate is not directly observed in the spectra.

(3) The lowest excited state of keto-AcThDP has charge-transfer character with the electron being transferred from the aminopyrimidine ring to the thiazolium ring delocalized towards the acetyl group. Another charge-transfer character is observed

for enolate-AcThDP, where an electron is excited from the π-system of the enolate moiety to the thiazolium ring.

(4) The excitation of enolate-AcThDP has a significantly larger oscillator strength compared to the one of the keto tautomer. Therefore, the kinetic overshooting can be explained by a fast formation of enolate-AcThDP tautomerizing slowly to keto-AcThDP. This intermediate still absorbs in the spectral region of interest, but less intense.

(5) The perturbative analysis of the enzyme residues on the spectra of keto and enolate-AcThDP revealed two mutant candidates for PK, which would have been capable to split their common absorption band. Unfortunately, the experimental realization could not confirm these predictions, most likely due to different protonation states of the residues or an influence on the reaction mechanism.

The interpretation of the spectra calculations in light of the experimental observations provided valuable information on the mechanism of the PK enzyme, identifying the enolate-AcThDP as a key residue. This intermediate then equilibrates slowly in absence of phosphate to keto-AcThDP with a rate constant of 4 s^{-1}, preventing the off-pathway hydrolysis. The presence of the phosphate even raises the rate constant of the reaction ($k_{\text{cat}} = 14$ s^{-1}), indicating a special role of this substrate. Its impact was studied through reaction pathway calculations for the tautomerization process in the content of the Master's project of S. Wille.[41] The results are featured in his thesis and here briefly discussed, since they complement the spectra calculations. Three different reaction mechanisms were considered in the work, employing a QM cluster model for AcThDP and selected residues in the active site of PK. These are schematically represented in Figure 5.5. Furthermore, the free activation energies for the conversion of enolate to keto-AcThDP are given, calculated at the B3LYP-D3(BJ)/def2-TZVP//B3LYP-D3(BJ)/def2-SVP level of theory.

The reference in the reaction pathway calculations was chosen to be enol-AcThDP, which is converted to the enolate form after proton transfer to the His553 residue with a negligible barrier. In the first mechanism (A) this residue then acts also as proton donor to build keto-AcThDP, whereby the phosphate was absent. The calculated free activation energy of 77.1 kJ · mol^{-1} agrees well with the observed kinetic constant of 4 s^{-1} which corresponds to a barrier of 70.0 kJ · mol^{-1}, employing the Eyring-Polanyi

equation at standard ambient temperature.[139] Including the phosphate, two possible scenarios were distinguished. In case of mechanism B, the His553 residue acts again as proton donor. Thereby, the activation barrier is strongly increased to $116.0 \, \text{kJ} \cdot \text{mol}^{-1}$, making this process unlikely. The reason for this is that the phosphate sterically hinders the His553 residue by its movement for the proton transfer. However, the proton for the tautomerization could also be supplied from the phosphate itself, which is ideally placed next to the acetyl substituent (mechanism C). This results in a decrease of the barrier to $45.0 \, \text{kJ} \cdot \text{mol}^{-1}$, explaining the larger catalytic rate constant compared to the rate for the tautomerization in absence of phosphate. Furthermore, the rate-determining step in the reaction is likely shifted, so that AcThDP is not accumulated and could not be spectroscopically detected.
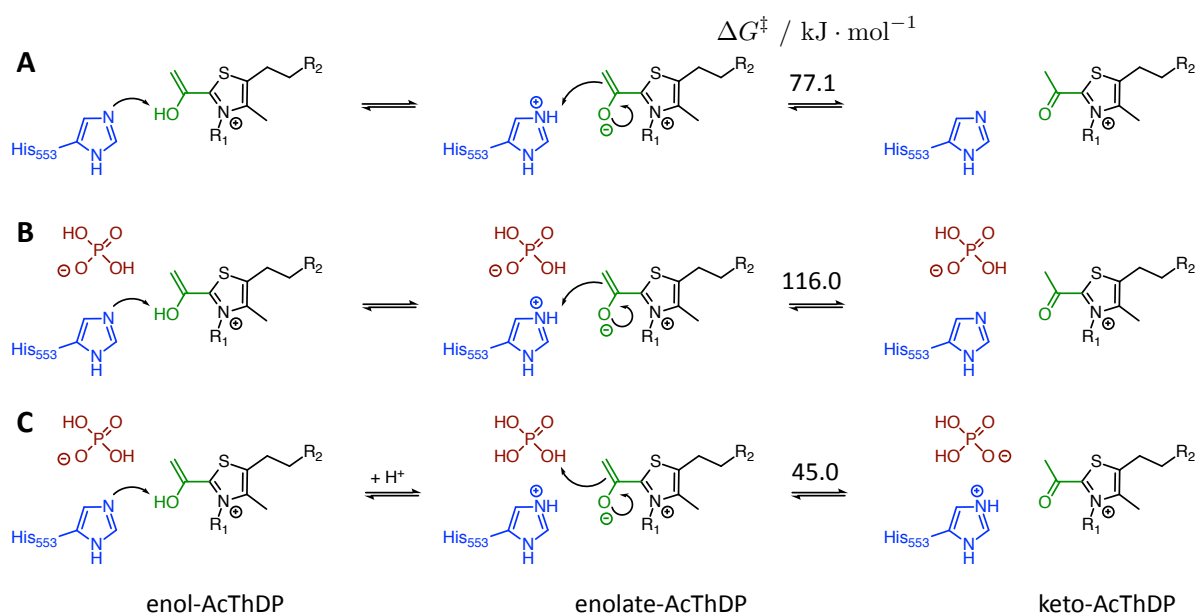
Figure 5.5.: Reaction mechanisms under study for the keto-enol tautomerization of AcThDP in *B. breve* PK.[41] In mechanism A the tautomerization is mediated by His553 in absence of the co-substrate phosphate. In mechanism B the phosphate is included in the enzyme pocket, but the tautomerization is still mediated by His553. Mechanism C consideres a co-substrate assisted enzyme catalysis with the phosphate being the effective proton donor for the formation of keto-AcThDP.

6

Transketolase

Transketolase (TK) is a ThDP-dependent enzyme, catalyzing the reversible transfer of a dihydroxyethyl unit from a donor ketose to the C1-position of an acceptor aldose. This is a key reaction required in the Calvin cycle of photosynthesis as well as in the nonoxidative pathway of pentose phosphates.[9;19;140] In several experiments this enzyme was investigated, including crystal structure analysis[105;141;142] and mutagenesis studies combined with kinetic measurements.[143–145] These investigations lead to a proposed reaction mechanism (Figure 6.1). The activated ylide-ThDP performs a nucleophylic attack at the ketose xylulose-5-phosphate (X5P) to build a substrate-ThDP conjugate. After C2–C3 bond cleavage of the sugar moiety, the aldose product, glyceraldehyde-3-phosphate (G3P), is released and 1,2-dihydroxyethyl-ThDP (DHEThDP) is formed. Opposed to the phosphoketolase enzyme, this intermediate is kinetically stabilized by the TK enzyme in order to react with the newly arrived aldose substrate,[141] either being ribose-5-phosphate (R5P) or erythrose-4-phosphate (E4P). Thereby, the ketoses sedoheptulose-7-phosphate (S7P) or fructose-6-phosphate (F6P) are formed and afterwards released as products.

The Tittmann group investigated this enzyme in further detail.[40] In a first set of experiments their focus was on the key intermediate 1,2-dihydroxyethyl-ThDP (DHEThDP). The latter was indentified with crystallographic measurements at high resolution (0.95 Å) of *Escherichia coli* (*E. coli*) TK. For this purpose, the enzyme was reacted with the substrate analog β-hydroxypyruvate, which is converted to DHEThDP after bonding to the activated cofactor and subsequent decarboxylation. The high resolution of the X-ray measurements allowed the detection of some hydrogen atoms with the pyrimidine ring

being characterized in the APH$^+$ state. However, the protonation state of the dihydrox-yethyl substituent could not be determined. Both a enol as well as enolate state of this enamine intermediate would be conceivable. The proton could easily be transferred to a nearby histidine residue. Additionally, UV-vis measurements revealed an uncommon broad absorption band in the spectral region of 2.25–3.54 eV (350–550 nm), most likely connected to DHEThDP (see Figure 6.2b). In this work, the focus was on the spectral behavior of DHEThDP, considering both states of the substituent. This leads to an explanation for the origin of the broad measured band.

The second set of experiments carried out in the Tittmann group considered the co-operativity in substrate binding of the dimeric human TK enzyme. X-ray structural analysis at high resolution (0.97 Å) revealed a proton wire connecting the two active sites over a distance of 25 Å. This includes six glutamate residues, three water molecules as well as a glutamine bridge (see Figure 6.2c). The wild-type enzyme exhibits positive

Figure 6.1.: Proposed mechanism of the reaction of TK with the donor ketose xylulose-5-phosphate (X5P) and the acceptor aldose ribose-5-phosphate/erythrose-5-phosphate (R5P/E4P). Thereby, the products glyceraldehyde-3-phosphate (G3P) and sedoheptulose-7-phosphate/fructose-6-phosphate (S7P/F6P) are released. Pyr and DP abbreviate the pyrimidine ring and diphosphate portion of ThDP, respectively.

cooperativity of substrate binding (Hill coefficient of 1.5–1.6), whereas mutation of any glutamate residue to a glutamine inhibits the proton transfer, so that no cooperativity is observed. Interestingly, two conformations were detected for the glutamine bridge together with an associated water molecule. The latter potentially serves as a relay for proton transfer moving over the two-fold symmetry axis and connecting the two sides of the proton wire. In this work, molecular dynamics simulations were carried out with the focus on the stability of the proton wire as well as the flexibility of the glutamine bridge.
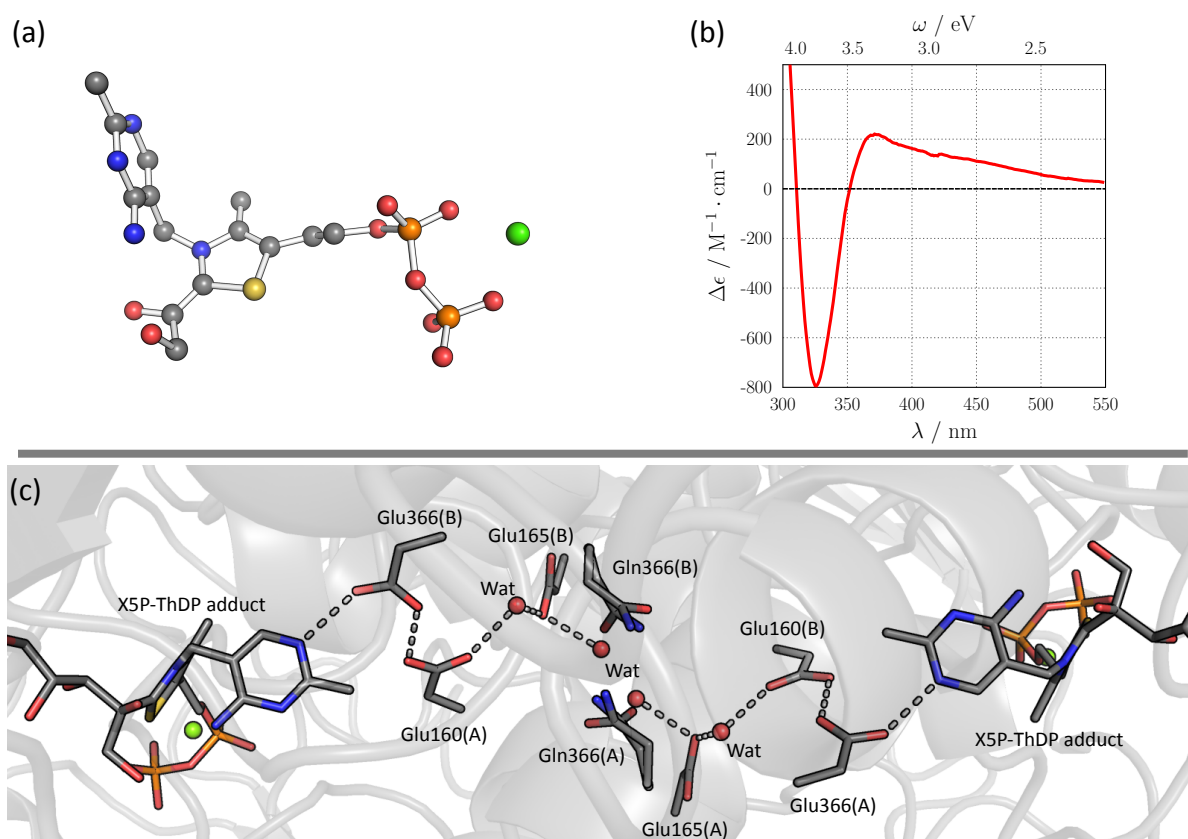
Figure 6.2.: Experimental observations for the TK enzyme from the Tittmann group. (a) Crystal structure of the DHEThDP intermediate in *E. coli* TK at 0.95 Å resolution. (b) Difference UV-vis spectrum of *E. coli* TK after addition of β-hydroxypyruvate with respect to the resting state of the enzyme. (c) Crystal structure (PDB code 4KXW, resolution of 0.97 Å) of the proton wire connecting the two active sites in the human TK enzyme. Two conformations of the glutamine bridge and the associated water molecule are shown. The hydrogen bond network is indicated by dashed lines. Experimental data for the graphics were provided by the Tittmann group from Göttingen. [40]

6.1. Computational Details

6.1.1. Spectra Calculations of the DHEThDP Intermediate

The spectra of the DHEThDP intermediate were studied in the active site of *E. coli* TK, employing a QM/MM cluster model. Here, the crystal structure data was provided by the Tittmann group,[40] including DHEThDP by soaking TK with β-hydroxypyruvate. The intermediate was saturated with hydrogen atoms, setting the pyrimidine ring to its APH$^+$ form. Two different states of the substituent were considered. In the first case, the O2 atom of the substituent was protonated (enol-DHEThDP), whereas an enolate form was used for the second state of DHEThDP. A residue shell of 5 Å around the cofactors was taken, including also six crystallographic water molecules. Free residues were capped by acetyl or *N*-methyl groups. Titratable residues were set to their reference state at a pH of 7 with a few exceptions. The residues Asp183 and Asp184 are located at the edge of the amino acid shell, missing the interactions with the second shell of residues. Therefore, these residues were chosen to be protonated (neutral). The Glu160 residue was found to be in direct contact to the canonical glutamate (Glu411). It was set protonated (neutral) in order to build a hydrogen bond. Furthermore, the His66 was doubly protonated (positively charged) due to its location in the proximity of the diphosphate moiety of DHEThDP. The His473 residue was the only one differing in the protonation state for enol and enolate-DHEThDP, since this residue is proposed to be the proton acceptor/donor for their interconversion. For this reason, it was set monoprotonated (neutral) at its Nδ atom in order to form a hydrogen bond to the hydroxyl group of enol-DHEThDP, whereas a doubly protonated state (positively charged) was used for enolate-DHEThDP. The AMBER ff10 force field was then employed for the parametrization of the enzyme environment,[118] whereas the GAFF force field was used for DHEThDP.[120] The hydrogen atoms of the water molecules were optimized at the MM level of theory, followed by a QM/MM optimization of the intermediates at the B3LYP-D3(BJ)/def2-TZVP level of theory.[61;62;111–113] Vertical excitation energies were then obtained with TD-DFT/TDA calculations (CAM-B3LYP/def2-TZVP) for DHEThDP in the pointcharge field of the amino acids.[65;113]

A more detailed analysis for the proton potential for the conversion from enol to enolate-DHEThDP was also carried out, employing the same setup as for the single point calculations. However, this time all hydrogen atoms were optimized at the MM

level of theory, restraining all other atoms in their Cartesian coordinates by a harmonic potential with a force constant of 5 kcal·mol^{-1}·Å^{-2}. A QM/MM optimization followed for enol-DHEThDP at the B3LYP-D3(BJ)/def2-TZVP level of theory, including also the side chains of His473 and Glu411 residues in the QM part. Hydrogen atoms were added as link atoms to the QM region for the cut covalent bond, while the charge-redistribution scheme of Sherwood and coworkers was applied for the adjacent MM atoms.[80] This was the setup of choice for all subsequent QM/MM calculations. The potential for the hydrogen movement was then calculated through incremental increase or decrease of the O–H distance by 0.5 bohr and reoptimization of only its position with a restraint to the O–H bond length in form of a harmonic potential ($k = 1$ kcal·mol^{-1}·Å^{-2}). The potential was then fitted to a eight-order polynomial function. An anharmonic frequency analysis followed,[146] thereby approximating the system as a rotation-free two-atomic O–H molecule. The vibrational wave functions were expanded with a basis set of 41 harmonic oscillators. Solution of this nuclear Schrödinger equation then lead to the vibrational states, so that the probability of presence of the hydrogen atom within the potential could be determined. Furthermore, TD-DFT/TDA spectra at the CAM-B3LYP/def2-TZVP level of theory were calculated at several points of the pathway.

Program packages involved in the computations were ORCA 3.0.3 (DFT),[123] Amber 11 (MM)[124] and ChemShell (QM/MM).[80;125] The anharmonic frequency analysis was supported by Benjamin Schröder from the Mata group in Göttingen.

6.1.2. Molecular Dynamic Simulations of the Human TK Enzyme

The dimeric human TK enzyme was modeled on the basis of the crystal structure provided by the Tittmann group (PDB code 4KXW, resolution of 0.97 Å), including the X5P-ThDP adducts in the active sites. These were taken and saturated with hydrogen atoms, considering the pyrimidine ring in the IP state. After an optimization at the B3LYP-D3(BJ)/def2-SVP level of theory,[61;62;111–113] whereby all angles and dihedrals were constrained, atomic charges were determined through an electrostatic potential fit (HF/6-31G*).[85] All other parameters were used from the GAFF force field.[120] As already described for unsubstituted IP-ThDP (see Chapter 4), two new atom types were introduced for the C4' and N4' atoms of the iminopyrimidine ring. From the environment the sodium ions and ethylene glycole molecules (arising from the crystallization process) were removed. The remaining protein was saturated with hydrogen atoms and

parametrized with the AMBER ff10 force field.[118] Thereby, the protonation states of titratable residues were assigned after inspection of their local environment or through information from the X-ray electron density map.[40] There were only a few notable exceptions with differences to their reference states at pH 7. The Asp424 residue was found to interact with the ThDP-bound substituent, forming a hydrogen bond. Therefore, it was assigned to its protonated (neutral) state. Two histidine residues (His258 and His416) are located in the proximity of the diphosphate moiety of ThDP and set doubly protonated (positively charged). The protonation states of the glutamate residues involved in the proton wire are later discussed in the text. The total system charge of +8 atomic units was then neutralized with chloride ions. Afterwards, the system was put into a TIP3P water box, having a distance of 12 Å between the enzyme and the boundary of the box. A short minimization followed for the hydrogen atoms, water molecules and chloride ions, thereby restraining all other atoms in their Cartesian coordinates with a harmonic potential with a force constant of $5 \ \text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$. This structure was then used as reference for the molecular dynamics (MD) simulations and their analysis.

After the preparation, the system was heated up in six stages over 60 ps to 300 K in a NVT simulation. The temperature was controlled by Langevin dynamics, employing a collision frequency of $2 \ \text{ps}^{-1}$. Furthermore, periodic boundary conditions together with a cutoff of 12 Å for non-bonded interactions was applied as well as the SHAKE algorithm to constrain bond lengths involving hydrogen atoms.[121;122] The latter allowed the application of a 2 fs time step. An equilibration phase followed for 500 ps in a NPT simulation. The pressure was controlled by isotropic position scaling ($p = 1$ bar). In order to check for convergence, the root-mean-square-deviation of the protein backbone atoms was monitored. The NPT simulation was then extended for 16.5 ns as the final production phase. All MM calculations were carried out with the Amber 11 program package.[124]

6.2. Results and Discussion

6.2.1. Spectra of DHEThDP

The broad absorption band in the range of 2.25–3.55 eV (350–550 nm) observed in UV-vis measurements of *E. coli* TK after addition of β-hydroxypyruvate is most likely linked

to the formation of the DHEThDP intermediate. The spectra of this key intermediate were studied in the active site of the enzyme, using an amino acid shell of roughly 5 Å around the cofactor. Hereby, the pyrimidine ring of DHEThDP was set to its APH$^+$ state with a deprotonated (negatively charged) canonical glutamate based on the information from the electron density map of the X-ray measurements. Two different setups were employed for the protonation state of the substituent of DHEThDP in connection to a hydrogen bonded histidine residue (His473). In the first case, the oxygen atom of the enamine moiety was protonated (enol-DHEThDP), forming a hydrogen bond to the deprotonated Nϵ atom of the (neutral) His473 residue. In the second case, the oxygen atom was deprotonated (enolate-DHEThDP), whereby the His473 residue was assumed to serve as proton acceptor and set doubly protonated (positively charged). Otherwise, the same setup for the enzyme environment was considered for both systems. After a QM/MM optimization of the intermediates at the B3LYP-D3(BJ)/def2-TZVP QM level of theory, vertical excitation energies were obtained with TD-DFT/TDA employing the CAM-B3LYP functional with the def2-TZVP basis set. The optimized structures and calculated excitation energies are shown in Figure 6.3.

(a) enol-DHEThDP

$\omega_1 = 2.92$ eV $\quad \lambda_1 = 424$ nm $\quad f_1 = 0.006$
$\omega_2 = 3.33$ eV $\quad \lambda_2 = 372$ nm $\quad f_2 = 0.004$

(b) enolate-DHEThDP

$\omega_1 = 1.86$ eV $\quad \lambda_1 = 668$ nm $\quad f_1 = 0.022$
$\omega_2 = 2.25$ eV $\quad \lambda_2 = 552$ nm $\quad f_2 = 0.003$
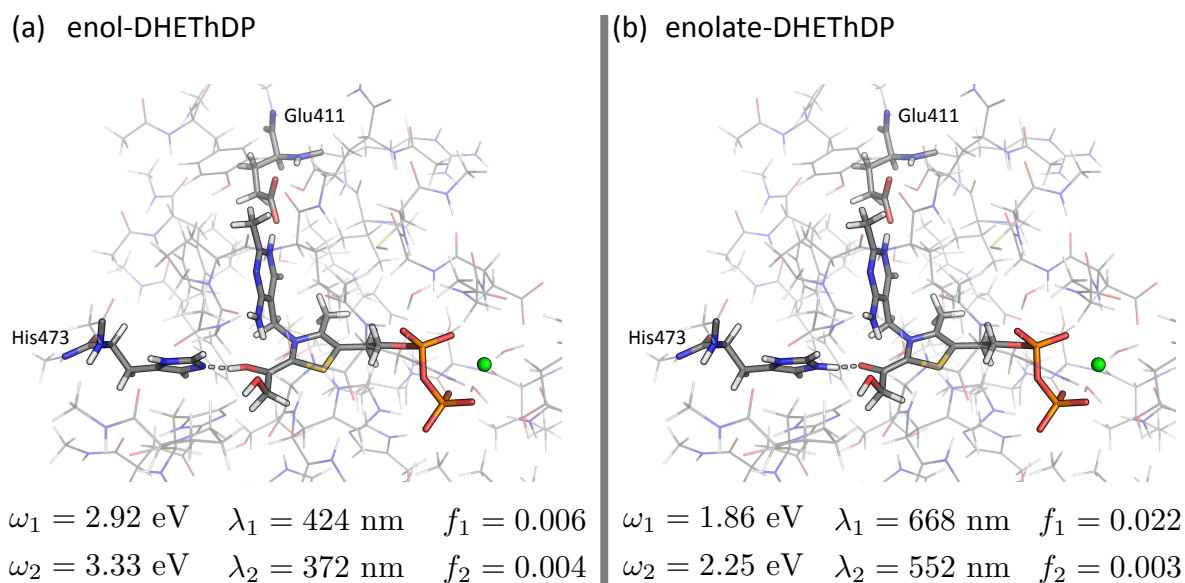
Figure 6.3.: QM/MM optimized structures of (a) enol and (b) enolate-DHEThDP in the active site of *E. coli* TK. The His473 residue is assumed to serve as proton acceptor/donor for their inter-conversion. The pyrimidine ring was set to its APH$^+$ state with a deprotonated canonical glutamate (Glu411). Furthermore, TD-DFT/TDA with CAM-B3LYP/def2-TZVP vertical excitation energies, their associated wavelengths and oscillator strengths are given for the first and second excited states for each system.

The spectra calculations revealed two low-energy excited states for enol and enolate-DHEThDP which could contribute to the observed band at 2.25–3.55 eV (350–550 nm). In particular, these excitation energies are found at the limits of the absorption band. In the high energy regime (2.9–3.3 eV) the enol-DHEThDP absorbs, whereas the excitation energies are strongly red shifted for enolate-DHEThDP to 1.86 and 2.25 eV. This large shift of about 1.1 eV roughly corresponds to the broadness of the measured band ($\approx$ 1.3 eV). The origin of this energy gap can be explained by inspection of the excitation characters. The dominant NTO pairs for the first and second excited states of enol and enolate-DHEThDP are shown in Figure 6.4. Equivalent excitation characters are observed for both states of DHEThDP. All excited states correspond to charge-transfer excitations from the π-system of the thiazolium-substituent moiety to the aminopyrimidinium ring. This is an opposite charge-transfer character compared to the spectral characteristics of unsubstituted ThDP and AcThDP featured in the preceeding chapters. In the case of DHEThDP the electronic structure is reversed, so that the electron rich thiazolium-substituent moiety acts here as electron donor and the positively charged
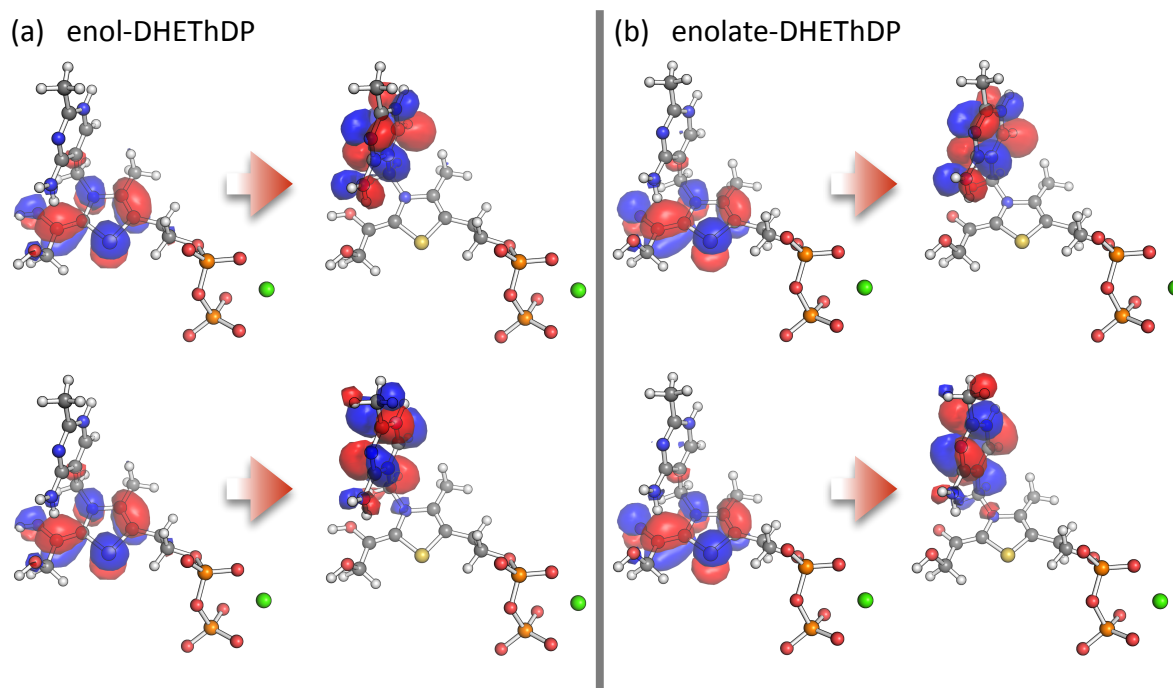
Figure 6.4.: Dominant CAM-B3LYP/def2-TZVP NTO pairs for the (a) enol and (b) enolate-DHEThDP. The NTO pairs for the first and second excited state are shown on the top and on the bottom, respectively.

pyrimidine ring (APH$^+$) being the electron acceptor. According to this excitation character, the introduction of a negative charge at the thiazolium-substituent moeity by moving from enol to enolate-DHEThDP facilitates the donation of an electron in the excitation process. This results in a significantly lower excitation energy for the enolate-DHEThDP intermediate.

The information of the first set of calculations can be summarized as follows. Both protonation states of the DHEThDP intermediate show excited states with similar charge-transfer characters in the energy range of the measured UV-vis band. These are found at the limits of the band, whereby the excitation energies of enol and enolate-DHEThDP are located in the high and low-energy regime, respectively. The connection between the two limits of the band could arise from a dynamical equilibrium of the intermediates or from a rather weak O–H bond in the enol case. For this reason, the 1D-potential for the hydrogen movement was calculated. Thereby, the optimized structures of enol-DHEThDP was used as reference. The O–H distance was then incrementally increased/decreased. The hydrogen position was then reoptimized with a restraint for its distance to the oxygen atom through a harmonic potential. All other atoms were constraint in their Cartesian coordinates. For these calculations the side chains of the His473 residue as well as the canonical glutamate (Glu411) were included in the QM region. An anharmonic frequency analysis followed in order to determine the probability of presence of the hydrogen atom along the calculated path. This was applied to scale the oscillator strengths of the excited states from TD-DFT/TDA calculations (CAM-B3LYP/def2-TZVP). A spectrum was then simulated by convolution of Gaussian bands (FWHM = 0.3 eV). The results are shown in Figure 6.5.

The potential for the hydrogen movement reveals a common shape for an anharmonic O–H stretch vibration. Employing this reaction coordinate, no minimum could be observed at large O–H distances, where the enolate-DHEThDP intermediate with a doubly protonated His473 residue would be expected. The anharmonic frequency analysis shows significant probability of presence in the O–H distance range of 0.9–1.1 Å with a maximum at 1.01 Å for the lowest quantum state. This distance is, as expected, somewhat larger than the determined minimum of the potential (0.99 Å), but still refers to a enol-DHEThDP intermediate. The first excited vibrational state is well separated by about 2800 cm^{-1} from the ground state. According to Boltzmann statistics, only the latter is significantly populated at a temperature of 300 K. Therefore, the probability of presence of the hydrogen position from the vibrational ground state was employed for the simu-

lation of the UV-vis spectrum (blue curve in Figure 6.5b). This includes only the two charge-transfer transitions since these are the ones contributing to the measured band. The simulated spectrum shows absorbance in the spectral range of 2.91–3.81 eV (325–425 nm). Similar to the single point calculations, this corresponds to the high-energy regime of the measured UV-vis band. In this model, the broadness of the band cannot be explained. However, there are important information resulting from these calculations. The observed band cannot solely originate from the enol-DHEThDP intermediate. For that case, a much sharper band at about 3.26 eV (380 nm) would be expected. In order to obtain such a broad band the enolate-DHEThDP intermediate has to be involved. The simple model of a moving hydrogen atom in a rigid environment is not enough to describe this system. A coupling to other degrees of freedom is required for their interconversion. This could be other proton transfers or conformational changes in the enzyme environment. This effect was, however, not investigated in further detail.
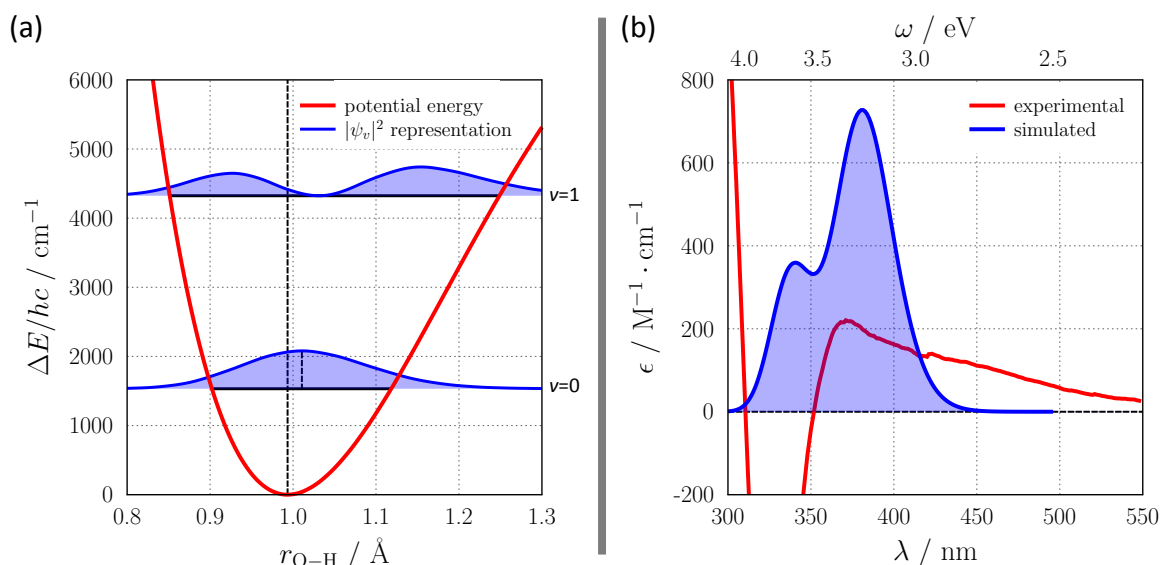
Figure 6.5.: (a) QM/MM potential energy curve for the hydrogen movement connecting the enol-DHEThDP intermediate with its enolate form in the active site of *E. coli* TK. Furthermore, the vibrational ground and first excited state levels are depicted together with the resulting probability of presence ($|\psi_v|^2$) of the hydrogen position. Dashed lines indicate the positions of the minumum of the potential and the maximum of $|\psi_0|^2$. (b) Simulated absorption spectrum for the charge-transfer states of DHEThDP in comparison to the measured difference absorbance.

6.2.2. Dynamics of the Proton Wire in the Human TK Enzyme

The dimeric human TK enzyme was studied through crystallographic measurements at high resolution (0.97 Å) in the Tittmann group. A proton wire was identified connecting the two active sites over a distance of 25 Å. It is proposed to synchronize the enzyme activity, exhibiting positive cooperativity of substrate binding for the wild-type enzyme. Additionally, two conformations for a glutamine bridge and an associated water molecule was observed which is potentially responsible for the proton transfer over the two-fold symmetry axis. The stability of the hydrogen bond network as well as the flexibility of the glutamine bridge was investigated through molecular dynamics (MD) simulations. The setup of the protonation states of the residues involved in the proton wire is depicted in Figure 6.6.

The protonation states were symmetrically assigned according to the symmetry axis. Starting with the pyrimidine moiety of the X5P-ThDP adduct, its N1' atom was found to be protonated from the X-ray electron density map, so that a IP state is assumed for the tetrahedral substituted ThDP intermediate.[29] A hydrogen bond is build to the canonical
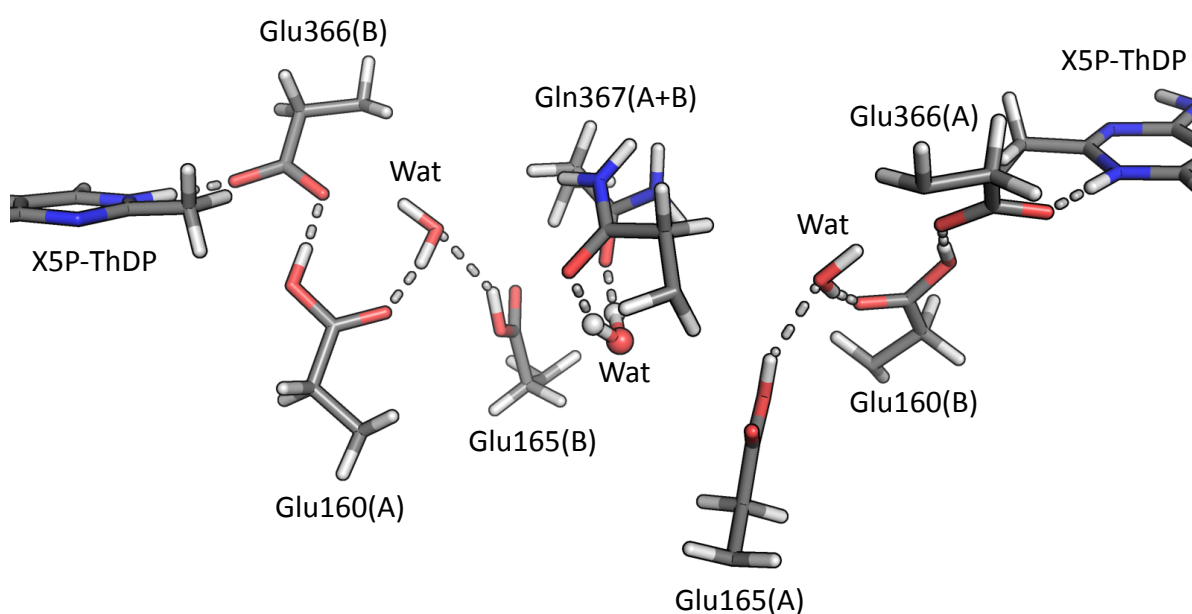
Figure 6.6.: Setup for the MD simulations of the protonation states of glutamate residues involved in the proton wire, connecting the two active in the human TK enzyme. Hydrogen bonds are indicated as dashed lines. The glutamine bridge and the associated water molecule was initialized in one of the observed conformations. The latter potentially serves as proton raly in the communication channel and is here highlighted in a ball-and-stick representation.

glutamate (Glu366), which is in close contact to the Glu160 residue. Their $O \cdots O$ distance is just 2.54 Å with a hydrogen located nearly halfway between them. This could indicate a low-barrier hydrogen bond,[147] having a special role in the communication channel. This subject was also studied in the collaboration of the Tittmann and Mata groups, but is not discussed in further detail within this thesis. However, for a classical MD simulation the proton has to be assigned to either of these residues. Here, the Glu160 residue was set protonated, so that the bridging water molecule could form a hydrogen bond to its other carboxyl oxygen atom. The hydrogen bond network was then finished by a protonated Glu165 residue interacting with the before mentioned water molecule. The glutamine bridge and the associated water molecule was initialized in one of the observed conformations. Employing this setup, the system was heated up and equilibrated. Finally, the system was simulated for a total production time of 16.5 ns at 300 K and 1 bar.

The MD simulation was analyzed by monitoring the key physical quantities (energies, temperature and pressure). The fluctuations were within the expected thresholds (e.g. $T = 300.0 \pm 0.8$ K), indicating a stable MD run. Additionally, the root-mean-square-deviation (RMSD) of the enzyme backbone atoms was monitored. This quantity is commonly employed to identify conformational changes of the enzyme structure.[148] In this MD simulation there are no strong deviations from the reference structure observable. The enzyme retains its initial conformation. Furthermore, visible inspection of the trajectory also revealed a well-behaved simulation, where no unphysical motions could be detected. In particular, the substrate-ThDP adducts and the residues involved in the proton wire were viewed. The simulation showed a stable hydrogen bond network as indicated in Figure 6.6. All hydrogen bonds and conformations of these residues are kept throughout the simulation time.

Having verified a stable MD simulation and hydrogen bond network within the communication channel, the focus was on the flexibility of the glutamine bridge with the associated water molecule. Two descriptors were employed to analyze the dynamics of this unit. On the one hand, the Cartesian coordinates RMSD of both Gln367 residues was monitored relative to the reference structure. Preceeding RMSD fits of the enzyme backbone atoms were carried out for each snapshot to ensure comparable structures. This value could then identify different conformations for these residues. On the other hand, the difference distance between the oxygen atom of the associated water molecule and the protonated oxygen atoms of the Glu165(A) and Glu165(B) residues were calcu-

lated. This was employed to follow the movement of the water molecule. The descriptors were calculated in 0.5 ps intervals including 33000 structures in total. The results are shown in Figure 6.7.

The analysis of the trajectory reveals three different conformations for the Gln367 residues, which are observable in their Cartesian coordinate RMSD (Figure 6.7c). The structures fluctuating around 1.0 and 2.5 Å correspond to the partially occupied sites found in the crystal. Typical snapshots for these conformations are shown in Figure 6.7a. Several events are observed for the conformational change during the simulation time, thereby the water seems also to move through the symmetry axis. This can be seen by comparison with the difference distance of the water to the Glu165 residues (Figure 6.7d). The oscillation occurs if the sign of this descriptor changes. Similar to the RMSD analysis several events are observed for this movement. A third conformation is found for the glutamine bridge in the time interval 4.0–10.2 ns with a RMSD of about 3.5 Å (Figure 6.7b). Here, another water molecule enters the unit and forms a hydrogen bond to one of the glutamine residues. Thereby, the bridge is deactivated. The shuttling water sticks at one site in this period. However, the movement is later recovered, so that the conformations observed in the crystal are restored.

The time scale for the preservation of the water position was estimated from the difference distance of the water molecule to the Glu165 residues. Hereby, the moving average over 25 ps was consulted in order to reduce the scattering of the data (see Figure 6.7d). A shuttling event was then defined if the difference distances exceeds 1.0 Å starting from the negative axis or deceeds -1.0 Å with origin on the other site. This description results in the observation of 23 shuttling events of the water molecule during the 16.5 ns simulation time. Based on these criteria, the time scale for the preservation of the water molecule is estimated at 0.7 ± 0.3 ns, corresponding to a shuttling frequency of 1.4 ± 0.6 ns^{-1}. Admittedly, the error range for the time scale is quite large due to the short simulation time as well as the period of the third conformation where the water sticks at one site. However, the important information is the determination of the oscillation within the nanosecond regime. This is much faster than the catalyzed reaction time, so that the hypothesis of the water molecule acting as proton carrier in the communication channel is maintained.

As mentioned above, the conformational change of the glutamine residues seem to correlate with the movement of the water molecule through the symmetry axis. In order to verify this assumption, a cluster analysis of the trajectory was carried out with the

density-based spatial clustering of applications with noise (DBSCAN).[149] Hereby, the data in the time intervals 0–4.0 ns and 10.2–16.5 ns were considered, excluding the period
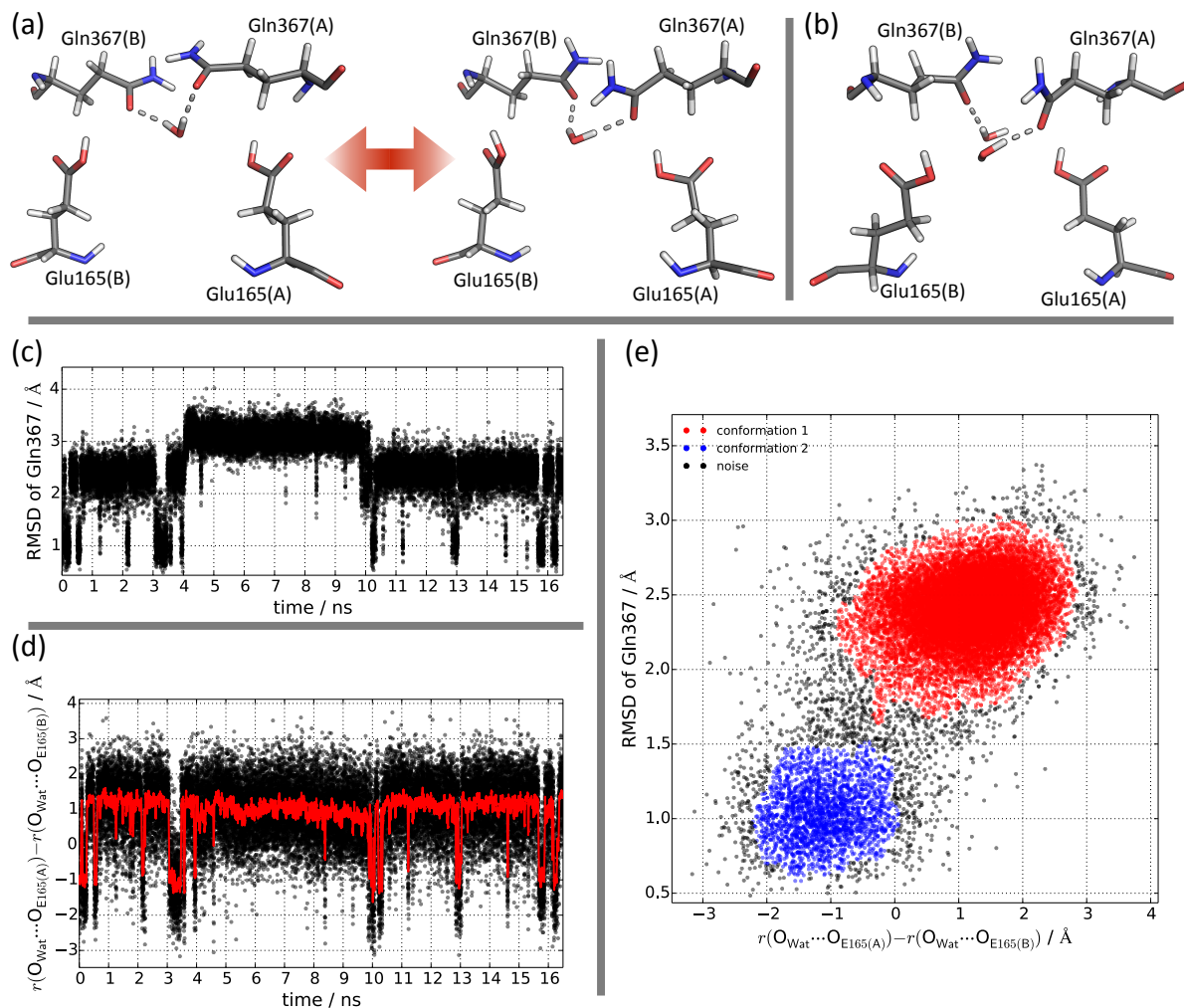
Figure 6.7.: (a) Snapshots of the MD simulation representing the two conformations of the glutamine bridge with the associated water molecule which are observed in the X-ray structure. Additionally, the neighboring Glu165 residues are shown. The oscillation of the water molecule through the symmetry axis is visible by moving from one to the other conformation. (b) Snapshot of a third conformation observed in the MD simulation being apparent approximately in the time interval 4.0–10.2 ns. Another water molecule builds a hydrogen bond, whereby the bridge is deactivated. (c) Cartesian coordinate RMSD of both Gln367 residues relative to the reference structure derived from the original crystal structure (see Section 6.1.2). (d) Difference distance between the oxygen atom of the water molecule and the protonated oxygen atoms of the Glu165(A) and Glu165(B) residues. The moving average over 25 ps is shown in red. (e) DBSCAN analysis of the trajectory with the two quantities from graphs (c) and (d) being the descriptors. The analysis employs the data for the time intervals 0–4.0 ns and 10.2–16.5 ns, thereby excluding the conformation shown in (b).

where the third conformation of the glutamine bridge was observed. The method finds two significant clusters, pairing the difference distance around 1.0 Å to 2.5 Å RMSD, as well as −1 Å to 1.0 Å RMSD approximately (Figure 6.7e). This confirms the correlated motion of the water molecule with the conformational change of the Gln367 residues. Both are found to be in one of the partially occupied sites observed in the X-ray structure at the same time.

Summarizing the analysis of the trajectory for the flexibility of the glutamine bridge, the simulation confirms both conformations observed in the crystal for this unit. Several events for switching between them are observed, correlated with the movement of the water molecule through the symmetry axis. Therefore, this oscillation takes place in the nanosecond regime, much faster than the catalyzed reaction time. This observation supports the hypothesis that a shuttling water could be involved in the proton wire, acting as proton relay for the communication between the two active sites of the TK enzyme.

6.3. Conclusions

The DHEThDP intermediate was experimentally observed in the *E. coli* TK enzyme and connected to an uncommon broad absorption band in the range of 2.25–3.55 eV (350–550 nm). In this work, this key intermediate was studied in the active site of TK through the use of TD-DFT/TDA in a QM/MM approach. Hereby, two different states of the substituent of DHEThDP were modeled where in the first case a protonated species (enol-DHEThDP) was considered. In the second case, an enolate form was assumed, thereby transferring the proton to the nearby His473 residue. Furthermore, the 1D-potential for this proton movement was calculated, followed by an anharmonic frequency analysis. The main findings can be summarized as follows.

(1.1) Both the enol and enolate-DHEThDP intermediates show excitations in the spectral region of interest. These are located at the limits of the measured UV-vis band with the enol and enolate-DHEThDP intermediates absorbing in the high and low-energy regimes, respectively.

(1.2) The excitation characters correspond to charge-transfer excitations, where an electron is transferred from the π-system of the thiazolium-substituent moiety to the aminopyrimidinium ring. This newly observed excitation character explains the

large red shift for enolate-DHEThDP through the introduction of a negative charge within the excitation hole.

(1.3) The calculated 1D-potential by scanning the only distance of the proton from enol-DHEThDP reveals a typical anharmonic O–H stretch vibration. The enolate-DHEThDP intermediate cannot be identified as minimum with this simple reaction coordinate.

(1.4) The spectrum derived from the latter potential shows only significant absorbance in the spectral range of 2.91–3.81 eV (325–425 nm). This cannot explain the measured UV-vis band solely arising from the enol-DHEThDP intermediate, so that the enolate-DHEThDP intermediate is required. A more complex reaction coordinate has to be considered, coupling other degrees of freedom of the system like conformational changes in the enzyme environment.

X-ray structural analysis on the dimeric human TK enzyme were carried out in the Tittmann group, leading to the observation of a proton wire which connects the two active sites. This communication channel is proposed to be responsible for positive cooperativity in substrate binding of the enzyme. In between a glutamine bridge with a associated water molecule is detected with two partially occupied sites, corresponding to two different conformations. In this work, the stability of the hydrogen bond network and the dynamics of the glutamine bridge were investigated through the use of classical MD simulations. The main findings can be summarized as follows.

(2.1) The hydrogen bond network between the X5P-ThDP adducts, glutamate residues and water molecules is stable during the whole simulation time. The conformation is also retained relative to the crystal structure. This is required for an efficient proton transfer in solution in order to exhibit communication between the two active sites.

(2.2) Both conformations of the glutamine bridge with the associated water molecule are observed in the trajectory, confirming a shuttling water molecule for the connection of the two sites. A third confirmation is found which deactivates the bridge. However, this movement is recovered in the progression of the system.

(2.3) The DBSCAN analysis reveals a correlated movement of the shuttling water through the symmetry axis with the conformational change of the Gln367 residues. These

are found to be in one of the partially occupied sites observed in the crystal structure at the same time.

(2.4) The oscillation of the glutamine bridge with the associated water molecule takes place in the nanosecond regime, being much faster than the catalyzed reaction time. This fulfills the requirement that the water molecule acts as a proton carrier in the communication channel.

7

Reaction Intermediates of ThDP-Dependent Enzymes

The investigation of ThDP and its derivatives featured in the preceeding chapters revealed new information on their spectral signatures. In particular, the 1',4'-imino-pyrimidine (IP) state of the ThDP cofactor was found to absorb in the range of 4.00–4.13 eV (300–310 nm), resulting from a charge-transfer excitation from the pyrimidine to the thiazolium ring. This is in contrast to the assigned local excitation within the pyrimidine ring deduced from experimental model studies.[23;30] Furthermore, enolate-AcThDP was identified as a key intermediate in the reaction pathway of the phosphoketolase enzyme with an absorption band at about 2.95 eV (420 nm). A different charge-transfer character was observed for this species where the electron is transferred from the substituent to the thiazolium ring. This state of the acetyl moiety was experimentally not related to any UV-vis or CD measurements until now.[24] These results were the motivation for the revision of the spectral assignments of several proposed intermediates in ThDP-catalyzed reactions. The reactants and their associated ThDP intermediates are shown in Figure 7.1.

The intermediates exhibit different structural elements with respect to the bonded substituent. In light of the preliminary findings of this work, the impact on the spectra can be estimated. The results highlight potential inconsistencies in the experimental assignments. The main issues can be summarized as follows:
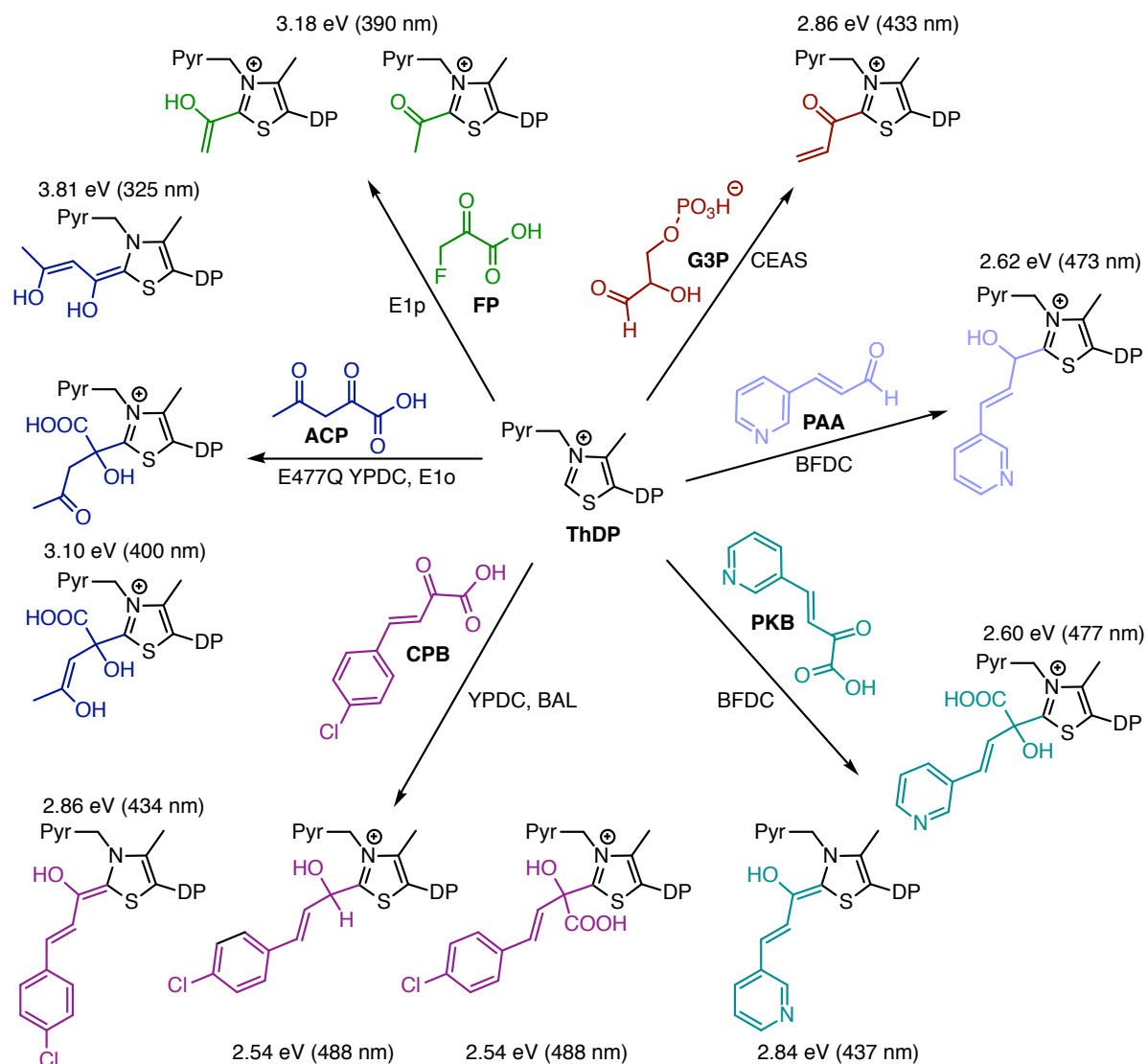
Figure 7.1.: Schematic representation of the proposed intermediates in ThDP-catalyzed reactions combined with their experimentally assigned absorption maxima from UV-vis or CD measurements. The reactants and the related substituents of the intermediates are color-coded. Abbreviations: Pyr – pyrimidine ring of ThDP, DP – diphosphate moiety of ThDP, FP – fluoropyruvate,[24] G3P – glyceraldehyde-3-phosphate,[25] PAA – 3-(pyridin-3-yl)acrylaldehyde,[26] PKB – (*E*)-2-oxo-4-(pyridin-3-yl)but-3-eonic acid,[26] CPB – (*E*)-4-(4-chlorophenyl)-2-oxo-buteonic acid,[24] ACP – acetyl pyruvate,[24] E1p – E1 component of the *E. coli* pyruvate dehydrogenase complex, CEAS – N^2-(2-carboxyethyl)arginine synthase, BFDC – benzoylformate decarboxylase, YPDC – yeast pyruvate decarboxylase, BAL – benzaldehyde lyase, E1o – E1 component of the *E. coli* 2-oxoglutarate dehydrogenase complex.

(1) Extending the degree of conjugation between the thiazolium ring and a co-planar substituent at the C2 atom can strongly shift the charge-transfer transition to lower energies. This was controversially discussed in the case of acryolyl-ThDP (reaction with G3P), where model compounds were used to explain the origin of the band.[25]

(2) The charge-transfer band of IP-ThDP in the range of 4.00–4.13 eV (300–310 nm) is proposed to remain for intermediates, having a tetrahedral coordinated carbon atom bonded to the C2 atom of ThDP. Such intermediates were frequently assigned to bands < 3.5 eV (see reactions with PAA, PKB or CPB in Figure 7.1).[24;26] The origin of these bands are questionable under these conditions.

(3) The protonation state or tautomeric form of a substituent at the C2 atom can significantly influence the spectral signatures of these intermediates and feature prominently in the reaction mechanism. This has often not been given enough consideration in the spectral assignments. The enolate states of the intermediates from the reactions with FP and ACP (see Figure 7.1) were not explicitly taken into account.[24]

(4) The enamine intermediates, resulting after C–C bond cleavage of the substrate-ThDP adducts, exhibit another electronic structure with a conjugated thiazolium-substituent π-system. Therefore, local excitations within this subunit may result in low-energy bands.[23]

(5) Side products were observed for the reaction of ThDP with PKB.[26] However, the associated intermediates were not taken into account for the interpretation of the spectra. In particular, oxidation of the enamine intermediate would strongly extend the degree of conjugation towards the thiazolium ring, possibly leading to low-energy charge-transfer excitations.

In this work, the focus was on the analysis of the excitation character of ThDP intermediates, thereby correlating structural properties with excitation energies. This should provide a guideline for the interpretation of UV-vis or CD spectra of ThDP-dependent enzymes. Additionally, the investigation of these systems offers the opportunity to test the incremental EOM-CCSD method, providing information on its performance to larger systems and charge-transfer excitations.

7.1. Computational Details

The intermediates from ThDP-dependent enzymes were modeled by replacing the phosphate group tail by a hydroxyl group, since this moiety is not included in any observable excitation and its inclusion in a vacuum model could otherwise result in ghost states. The states of the pyrimidine ring were chosen according to the substituent as proposed by experimental studies.[23;29] If the carbon atom of the substituent, bonded to the C2 atom of ThDP, was tetrahedral coordinated, the pyrimidine ring was set to its imino form (IP). Otherwise, the 4'-aminopyrimidine state (AP) of ThDP was chosen. This is almost always the case, if the substituent is conjugated towards the thiazolium ring.

The model compounds were optimized in the V conformation at the B3LYP-D3(BJ)/ def2-SVP level of theory.[61;62;111–113] For all density functional theory (DFT) calculations, the resolution of the identity (RI) algorithm with the corresponding auxiliary basis set was employed to accelerate the calculation of the Coulomb integrals.[114;115] Futhermore, the chain-of-spheres (COSX) algorithm was applied for the exchange part.[116] The spectra were then calculated with time-dependent density functional theory (TD-DFT) within the Tamm-Dancoff approximation (TDA).[72] For this purpose, the CAM-B3LYP functional with the def2-TZVP basis set was applied,[65;113] including the COSMO solvation model for the solvents toluene ($\epsilon = 2.4$, $n = 1.497$) and ether ($\epsilon = 4.34$, $n = 1.35$).[117;123] The subsequent excited state analysis followed the proposed method of Plasser and Lischka.[150] Further details are described below. Additionally, the incremental EOM-CCSD method was applied to selected systems and compared to gas phase TD-DFT/TDA excitation energies. All DFT calculations were performed with the ORCA 4.0.0.2 program package.[123] The incremental EOM-CCSD calculations were carried out with a development version of the MOLPRO 2015.1 program package.[88;89]

In order to verify the level of theory, wave functions methods were employed to selected systems. LR-CC2 excitation energies with a modified aug-cc-pVDZ basis set (A'VDZ),[51;86;87] where the diffuse functions were removed from the hydrogen atoms, were used as reference. Here, the geometries were optimized at the B3LYP-D3(BJ)/def2-TZVP level of theory. For comparison, TD-DFT calculations with and without the TDA approximation, using CAM-B3LYP/def2-TZVP, were carried out on the same structures. Furthermore, calculated spectra of optimized structures with the smaller def2-SVP basis set were considered. The LR-CC2 calculations were carried out with the DALTON 2013 program package.[126;127]

7.2. Method

The method for the analysis of the spectra of ThDP intermediates was tested for three benchmark systems (see Figure 7.2). The unsubstituted ThDP cofactors in both states, 4'-aminopyrimidine (AP) and 1',4'-iminopyrimidine-(IP)-ThDP, were considered, which were already featured in Chapter 4. The spectral signatures in the low-energy regime were found to arise from charge-transfer excitations, where an electron is transferred from the pyrimidine ring to the thiazolium ring. Since these molecules can appear in the resting states of all ThDP-dependent enzymes, they are considered as references to characterize the spectral behavior for substituted (on-pathway) ThDP derivatives. Such an intermediate constitutes enolate-AcThDP, considered as third benchmark molecule. It is observed in the reaction pathway of the phosphoketolase enzyme (see Chapter 5). The introduction of a negatively charged substitutent caused the assignment of a different charge-transfer character for its first excited state. The electron is transferred from the π-system of the enolate moiety to the thiazolium ring. For these three systems it was shown that model calculations including the COSMO solvation model for the environment can qualitatively (and at least semi-quantitatively) describe their spectral signatures. As a first step for the analysis, the benchmark calculations of the different electronic structure methods for the computation of excitation energies were revisited and complemented. LR-CC2 calculations were taken as reference and compared to TD-DFT excitation energies with different technical details employing the CAM-B3LYP functional. The results are listed in Table 7.1.

TD-DFT/TDA calculations of the charge-transfer states of the unsubstituted cofactors, AP and IP-ThDP, were already shown to be in good agreement with the reference LR-CC2 excitation energies. Here, the effect of the TDA approximation was also con-
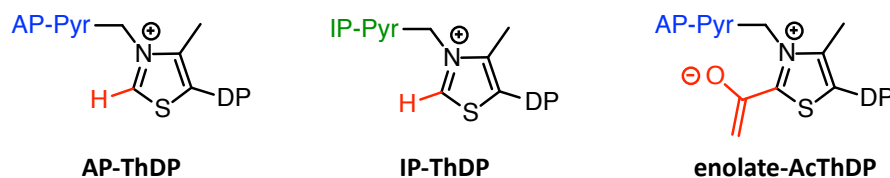
Figure 7.2.: Benchmark systems for the computation of excitation energies and excited state analysis. DP abbreviates the diphosphate portion of ThDP, which is modeled as a hydroxyethyl unit in the calculations. Pyr denotes the pyrimidine ring, being either in the 4'-aminopyrimidine (AP) or 1',4'-iminopyrimidine (IP) state.

Table 7.1.: Vertical electronic excitation energies for the three benchmark molecules shown in Figure 7.2, using different level of theory. All energies are given in electronvolts (eV). The corresponding wavelengths in nanometers (nm) are listed in parentheses.

spectra	TDA	optimization[a]	AP-ThDP	IP-ThDP	2-enolate-AcThDP
LR-CC2[b]	—	def2-TZVP	4.41 (281)	4.36 (284)	2.58 (481)
CAM-B3LYP[c]	yes	def2-TZVP	4.26 (291)	4.33 (286)	2.99 (415)
CAM-B3LYP[c]	no	def2-TZVP	4.25 (292)	4.29 (289)	2.84 (436)
CAM-B3LYP[c]	yes	def2-SVP	4.39 (282)	4.32 (287)	3.01 (412)

[a] Optimizations were carried out at the B3LYP-D3(BJ) level of theory.

[b] Basis: A'VDZ

[c] Basis: def2-TZVP

sidered. The approximation shows nearly no effect with deviations of less than 0.05 eV for both molecules. This is a typical observation for charge-transfer states, where the hole-particle densities are well separated.[137] Furthermore, the spectra were recalculated for optimized structures with a smaller basis set. In the case of AP-ThDP, an increase of the excitation energy by about 0.1 eV is observed. The reason for this is a slightly different orientation between the pyrimidine and thiazolium ring. However, the calculations are still in acceptable agreement. For IP-ThDP the optimization method shows structurally and energetically no effect.

The first excited state of enolate-AcThDP is identified by all methods as a different charge-transfer state, thereby lowering significantly the excitation energy compared to the unsubstituted cofactors. The LR-CC2 excitation energy is even lower by 0.4 eV than the corresponding TD-DFT/TDA results. However, both methods predict the same excitation to be lowest in energy, being well separated to the second excited state by at least 0.6 eV. The TDA approximation raises the excitation energy by 0.15 eV. This is somewhat larger than for the AP and IP-ThDP molecules. In the case of enolate-AcThDP the hole-particle densities are spatially closer, leading to an increase in the error. However, this is a well known fact and the deviation is still in an acceptable range.[72;137] Furthermore, there is no effect observable for the optimized structure with the smaller basis set. The more efficient optimization method is satisfactory for these class of molecules. Altogether, the TD-DFT/TDA with CAM-B3LYP/def2-TZVP//B3LYP-D3(BJ)/def2-SVP level of theory is suitable in describing the spectral signatures of ThDP and its derivatives, so that this method is employed in all further calculations.

Having established a reliable level of theory, the influence of the environment was considered. The COSMO solvation model was employed for the calculation of the first excited states of the benchmark molecules. Since the environment of a hydrophobic enzyme pocket is best described with low dielectric constants ($\epsilon \approx 2 - 4$),[129] toluene and ether were chosen as solvents. The results are shown in Figure 7.3a.

The three molecules show a similar response to the polarization of the environment for their first excited states. The excitation energies are red shifted in toluene relative to the gas phase by 0.25, 0.35 and 0.08 eV for AP-ThDP, IP-ThDP and enolate-AcThDP, respectively. The solvation effect is larger for the unsubstituted cofactors, possibly due to its different charge-transfer character where the hole-particle densities are more separated. Increasing the dielectric constant by moving from toluene to ether raises also the excitation energies. The strongest effect is observed for AP-ThDP with a fairly large blue shift of 0.3 eV. This is not unexpected since charge-transfer excitation energies typically shifts most in the low ϵ-regime.[151] Since the aim is to characterize the spectral signatures of several intermediates proposed to be in different ThDP-dependent enzymes each with its own hydrophobic active site, the arithmetic mean to the COSMO results in toluene and ether was used for the analysis, including a range of two standard deviations ($\omega_{\text{theor}} = \langle\omega\rangle \pm 2\sigma_\omega$). This results in the excitation energy range of 4.23 ± 0.30, 4.07 ± 0.20 and 2.98 ± 0.09 eV for AP-ThDP, IP-ThDP and 2-enolate-AcThDP, respectively.

With the appropriate method for the calculation of the excitation energies, the excited state characters are analyzed in order to identify correlations between the structure of the ThDP intermediates and the resulting spectra. For this purpose the proposed method of Plasser and Lischka was applied.[150] The analysis is based on the fragmentation of a system and the corresponding parts of the one-electron transition density matrix of an excitation. The contribution of the excitation from a fragment A to B is calculated by

$$\Omega_{AB}^{0\alpha} = \frac{1}{2} \sum_{\substack{\mu \in A \\ \nu \in B}} \left(\mathbf{T}^{0\alpha}\mathbf{S}\right)_{\mu\nu} \left(\mathbf{S}\mathbf{T}^{0\alpha}\right)_{\mu\nu} \quad , \tag{7.1}$$

with $\mathbf{T}^{0\alpha}$ being the one-electron transition density matrix for an excitation from the ground state 0 to the excited state α in the AO basis. $\mathbf{S}$ denotes the AO overlap matrix and the indices μ and ν are running over all AOs, which are located at atoms associated with the fragments A and B, respectively. The off-diagonal elements of this $\mathbf{\Omega}$ matrix (Ω_{AB}) can then be interpreted as charge-transfer numbers of fragment A to B, whereby

the diagonal elements Ω_{AA} represent the weight of the local excitation within fragment A to the total excitation.

In the case of ThDP intermediates, the systems were divided into four fragments (see Figure 7.3b): the pyrimidine ring (P), the CH_2-linkage (L), the thiazolium moiety (T) and the substituent (S). This fragmentation should characterize the spectral signatures
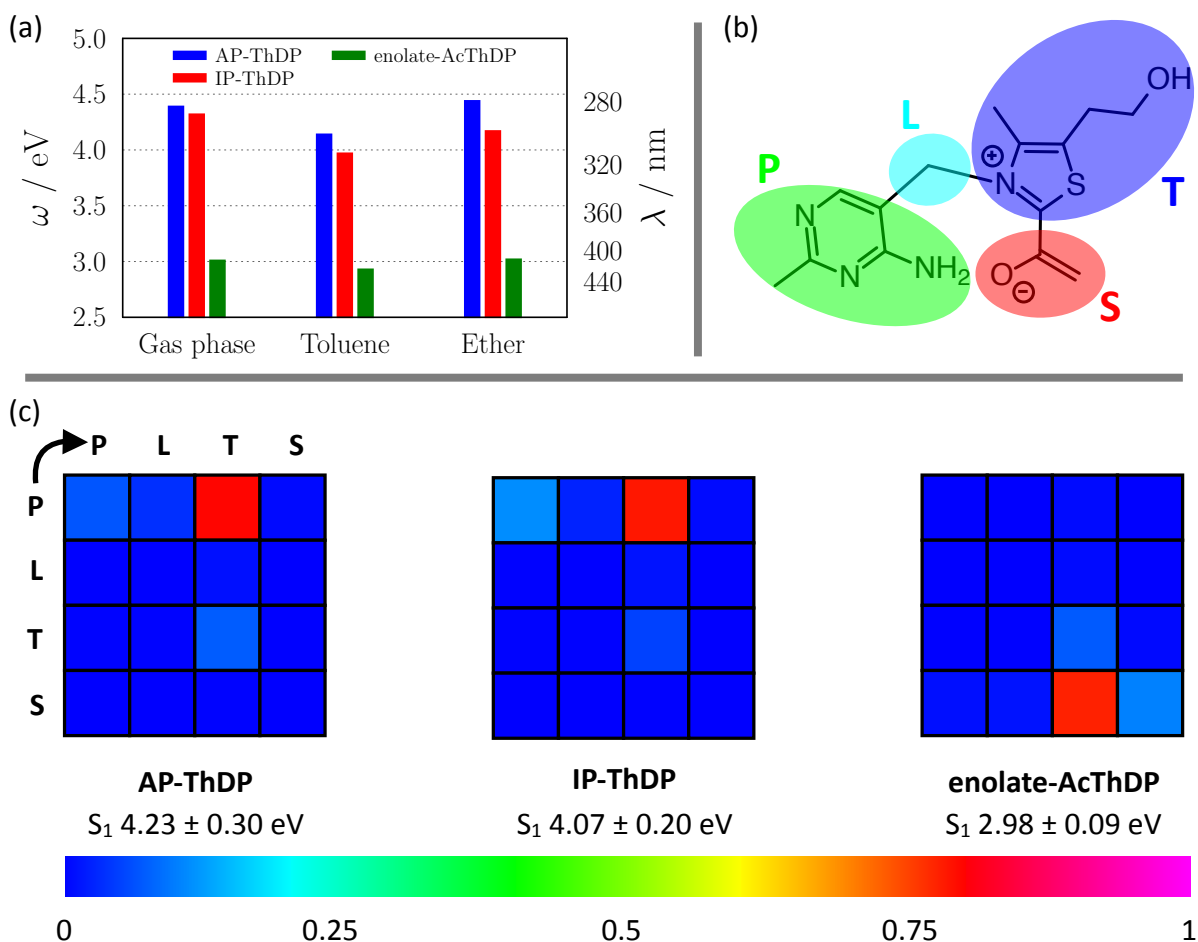
Figure 7.3.: (a) CAM-B3LYP/def2-TZVP//B3LYP-D3(BJ)/def2-SVP vertical excitation energies for the first excited states of the three benchmark systems in different environments, using the COSMO solvation model. (b) Fragmentation scheme of ThDP derivatives for the excited state analysis, using the example of enolate-AcThDP. The systems are divided into four fragments: the pyrimidine ring (P), the CH_2-linkage (L), the thiazolium moiety (T) and the substituent (S). (c) Graphical representation of the charge-transfer matrices for the excited states of the three benchmark molecules. The donor fragments are arranged row-wise, whereby the acceptor fragments are organized column-wise. Furthermore, the average excitation energies in toluene and ether are given.

of the different intermediates best, since their main differences arise from the substituent. The charge-transfer numbers are obtained by averaging the individual values for the two solvents, toluene and ether. In the graphical representation of the matrix, the elements are shown as colored squares. The donor fragments are arranged row-wise, whereby the acceptor fragments are organized column-wise. The results for the benchmark molecules are shown in Figure 7.3c.

The different character of charge-transfer states of the benchmark molecules can be clearly distinguished. For the AP and IP-ThDP cofactors the only observable contribution comes from the Ω_{PT} elements (charge-transfer from the pyrimidine to the thiazolium ring). Hereby, the substituent elements are just associated with the hydrogen atom bonded to the C2 atom of ThDP. Therefore, their corresponding charge-transfer numbers are negligibly small. The CH_2-linkage makes also no contribution, since it is hardly involved in any excitation of this class of molecules. These results should be considered as references for the following study of enzymatic on-pathway intermediates.

The situation differs for the lowest excited state of enolate-AcThDP. The contribution of a charge-transfer from the pyrimidine to the thiazolium ring is negligible. The character of the excitation has changed. This is well reflected in the charge-transfer numbers and their graphical representation. The main contibution arises now from the charge-transfer of the electron-rich substituent to the thiazolium moiety (Ω_{ST}). All other elements of the $\mathbf{\Omega}$ matrix are close to zero.

7.3. Results and Discussion

7.3.1. Overview

The method described in the preceeding section was applied to the model compounds of the proposed intermediates in ThDP-catalyzed reactions. Additionally, further protonation states of the substituents as well as possible intermediates from side reactions were considered. The structures are schematically represented in Figure 7.4. Furthermore, the calculated excitation energies for their first excited states are given and compared to the experimentally assigned band maxima.

A

2.80 ± 0.26 eV (2.86 eV)

443 ± 42 nm (433 nm)

B

3.12 ± 0.23 eV (3.18 eV)

398 ± 30 nm (390 nm)

C

3.66 ± 0.33 eV (3.18 eV)

338 ± 30 nm (390 nm)

D

2.98 ± 0.09 eV (———)

417 ± 12 nm (———)

E

4.21 ± 0.05 eV (3.10 eV)

295 ± 3 nm (400 nm)

F

3.92 ± 0.14 eV (3.10 eV)

316 ± 11 nm (400 nm)

G

2.44 ± 0.22 eV (———)

507 ± 46 nm (———)

H

3.82 ± 0.01 eV (3.81 eV)

325 ± 1 nm (325 nm)

I

3.14 ± 0.01 eV (2.86 eV)

395 ± 1 nm (434 nm)

J

3.95 ± 0.30 eV (2.54 eV)

314 ± 24 nm (488 nm)

K

2.85 ± 0.26 eV (———)

435 ± 40 nm (———)

L

4.10 ± 0.31 eV (2.60 eV)

303 ± 23 nm (477 nm)

M

3.12 ± 0.01 eV (2.84 eV)

398 ± 1 nm (437 nm)

N

3.84 ± 0.27 eV (2.62 eV)

322 ± 23 nm (473 nm)

O

3.95 ± 0.22 eV (———)

314 ± 18 nm (———)

P

3.03 ± 0.18 eV (———)

409 ± 24 nm (———)

Figure 7.4.: Overview of the investigated ThDP intermediates. The calculated excitation energies at the TD-DFT/TDA level of theory (CAM-B3LYP/def2-TZVP), and the corresponding wavelengths, are given below each intermediate. The excitation energies are averaged for the solvents toluene and ether, employing the COSMO solvation model for the representation of the environment. The error bar is purely provided for solvent effects. The experimentally assigned values are shown in parentheses.[24–26]

In the reaction of ThDP with glyceraldehyde-3-phosphate (G3P) a band at 2.86 eV (433 nm) was observed and connected to the formation of acryolyl-AcThDP (**A**).[25] The calculated excitation energy of 2.80±0.26 eV (433±42 nm) almost perfectly matches this value. The excitation is found to be of charge-transfer character whereby the electron is transferred from the pyrimidine to the thiazolium ring delocalized to the substituent (similar to the unsubstituted cofactors). This was controversially discussed in the work of Townsend *et al.*[25] The authors deemed a shift of about 1 eV related to the unsubstituted cofactors to be too large to arise from charge-transfer excitations. However, the calculations clearly prove the opposite. The extension of the degree of conjugation of the thiazolium ring towards the substituent strongly influences this charge-transfer excitation, resulting in shifts as large as those observed for acryolyl-ThDP.

The reaction with fluoropyruvate (FP) with ThDP in the E1 component of the *E. coli* pyruvate dehydrogenase complex revealed an absorption band at 3.18 eV (390 nm).[24] This band was associated with the formation of acetyl-ThDP (AcThDP), which was already extensively studied in the reaction mechanism of the phosphoketolase enzyme within this work (see Chapter 5). In the latter the excitation energy was found to be slightly red shifted to 2.95 eV (420 nm), most likely due to a different polarization of the cofactors in the enzyme environments. However, in the work of Patel *et al.* only the keto and enol states of AcThDP (**B** and **C**) were explicitly considered in the assignment, building no distinction between them.[24] The results show that the keto-AcThDP intermediate absorbs in the spectral region of interest, whereas the excitation of enol-AcThDP requires significantly larger energies. Both tautomers identify a similar charge-transfer character compared to the unsubstituted cofactors for these excited states. Furthermore, the enolate form of AcThDP (**D**) was not considered in the interpretation of the experimental spectra. A different charge-transfer character (from the substituent to the thiazolium ring) causes a low-energy band in the range of 2.88 ± 0.09 eV (414 ± 12 nm), somewhat close to the band maximum.

A related issue is observed for the intermediates arising from the reaction of acetyl pyruvate (ACP) with ThDP. The measured absorption band at 3.10 eV (400 nm) was assigned to the ACP-ThDP adducts in their keto or enol form (**E** and **F**).[24] The enolate state (**G**) of the substituent was not considered. The first excited state of the keto-ACP-ThDP adduct shows the same charge-transfer character as the IP-ThDP cofactor, resulting in comparable excitation energies (4.21 ± 0.05 eV; 295 ± 3 nm). This is expected since the substituent with its tetrahedral coordinated carbon atom (bonded

to the C2 atom of ThDP) almost retains the electronic structure within the thiazolium ring. Therefore, the intermediate can be excluded for the assignment of the absorption band at 3.10 eV (400 nm). Moving to the enol species, the character of the first excited state changes. Here, a charge-transfer from the enol moiety to the thiazolium ring is observed. However, the excitation energy of 3.92 ± 0.14 eV (316 ± 11 nm) is still too high to explain the origin of the band. The enolate-ACP-ThDP adduct also has to be considered. The introduction of a negative charge causes the calculated excitation energy to drop down to 2.44 ± 0.22 eV (507 ± 46 nm). The effect is certainly overestimated in the employed model, where specific hydrogen bonds are missing in the continuum representation of the enzyme environment. The latter would raise the excitation energy towards the observed band maximum at 3.10 eV (400 nm). Moreover, a dynamic equilibrium between the enol and enolate-ACP-ThDP adducts would be conceivable, so that the absorption could even arise in between the two limiting cases. However, the crucial point for the interpretation of the spectra is the inclusion of the enolate species in the reaction mechanism.

The decarboxylation of the ACP-ThDP adduct results in the enamine intermediate (**H**), exhibiting a different electronic structure with a delocalized and neutral π-system of the thiazolium-substituent portion. The band maximum was experimentally estimated to 3.81 eV (325 nm).[24;152] The agreement with the calculated excitation energy of 3.82 ± 0.01 eV (325 ± 1 nm) is striking. A local $\pi \rightarrow \pi^*$ character is, as expected, determined for this excitation. A similar behavior is observed for the postdecarboxylation intermediate for the reaction of (*E*)-4-(4-chlorophenyl)-2-oxo-buteonic acid (CPB) with ThDP (**I**). The extension of the π-system causes the local $\pi \rightarrow \pi^*$ excitation to decrease to 2.86 eV (434 nm). The deviation of about 0.2 eV to the calculated excitation energy is somewhat larger, but still acceptable given the intrinsic error of the employed method.

Besides the assigned absorption band for the enamine intermediate in the reaction of CPB with ThDP, another band was observed with maximum at 2.54 eV (488 nm).[24] This low-energy band was connected to the formation of the second postdecarboxylation intermediate (**J**). This one arises from the protonation of the carbon atom bonded to the thiazolium ring in the enamine state, thereby removing the conjugation between the thiazolium ring and the substituent. The electronic structure is now comparable to the enol-ACP-ThDP adduct (**F**). As a result, the first excited state of the intermediate **J** shows charge-transfer character from the π-system of the substituent to the

thiazolium ring, but the excitation energy of 3.95 ± 0.30 eV (314 ± 24 nm) is also much too high in order to explain the measured band. Missing the possibility of the formation of an enolate species, an intermediate of a side reaction was considered. The oxidation of enamine-CPB-ThDP (**I**) would lead to an intermediate with a fully conjugated thiazolium-substituent π-system (**K**). Consequently, the charge-transfer energy (from the pyrimidine ring to the thiazolium ring delocalized to the substituent) decreases significantly compared to the unsubstituted cofactors, as already seen for acryolyl-ThDP (**A**) and keto-AcThDP (**B**). The calculated excitation energy of 2.85 ± 0.26 eV (435 ± 40 nm) is much closer to the observed band maximum at 2.54 eV (488 nm). This is the best candidate for the assignment to the low-energy absorption band. The occurence of the side reaction is not unlikely, since the equivalent side products were detected in the similar reaction of ThDP with (*E*)-2-oxo-4-(pyridin-3-yl)but-3-eonic acid (PKB).[26]

The ThDP derivatives arising from the reaction with PKB (**L**–**P**) completes the set of studied intermediates. Thereby, many similarities are found in comparison to the previously discussed systems. The PKB-ThDP adduct (**L**) was experimentally assigned to the low-energy band with maximum at 2.60 eV (477 nm).[26] The structural motif of this intermediate is comparable to the enol-ACP-ThDP adduct (**F**) and the second postdecarboxylation intermediate (**J**) in the reaction with CPB. Consequently, the lowest calculated excitation energy of 4.10 ± 0.31 eV (303 ± 23 nm) is much too high in order to explain the band. The same holds for the intermediate **N**, where the carboxyl group is just replaced by a hydrogen atom. However, ThDP derivatives from side reactions with the enamine intermediate (**M**) as central point would constitute alternatives. A protonation of the Cγ atom of the substituent results in a partly conjugated thiazolium-substituent π-system (**O**), but this is not sufficient to shift the charge-transfer state related to the unsubstituted cofactors as much as experimentally observed. On the other hand, the oxidation of enamine-PKB-ThDP (**M**) reveals an intermediate with a fully conjugated thiazolium-substituent π-system (**P**), resulting in a low-energy charge-transfer excited state.

The results discussed so far suggest that the spectral signatures of ThDP derivatives can be divided into three classes. Firstly, the charge-transfer excitations from the pyrimidine ring to the thiazolium ring (delocalized to the substituent) which are strongly dependent on the degree of conjugation between the thiazolium ring and the substituent. Secondly, charge-transfer states where the electron is transferred from the substituent to the thiazolium ring and at last local $\pi \rightarrow \pi^*$ transitions for the enamine interme-

diates. These excitation classes can clearly be distiguished by the analysis method of Plasser and Lischka,[150] described in the preceeding section. This method was applied to the excited states of the intermediates. The results are presented for each class in the following sections.

7.3.2. Excitation Class I

The excited state analysis method of Plasser and Lischka was applied to the spectra calculations of ThDP intermediates.[150] Thereby, the systems were divided into four fragments: the pyrimidine ring (P), the CH_2-linkage (L), the thiazolium moiety (T) and the substituent (S). The calculated charge-transfer numbers Ω_{AB} between these fragments were employed to classify their spectral signatures. The first excitation class considers the charge-transfer from the pyrimidine to the thiazolium ring (delocalized to the substituent), being the equivalent excitations related to the AP and IP-ThDP cofactors in the resting states of enzymes. The graphical representation of the charge-transfer matrices $\mathbf{\Omega}$ for these excited states combined with their calculated excitation energies are collected in Figure 7.5.

Regarding the representation of the charge-transfer matrices for this excitation class, the systems can be further divided into two groups. These reflect two structural motifs of the intermediates. In the first case the substituent is at least partly coplanar towards the thiazolium ring (**A**, **B**, **C**, **K**, **O** and **P**), sharing a conjugated thiazolium-substituent π-system. Besides the prominent Ω_{PT} element of this excitation class, these intermediates feature a significant contribution of the charge-transfer from the pyrimidine ring to the substituent (Ω_{PS}). The result is an expansion of the particle space of the excitation, while the hole space remains nearly constant. Therefore, the charge-transfer excitation energy is red shifted according to the unsubstituted cofactors, which were observed in the energy range of 4.0–4.2 eV in the model calculations. Furthermore, a correlation is observed for the size of the red shift of the excitation energy and the contribution Ω_{PS}, being associated with the degree of conjugation between the thiazolium ring and the substituent. Increasing the charge-transfer number from the pyrimidine ring to the substituent (Ω_{PS}), the excitation energy is lowered. This is well reflected, considering just the first row in Figure 7.5. Moving from acryolyl-ThDP (**A**) over keto-AcThDP (**B**) to enol-AcThDP (**C**) the calculated excitation energies are 2.80 ± 0.26, 3.12 ± 0.23 and 3.66 ± 0.33 eV, respectively. Another issue is observed for the charge-transfer matrix

Figure 7.5.: Graphical representation of the charge-transfer matrices for the excited states of ThDP intermediates, showing a charge-transfer character from the pyrimidine to the thiazolium ring. The donor fragments are arranged row-wise, whereby the acceptor fragments are organized column-wise. The notation follows the labeling in Figure 7.4. Furthermore, the calculated excitation energies at the TD-DFT/TDA level of theory (CAM-B3LYP/def2-TZVP), averaged for the solvents toluene and ether, are given.

of intermediate **P**. Here, a non-negligible contribution of a local excitation within the substituent (Ω_{SS}) arises. Clearly, this raises the charge-transfer energy. For the similar intermediate from the reaction with CPB (**K**) the Ω_{SS} is close to zero, so that its excitation energy is lower by about 0.2 eV. This feature may, however, be an artifact of the employed computational model, including just a continuum representation of the enzyme environment. The crucial point is the correct trend for both systems with a strong red shift according to the unsubstituted cofactors.

Summarizing the results for the first group of this excitation class, an extension of the conjugation of the π-system of the thiazolium ring through the substituent generally decreases the charge-transfer excitation energy compared to the reference AP and IP-ThDP systems. The size of this red shift correlates with the degree of conjugation (contribution Ω_{PS}). The shift can be very strong with even more than 1 eV. In conclusion, such intermediates have to be considered for the interpretation of experimental spectra, where absorption bands with maxima <3.5 eV (>354 nm) are observed.

The second group of intermediates within this excitation class reveals a different structural motif (**E**, **F**, **J**, **L** and **N**). Here, the carbon atom bonded to the thiazolium ring is tetrahedral coordinated, so that the degree of conjugation of the thiazolium ring is retained with respect to the unsubstituted cofactors. In consequence, the charge-transfer from the pyrimidine to the thiazolium ring will not be delocalized towards the substituent. This is well reflected in the charge-transfer matrices, having negligible Ω_{PS} contributions for all these systems. This results in a different trend for the excitation energies compared to the first group of molecules. For the second group of intermediates, the charge-transfer energy remains constant or is even blue shifted with respect to the references. The size of the blue shift seems now to correlate with a contribution of a local excitation within the pyrimidine ring (Ω_{PP}). Obviously, the larger the Ω_{PP} contribution the stronger is the blue shift (compare for example intermediates **L** and **N** in Figure 7.5). If this effect agrees indeed with the reality or is just an artifact of the employed computational model is unclear as already discussed before. Here, the crucial point is that the charge-transfer energy remains at least constant, comparing to the reference states. Intermediates included in this set are typically substrate-ThDP adducts or related compounds which were frequently assigned to low-energy bands with maxima <3.5 eV (>354 nm).[24;26] These assignments cannot be rationalized by the charge-transfer excitation from the pyrimidine to thiazolium ring. An alternative would be the occurence of a different excitation character for these systems, leading to the

second excitation class in the next section.

7.3.3. Excitation Class II

The second excitation class of ThDP intermediates is characterized by charge-transfer excitations from the substituent to the thiazolium ring. Therefore, the excited state analysis identifies these states through a prominent Ω_{ST} element in their charge-transfer matrices. These are collected in Figure 7.6 and combined with the calculated excitation energies.

The dominant element of the charge-transfer matrices arise naturally from the transition from the substituent to the thiazolium ring (Ω_{ST}). For some intermediates (**J**, **K** and **O**) contributions of local excitations within the thiazolium ring or the substituent are observed. However, no clear correlation between the elements of the charge-transfer matrices and the calculated excitation energies can be detected. This issue is well reflected, taking enol and enolate-ACP-ThDP adducts as example (**F** and **G**). The only significant contribution for the excited state of enol-ACP-ThDP is the Ω_{ST} element, representing the charge-transfer from the π-system of the enol moiety to the thiazolium ring with an excitation energy of 3.92 ± 0.14 eV. Moving to the enolate form the charge-transfer matrix reflects the identical excitation character, but the introduction of a negative charge for the electron donor decreases, as expected, significantly the excitation energy to 2.44 ± 0.22 eV.

Although the excited state analysis is not competent to identify a correlation between the excitation energy and the charge-transfer matrix for this excitation class, a relation can be observed for the structural motifs of the intermediates and their spectral signatures. For this purpose, the systems with a neutral π-system within the substituent are considered, where the carbon atom bonded to the thiazolium ring is tetrahedrally coordinated (**F**, **J**, **L**, **N** and **O**). The latter means that the π-system of the thiazolium ring is retained with respect to the unsubstituted cofactors. Therefore, these excitations can be regarded as related charge-transfer transitions to the references with the π-system of the substituent mimicking the pyrimidine ring as electron donor, while the thiazolium ring as electron acceptor is retained. Assuredly, the excitation energies will be shifted dependent on the electronic structure of the substituent comparing to the unsubstituted cofactors, but this effect is not expected to be large. The calculated excitation energies of the intermediates are all above 3.5 eV (354 nm). Low-energy bands (<3.5 eV) frequently

observed in experiments and assigned to ThDP intermediates containing these structural motifs cannot be explained by the charge-transfer excitation from the substituent to the thiazolium ring. In contrast to the reference AP/IP-ThDP molecules an opportunity can be provided by the substituent to introduce a negative charge in the electron donor unit through proton transfer to the environment. This leads to a significant decrease
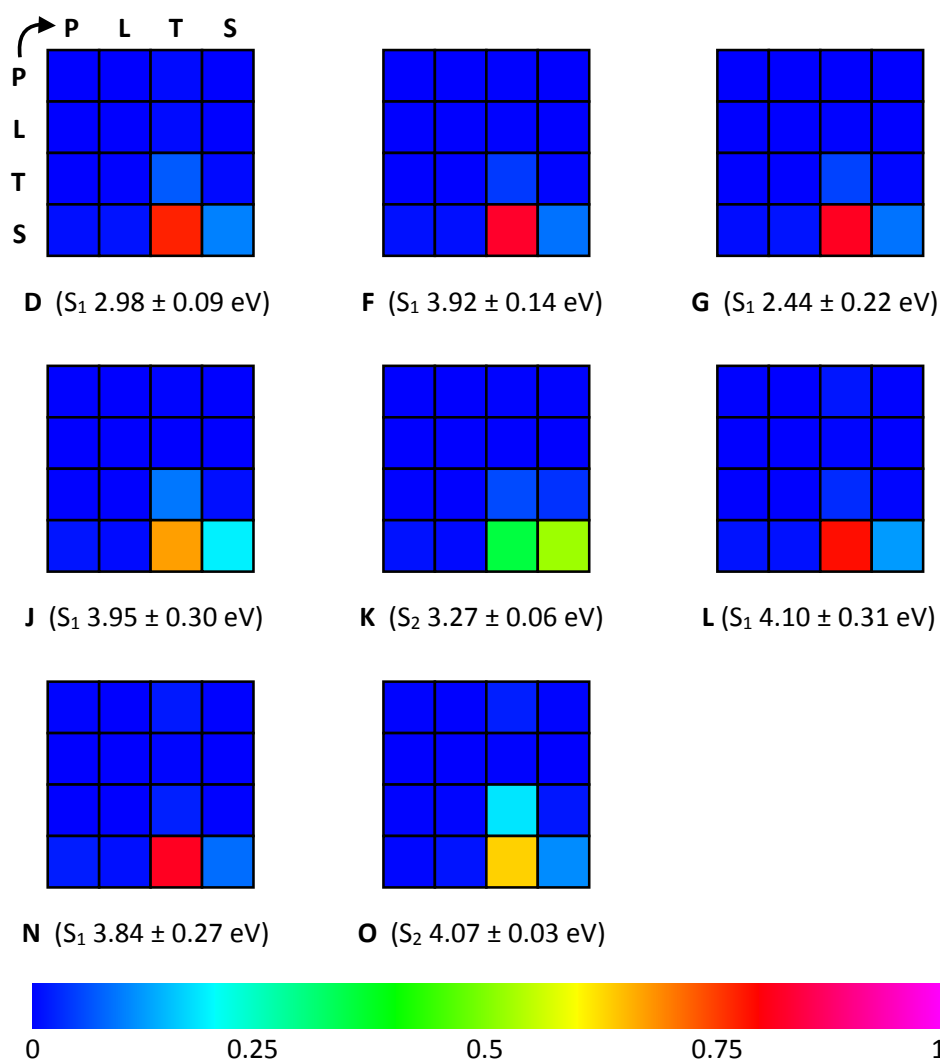
Figure 7.6.: Graphical representation of the charge-transfer matrices for the excited states of ThDP intermediates, showing a charge-transfer character from the substituent to the thiazolium ring. The donor fragments are arranged row-wise, whereby the acceptor fragments are organized column-wise. The notation follows the labeling in Figure 7.4. Furthermore, the calculated excitation energies at the TD-DFT/TDA level of theory (CAM-B3LYP/def2-TZVP), averaged for the solvents toluene and ether, are given.

of the charge-transfer energy as observed for the enolate species in the molecule set (**D** and **G**). Therefore, such states of ThDP intermediates have to be considered for the interpretation of absorption spectra, where low-energy bands are observed.

A special case in this excitation class constitutes the proposed intermediate of a side reaction of CPB with ThDP (**K**), where the oxidation of enamine-CPB-ThDP (**I**) leads to this intermediate with a fully conjugated thiazolium-substituent π-system. The first excited state is characterized by a charge-transfer from the pyrimidine to the thiazolium ring. This was already discussed in the preceeding sections. Here, the focus is on the second excited state. The Ω_{ST} element in the charge-transfer matrix holds the largest value, so that it is taken into the second excitation class. However, the local $\pi \rightarrow \pi^*$ contribution of the substituent is nearly equivalent. Therefore, this transition can also be regarded as a mixed charge-transfer/local excitation. The conjugation of the substituent to the thiazolium ring causes the excitation energy to be red shifted with respect to the references. The charge-transfer contribution arises through the formally positive charge at the thiazolium. In order to obtain a local $\pi \rightarrow \pi^*$ transition for the thiazolium-substituent moiety the positive charge has to be removed. This is the case for the enamine intermediates, leading to the third excitation class.

7.3.4. Excitation Class III

The third excitation class collects local $\pi \rightarrow \pi^*$ transitions of the conjugated thiazolium-substituent π-system. These are commonly observed for enamine intermediates, formed after C–C bond cleavage of the substrate-ThDP adducts.[23;24;26;152] Three systems were studied (**H**, **I** and **M**), exhibiting this structural motif. The charge-transfer matrices combined with the calculated excitation energies are shown in Figure 7.7.

In the case of local excitations the charge-transfer matrices have to be interpreted with some care. As expected, the analysis reveals significant contributions of the local Ω_{TT} and Ω_{SS} elements. Additionally, there are non-negligible contributions of the charge-transfer elements from the thiazolium ring to the substituent (Ω_{TS}) and *vice versa* (Ω_{ST}). These terms can here not be attributed to a charge-transfer character in the ordinary sense, since they artificially stem from the applied fragmentation scheme. The π-system of the conjugated thiazolium-substituent moiety is cut within this scheme, while the hole-particle spaces of the local $\pi \rightarrow \pi^*$ are extended over this region. Therefore, these elements should be accounted as local contributions for the third excitation class. The
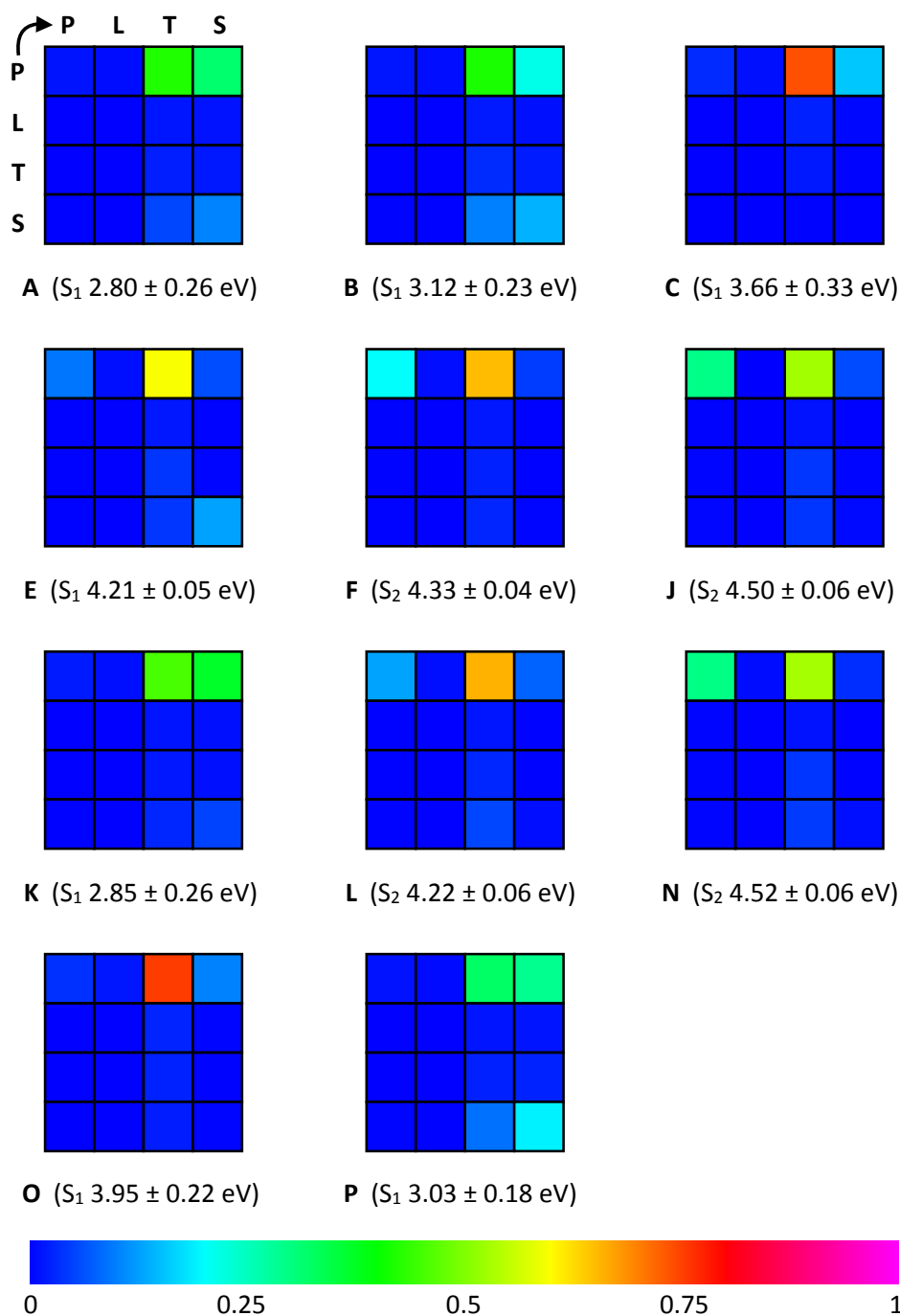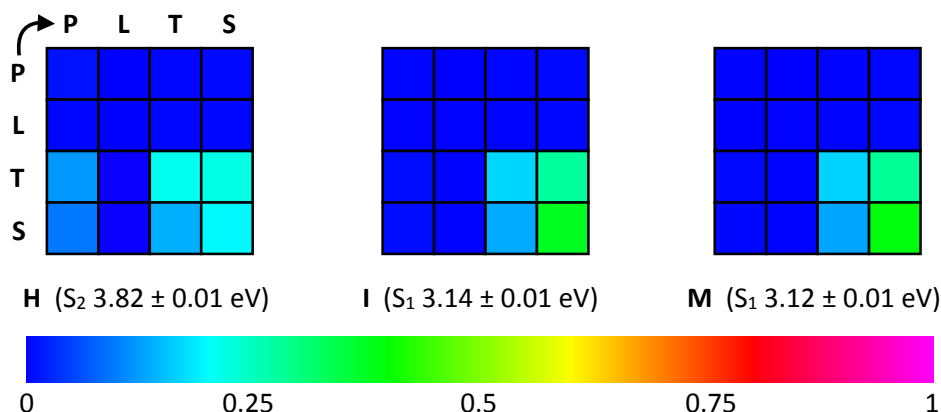
Figure 7.7.: Graphical representation of the charge-transfer matrices for the excited states of ThDP intermediates, showing a local $\pi \to \pi^*$ character within the thiazolium-substituent moiety. The donor fragments are arranged row-wise, whereby the acceptor fragments are organized column-wise. The notation follows the labeling in Figure 7.4. Furthermore, the calculated excitation energies at the TD-DFT/TDA level of theory (CAM-B3LYP/def2-TZVP), averaged for the solvents toluene and ether, are given.

locality is well reflected by summing up the four elements of the bottom right block of the charge-transfer matrices. In this way, the weight for the excitation within the thiazolium-substituent unit is larger than 0.95 for the three systems, identifying the proposed local character.

Although the applied fragmentation scheme is at first glance inconsistent for this class of excitations, advantage can also be taken from this. Regarding the results in more detail, a correlation between the calculated energies and the diagonal Ω_{SS} element can be identified. In fact, a larger contribution of the local $\pi \to \pi^*$ transition within the substituent indicates a larger conjugated π-system for the enamine intermediates, so that the required excitation energy is decreased. The transition energy of the enamine-ACP-ThDP system (**H**) points at 3.82 ± 0.01 eV. Moving to the similar enamine-CPB-ThDP (**I**) and enamine-PKB-ThDP (**M**) intermediates, the excitation energies are accordingly red shifted to 3.14 ± 0.01 and 3.12 ± 0.01 eV, respectively. Therefore, such intermediates can be responsible for the observation of low-energy bands (<3.5 eV), whereby the position of the band depends strongly on the size of the conjugated π-system.

A last remark should be discussed for the spectral signatures of enamine-ThDP intermediates. For the 1,2-dihydroxyethyl-ThDP (DHEThDP) intermediate, studied in the transketolase enzyme (see Chapter 6), a charge-transfer from the thiazolium-substituent π-system to the pyrimidine ring was observed. Thereby, the pyrimidine ring was found

to be in its APH$^+$ state. The introduction of a positive charge within the electron acceptor unit causes the excitation to be lower than the local $\pi \to \pi^*$ transition. For the analysis of the spectral signatures of the enamine intermediates featured in this chapter, the state of the pyrimidine ring was tentatively assumed to be in the neutral AP state. This raises the charge-transfer excitation, so that it was excluded for the analysis. However, the investigation of the DHEThDP intermediate also revealed the requirement of an enolate state of the substituent in order to absorb in the low-energy regime (<3.5 eV). Following the argumentation discussed for the charge-transfer excitations from a neutral π-system of the substituent to the thiazolium ring from the preceeding section, a neutral thiazolium-substituent unit combined with an APH$^+$ state of the pyrimidine ring is not expected to produce low-energy bands. This hypothesis was, however, not investigated in further detail.

7.3.5. Application of the Incremental EOM-CCSD Method

The incremental EOM-CCSD method refined within this work (Chapter 3) was tested for the spectral signatures of ThDP and its derivatives. This should provide further information on the performance for larger systems and different excitation characters as charge-transfer states, not being considered before. Since the virtual space is not restricted in the current implementation of the incremental approach, the calculations for these systems are computationally demanding. Therefore, only a subset of molecules and excited states were considered. This collection includes the charge-transfer excitations of the unsubstituted cofactors in both tautomeric states. For AP-ThDP, a slightly different optimized structure was used, compared to the preceeding calculations. Hereby, a small imaginary frequency was detected, which should, however, not influence the comparison of vertical excitation energies. Additionally, acryoyl-ThDP (**A**), keto-AcThDP (**B**), enol-AcThDP (**C**) as well as the oxidized enamine intermediate from the side reaction with PKB (**P**) were added to the set from excitation class I. Furthermore, two systems from excitation class II (charge-transfer from the substituent to the thiazolium ring) were considered, the enolate-AcThDP itermediate (**D**) and product-ThDP adduct from the reaction with CPB (**J**). The system set was completed by the local $\pi \to \pi^*$ transitions (excitation class III) of the enamine intermediates from the reactions with ACP (**H**) and PKB (**M**).

The incremental EOM-CCSD calculations were carried out with the group construc-

tion algorithm B as denoted in Chapter 3. This was the one showing the best performance for the benchmark calculations. The algorithm is based on the mixed NTO-LMO orbital space from the CIS reference calculation for the determination of the dominant NTOs and the subsequent orbital localization of the remaining valence orbitals. The main fragment is then built according to the overlap between the hole/particle and LMO densities with a group size limit of six orbitals, whereas the other fragments are constructed by the LMO-LMO exchange interactions with a group size limit of four orbitals. Vertical excitation energies were then calculated in a two-body (2B) approach at the EOM-CCSD/A'VDZ level of theory for the increments.[39;86;87] Since the full EOM-CCSD calculations were not accessible for these system sizes, comparisons are made to gas phase TD-DFT/TDA excitation energies employing CAM-B3LYP/def2-TZVP. These methods should at least give similar results, so that large deviations to the incremental approach could indicate a problem of the procedure. The results are listed in Table 7.2. The table contains additional data from *a posteriori* analysis of the incremental EOM-CCSD calculations discussed further below.

The comparison between the incremental EOM-CCSD excitation energies and the TD-DFT/TDA results identifies two sets of the studied excited states. In the first case, the calculated excitation energies are strongly overestimated by the incremental approach with deviations larger than 0.5 eV (AP-ThDP, IP-ThDP, **A**, **B**, **P** and **J**). These excited states have all charge-transfer character (excitation classes I and II). In contrast to this, the local $\pi \rightarrow \pi^*$ transitions (excitation class III) reveal only small deviations of $|\Delta\omega| < 0.2$ eV as well as the different charge-transfer excitations of enol-AcThDP (**C**) and enolate-AcThDP (**D**). This energy difference is by far in the error range of the applied methods, indicating a successful performance of the incremental approach. In order to identify the origin of this discrepancy, the dominant NTOs were compared resulting from the CIS reference calculation of the incremental method and the TD-DFT/TDA calculations. Exemplarily, these are shown for keto-AcThDP (**B**) and enol-AcThDP (**C**) in Figure 7.8.

The occupied CIS NTO of keto-AcThDP (**B**) is delocalized to some extent towards the thiazolium ring compared to the TD-DFT/TDA NTO, while the virtual NTOs share a similar shape. This delocalization of the hole density of the reference CIS calculation towards the acceptor moiety is sensitive to the charge-transfer and raises significantly the excitation energy. In the case of enol-AcThDP (**C**) the occupied NTOs of both methods are quite similar and well localized at the pyrimidine ring. Here, the virtual

Table 7.2.: Vertical excitation energies of ThDP and its derivatives calculated with the incremental EOM-CCSD method ($\omega_{\mathrm{iEOM}}^{\mathrm{2B}}$) and compared to gas phase TD-DFT/TDA results at the CAM-B3LYP/def2-TZVP level of theory (ω_{CAM}). Their deviations are denoted by $\Delta\omega$. All energies are given in electronvolts (eV). The labels of the ThDP intermediates follow the notation introduced in Figure 7.4. The excitation characters are described by the classification from the preceeding sections. Furthermore, the number of relevant NTOs from the CIS calculations ($n_{\mathrm{NTO}}^{\mathrm{CIS}}$) are given, which were determined according to a threshold of 0.05 for their singular values. The descriptor for the deviation between the dominant occupied NTOs from the CIS and TD-DFT/TDA calculations is denoted by $\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$ (details are described in the text). The energy deviations for the exclusion of 2B increments ($\Delta\omega_{\mathrm{excl}}^{\mathrm{2B}}$) as well as the fraction of excluded groups ($n_{\mathrm{excl}}^{\mathrm{2B}}/n_{\mathrm{tot}}^{\mathrm{2B}}$) are given. The exclusion criterion is described in the text.

molecule	class	$\omega_{\mathrm{iEOM}}^{\mathrm{2B}}$	ω_{CAM}	$\Delta\omega$	$n_{\mathrm{NTO}}^{\mathrm{CIS}}$	$\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$	$\Delta\omega_{\mathrm{excl}}^{\mathrm{2B}}$	$n_{\mathrm{excl}}^{\mathrm{2B}}/n_{\mathrm{tot}}^{\mathrm{2B}}$
AP-ThDP	I	5.07	4.51	+0.56	3	0.69	0.02	2/15
IP-ThDP	I	4.88	4.32	+0.56	2	0.23	0.02	2/13
A	I	3.60	2.84	+0.76	1	0.51	0.01	2/18
B	I	3.85	3.11	+0.74	1	0.53	0.02	2/18
C	I	3.81	3.79	+0.02	1	0.40	0.01	2/17
P	I	3.79	3.09	+0.70	2	0.65	0.01	3/21
D	II	3.01	3.01	±0.00	1	0.30	0.03	7/16
J	II	5.06	3.85	+1.21	3	0.87	0.01	2/22
H	III	3.66	3.85	−0.19	1	0.22	0.01	4/18
M	III	3.08	3.24	−0.16	1	0.32	0.02	7/22

CIS NTO is delocalized between the two rings. However, the virtual space is much more flexible, so that this can be compensated within the EOM-CCSD calculations. These findings explain the better performance of the incremental approach for the charge-transfer excitation of enol-AcThDP (**C**). Indeed, the delocalization of the hole density of the CIS reference towards the acceptor unit of the charge-transfer states were observed for nearly all cases where the excitation energies of the incremental EOM-CCSD method are significantly overestimated. In contrast to this, the occupied NTOs of both methods share a similar shape when the incremental approach showed a succesful performance. In order to quantify these observations and provide a descriptor, which indicates the quality of the CIS reference for charge-transfer states, the densities of the dominant occupied NTOs from both methods were considered. The difference density was then built and its absolute values were numerically integrated on a grid, according to the

following equation

$$\Delta n_{\mathrm{NTO}}^{\mathrm{h}} = \sum_{i}^{N_{\mathrm{grid}}} \left| |\phi_{\mathrm{NTO}}^{\mathrm{CIS}}|_i^2 - |\phi_{\mathrm{NTO}}^{\mathrm{CAM}}|_i^2 \right| \Delta V_i \, , \qquad (7.2)$$

with N_{grid} and ΔV_i being the number of grid points and the volume element, respectively. $\phi_{\mathrm{NTO}}^{\mathrm{CIS}}$ and $\phi_{\mathrm{NTO}}^{\mathrm{CAM}}$ denote the dominant occupied NTOs from the CIS and TD-DFT/TDA calculations, respectively. Since the NTOs are normalized, the descriptor $\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$ is restricted to the interval $[0, 2]$. The lower limit indicates a complete coincidence of the densities, whereas the upper limit describes the opposite. The results are listed in Table 7.2 from above.

The descriptor $\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$ reflects well the overestimation of the excitation energies of the incremental approach associated with a poor CIS reference compared to the TD-DFT/TDA results. The systems with a good performance show all values below 0.4 (**C**, **D**, **H** and **M**), indicating comparable occupied NTOs from both methods. For
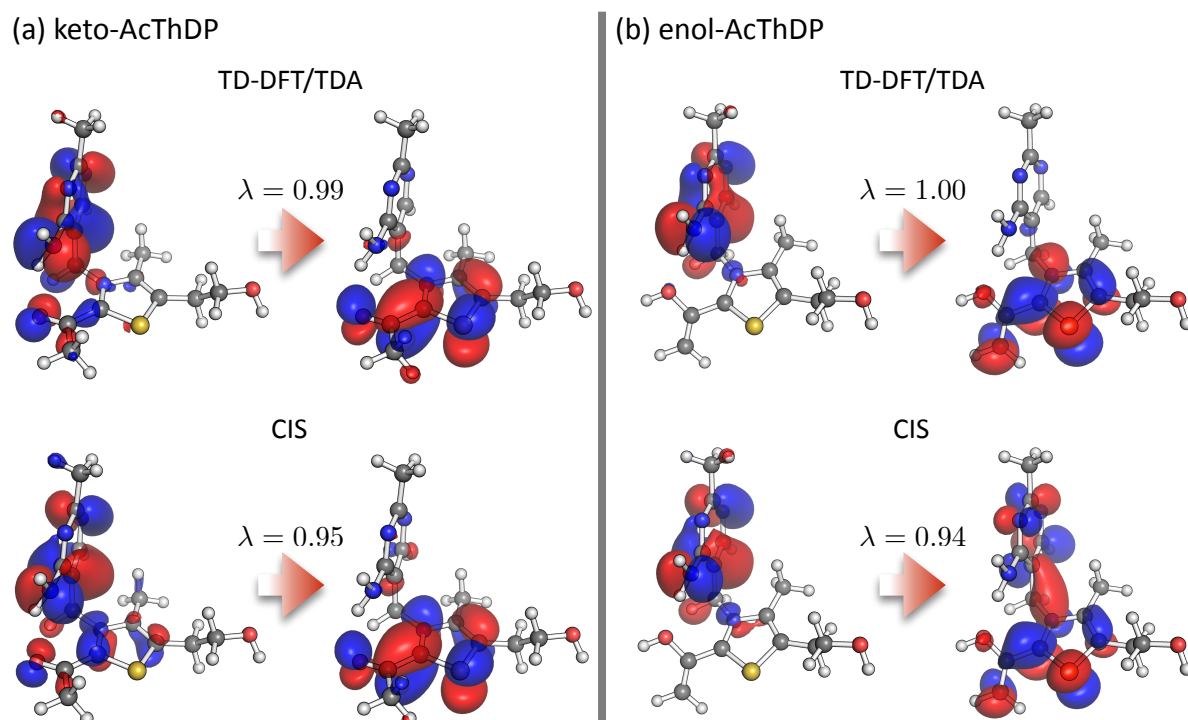
Figure 7.8.: Comparison of the dominant NTO pairs from the CIS/A'VDZ and TD-DFT/TDA (CAM-B3LYP/def2-TZVP) calculations for (a) keto-AcThDP (**B**) and (b) enol-AcThDP (**C**). Additionally, the singular values λ of the NTO pairs are given.

the overestimated charge-transfer excitations these values are generally larger, reflecting the delocalization of the hole densities towards the thiazolium ring. Furthermore, the description of these excitations require a larger number of NTO pairs of the CIS reference for some systems (see Table 7.2). In contrast to this, the TD-DFT/TDA transition density matrices for all studied excitations could be decomposed to single NTO pairs with weights above 0.95.

One notable exception for the analysis based on the descriptor $\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$ is observed. The charge-transfer excitation of IP-ThDP is significantly overestimated in the incremental EOM-CCSD calculation although the dominant NTOs between CIS and TD-DFT/TDA are comparable. This issue was investigated in further detail through analysis of the EOM-CCSD calculations. The results are shown in Figure 7.9.

The dominant occupied CIS NTO of IP-ThDP is well localized at the pyrimidine ring as it is also for the TD-DFT/TDA NTO. This is identified by the descriptor $\Delta n_{\mathrm{NTO}}^{\mathrm{h}}$ through a rather low value of 0.23. For the origin of the overestimated excitation energy the
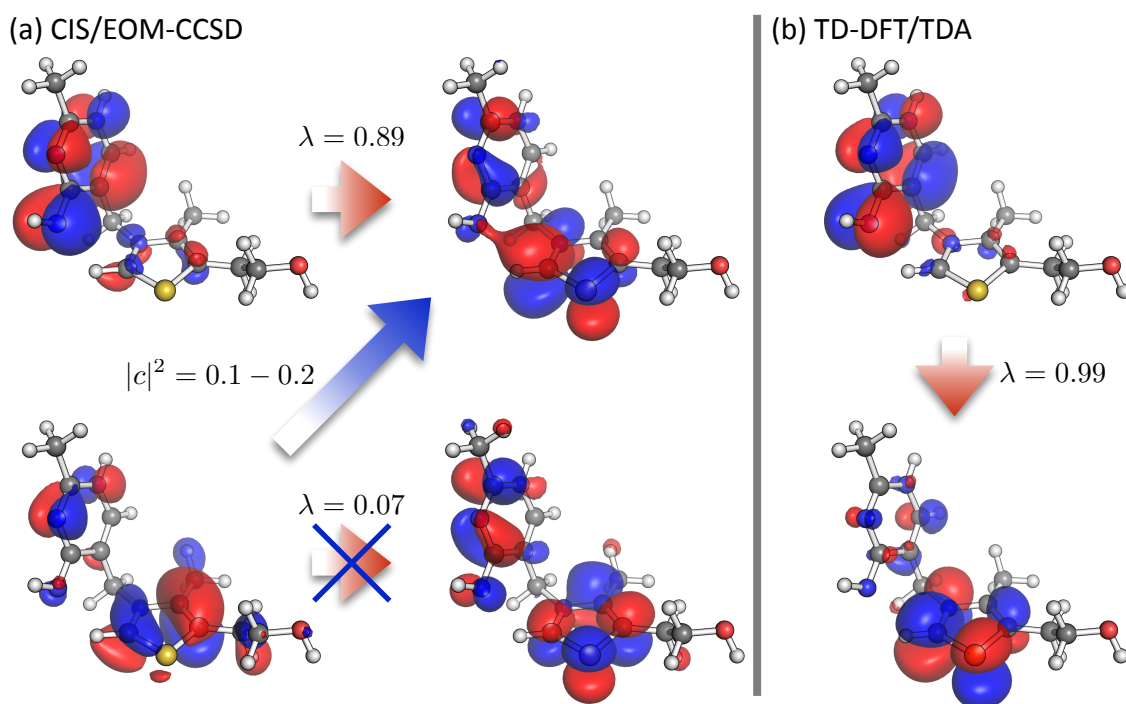
Figure 7.9.: Comparison of the dominant NTO pairs from the (a) CIS/A'VDZ and (b) TD-DFT/TDA (CAM-B3LYP/def2-TZVP) calculations for IP-ThDP. Additionally, the singular values λ of the NTO pairs are given. The blue colored arrow and cross indicate the contributions observed for the EOM-CCSD/A'VDZ calculations with the weight $|c|^2$ to the excited state.

second occupied CIS NTO combined with the EOM-CCSD results has to be considered. This NTO corresponds to a π-orbital of the thiazolium ring. The contributions with its virtual counterpart are found to be negligible in the increment calculations. However, the excitation from the second occupied NTO to the first virtual one determines a significant weight of 0.1–0.2 to the excited state. This contribution clearly raises the studied charge-transfer state. For the other systems showing overestimated excitation energies this contribution is naturally included due to the delocalization of the dominant occupied NTO towards the thiazolium ring.

A last analysis was carried out for the exclusion of groups from the incremental approach, which are far away from the excitation space. For the excited states of the capped histidine molecule it was already shown that the sum of the density overlaps between the LMOs of a group and the excitation space ($S_{\text{exc},G} = \sum\limits_{i \in G} S_{\text{exc},i}^{\text{h/p}}$) can be employed as criterion for their identification, whereby their exclusion was found to virtually have no effect on the calculated excitation energies (for details see Chapter 3). This was also tested for the studied excited states of ThDP and its derivatives, employing the same threshold of 0.01 for the criterion $S_{\text{exc},G}$. The energy penalties for the exclusion of such identified groups $\Delta \omega_{\text{excl}}^{\text{2B}}$ are listed in Table 7.2 from above. Furthermore, the number of excluded groups related to the total number of 2B increments $n_{\text{excl}}^{\text{2B}}/n_{\text{tot}}^{\text{2B}}$ are given.

The energy penalties for the exclusion of groups being far apart from the excitation space is found to be negligible. The largest one is 0.03 eV for enolate-AcThDP (**D**) which is by far lower than the intrinsic error of the method. For this excited state even 7 of 16 groups can be neglected from the treatment with virtually no loss of accuracy. This would make the approach much more efficient. For most of the other excited states the fraction of groups which can be excluded is smaller compared to the enolate-AcThDP intermediate (**D**). The reason for this is the charge-transfer character itself. Hole and particle densities are separated and extend together almost the entirety of the molecule. Therefore, the LMOs have a significant overlap to the excitation space and cannot be neglected in the many-body expansion. Another issue can be the delocalized CIS NTOs as discussed above. This results in a larger number of LMOs having a significant overlap to the hole or particle densities. However, the results clearly demonstrate that the employed criterion is reasonable for the identification of groups which have a minor impact on the excitation energies.

Summarizing the results for the application of the incremental EOM-CCSD method to

the excited states of ThDP and its derivates, the CIS reference was identified as a bottleneck of the incremental approach. In most cases, the charge-transfer excited states were described qualitatively wrong in the reference with the excitation hole being delocalized to the electron acceptor moiety. This leads to overestimated charge-transfer excitation energies. A solution to this problem could be the application of a TD-DFT/TDA calculation taken as reference. These are found to correctly describe the charge-transfer states. A subsequent incremental EOM-CCSD approach could then result in accurate excitation energies. This methodology was, however, not investigated in further detail.

7.4. Conclusions

The spectral signatures of proposed intermediates in ThDP-catalyzed reactions were studied, whereby the unsubstituted cofactors were taken as reference systems. Excitation energies were calculated at the TD-DFT/TDA level of theory (CAM-B3LYP/def2-TZVP) including a continuum representation of the enzyme environment. The excited state characters were then analyzed with the proposed method of Plasser and Lischka.[150] With this approach, structural properties of the ThDP intermediates could be related to their spectral behavior. Additionally, the incremental EOM-CCSD method, refined within this work, was applied to several excited states of ThDP and its derivatives. The main findings can be summarized as follows.

(1) The analysis of the low-lying excited states of the ThDP intermediates revealed three different classes of excitation characters. The first excitation class considers the charge-transfer excitations, where the electron is transferred from the pyrimidine to the thiazolium ring as observed for the unsubstituted cofactors. Excitation class II collects the charge-transfer excitations with the substituent being the electron donor whereas the thiazolium ring acts as electron acceptor. Lastly, local $\pi \to \pi^*$ transitions within the thiazolium-substituent moiety of the enamine intermediates build the third excitation class.

(2) Two distinct spectral behaviors are observed for excitation class I dependent on the type of the bonded substituent. The charge-transfer states are strongly red shifted with respect to the unsubstituted cofactors for coplanar substituents sharing a conjugated π-system with the thiazolium ring. These intermediates have to be considered in the interpretation of absorption spectra where low-energy bands

($<$3.5 eV) are observed. In contrast to this, substituents with a tetrahedral coordinated carbon atom bonded to the C2 atom of IP-ThDP almost retain the electronic structure of the thiazolium ring. As a consequence, the charge-transfer excitations are expected to be influenced to minor extent, so that the IP band in the spectral range of 4.00–4.13 eV (300–310 nm) should be observed.

(3) Charge-transfer excitations from the π-systems of the substituents to the thiazolium ring (excitation class II) are frequently observed for substituents with tetrahedrally coordinated carbon atoms bonded to the C2 atom of ThDP. For neutral substituents, these excitations can be related to the spectral signatures of unsubstituted ThDP by interpreting the π-system of the substituent as analogue to the one of the pyrimidine ring, while the thiazolium ring is retained as electron acceptor. According to the electronic structure of the substituent, the excited state is shifted with respected to unsubstituted ThDP, but the effect is not expected to be too large. However, there is the opportunity to introduce a negative charge in the electron donor unit through proton transfer to the enzyme environment, leading to a significant decrease of the charge-transfer excitation energies as observed for the enolate intermediates. Such states of the substituents have to be considered in the interpretation of spectra with low-energy bands ($<$3.5 eV).

(4) The excitation energies of the local $\pi \rightarrow \pi^*$ transitions (excitation class III) of the enamine intermediates depend, as expected, on the degree of conjugation of the thiazolium-substituent moiety. Extending the π-system leads to lower excitation energies, so that such intermediates could be observed in the low-energy regime of absorption spectra.

(5) The application of the incremental EOM-CCSD method to ThDP and its derivatives reveals an issue of the method for charge-transfer excitations. The CIS reference is hereby found to be the bottleneck of the approach. It is not capable to describe the excited states qualitatively correct. The occupied CIS NTOs are frequently found to be delocalized to the electron acceptor unit. This results in significantly overestimated excitation energies for the incremental approach.

The presented results should provide a guideline for the interpretation of measured absorption spectra of ThDP-dependent enzymes. Structural motifs of the intermediates are identified which are required for the observation of low-energy bands ($<$3.5 eV).

These are, for example, enolate tautomeric states of the substituent leading to low-lying charge-transfer states which were often not been given enough consideration in the spectral assignments.[24] The importance of such intermediates were, however, highlighted in the investigation of the phosphoketolase enzyme mechanism (Chapter 5), where enolate-AcThDP was found to be a key intermediate preventing off-pathway hydrolysis. Taking the presented results into account, these intermediates can be observed in the spectra if they have a significant role in the reaction mechanism. Furthermore, ThDP intermediates, with the carbon atom bonded to the C2 atom of ThDP being tetrahedrally coordinated, are not found to absorb in the low-energy regime. These intermediates were, however, frequently assigned to such absorption bands.[24;26] Here, other ThDP derivatives have to be considered in order to explain the spectral measurements, even intermediates from side reactions.

8
Summary

Electronic absorption spectroscopy (UV-vis or CD) is one of the major bioanalytical tools in use to identify intermediates in ThDP-catalyzed reactions, thereby rationalizing proposed mechanisms. Therefore, a special focus of this work was on the calculation of the spectral signatures of the cofactor and on-pathway intermediates. In a first step, the spectra of the pre-equilibrium states of ThDP were extensively studied, which are considered as reference systems since they could appear in all ThDP-dependent enzymes. These were experimentally assigned to a negative CD band at 3.65–3.87 eV (320–340 nm) and a positive one at 4.00–4.13 eV (300–310 nm) for the AP and IP tautomers, respectively.[23] Additionally, a hitherto unassigned negative band in the spectral range of 4.28–4.43 eV (280–290 nm) was observed in revisited CD measurements for *Zymomonas mobilis* pyruvate decarboxylase (*Zm*PDC) by the Tittmann group. For the AP band a charge-transfer character was deduced from chemical model measurements,[29] whereas a local $n \to \pi^*$ transition in the IP ring was assumed for its excitation character.[30] However, model compound calculations including a continuum representation of the environment revealed that these signals arise from charge-transfer excitations with the electron being transferred from the pyrimidine to the thiazolium ring for both tautomers. A local excitation for the IP-ThDP molecule could be excluded. Furthermore, it was shown that the *V* conformation of ThDP adapted in enzymes decreases the charge-transfer excitation energy compared to its *F* conformation observed for free ThDP in solution. This effect is strengthened through the different polarization of the enzyme environment and the aqueous solution, explaining why no absorption in the spectral region of 3.65–4.13 eV (300–340 nm) is observed in the latter.[130]

Having laid the ground work for the spectral signatures of ThDP, the tautomers were investigated in the active site of *Zm*PDC. A perturbative residue-by-residue analysis was applied to determine the impact of the first amino acid shell on the cofactor spectra. Hereby, the fundamental role of the protonation state of the canonical glutamate (Glu50) was highlighted. In the IP case, this residue was assumed to be deprotonated (negatively charged), thereby lowering significantly the charge-transfer state. In contrast to this, a protonated (neutral) Glu50 residue was natively assigned for AP-ThDP in order to build a hydrogen bond with its N1' atom. This assumption left the charge-transfer almost untouched. Since all experimental observations placed the AP band energetically below the one of IP-ThDP, a protonated canonical glutamate at the AP occupied enzyme pocket became questionable. MD-QM/MM calculations were then carried out for a functional dimer of *Zm*PDC in order to clarify this fact. The simulated absorption spectra revealed that the negative CD band at 3.65–3.87 eV (320–340 nm) is associated with AP-ThDP combined with a deprotonated Glu50 residue. Furthermore, the combination of an AP tautomer and a protonated canonical glutamate could even be linked to the hitherto unassigned band (4.28–4.43 eV; 280–290 nm). This assignment was additonally confirmed by CD measurements of the Glu50Gln variant of *Zm*PDC from the Tittmann group, where AP-ThDP should be the only accessible tautomer forming a hydrogen bond with its N1' atom and the carboxyamide group of the mutant residue. The unassigned band was the only observable signature in the spectra. Lastly, the assignment of the IP band at 4.00–4.13 eV (300–310 nm) could be confirmed, but also the calculations in the enzyme environment addressed a charge-transfer character.

The first reaction intermediate studied in this work was 2-acetyl-ThDP (AcThDP), detected in the reaction of the *B. breve* phosphoketolase (PK) enzyme with fructose-6-phosphate through X-ray structural analysis in the Tittmann group.[40] An absorption band at 2.95 eV (420 nm) combined with a kinetic overshooting were observered and linked to this intermediate. However, the experimental studies had not allowed to clearly determine the tautomeric state of the acetyl moiety. Model compound calculations including a continuum representation of the enzyme environment revealed that both keto and enolate-AcThDP absorb in the spectral region of interest. Enol-AcThDP showed a significantly larger excitation energy, excluding this tautomeric state from the observation in the spectra. The excitation characters for the absorbing intermediates are both of charge-transfer nature, but differ in their origin. The first excited state of keto-AcThDP reports a similar character as observed for the unsubstituted cofactors. The

extension of the π-system towards the substituent in the electron acceptor moiety leads to a significant red shift of about 1 eV compared to the ThDP bands. On the other hand, enolate-AcThDP shows a charge-transfer state with the electron being transferred from the electron-rich π-system of the enolate moiety to the thiazolium ring. Since the absorption intensity was computed to be larger for enolate than for the keto tautomer, the spectral observations could be explained in the following way: The enolate-AcThDP intermediate is initially formed in the course of the reaction, resulting in the absorption band at 2.95 eV (420 nm). Afterwards, this intermediate slowly tautomerizes to its keto form, absorbing in the same spectral region but less intense. These conclusions could be confirmed through reaction pathway calculations, which were subject of the Master's project of S. Wille.[41] Thereby, the enolate-AcThDP intermediate is found to be built through an almost barrier-free proton transfer from the enol tautomer to the His553 residue of *B. breve* PK. The latter then serves as proton donor for the tautomerization to keto-AcThDP in the absence of the co-substrate phosphate. This conversion is kinetically decelerated with a calculated free activation energy of 77.1 kJ $\cdot$ mol^{-1}, requiring a rather large movement of the histidine residue. In contrast to this, the inclusion of the co-substrate phosphate, acting hereby itself as the proton donor for the tautomerization, decreases the free activation to 45.0 kJ $\cdot$ mol^{-1} for an efficient enzyme catalysis. All these findings identified enolate-AcThDP as a key intermdiate in the enzymatic pathway, preventing the system from off-pathway hydrolysis in the absence of the co-substrate.

The preceeding calculations of the spectra of the AcThDP tautomers showed the coincidence of the absorbances of keto and enolate-AcThDP. As an attempt to split their common band, the impact of the first shell of amino acids was determined by a perturbative residue-by-residue analysis in the active site of *B. breve* PK. This approach allowed the identification of residues, varying in their contributions to the different charge-transfer states of both tautomers. Based on this information two mutant candidates, the variants Glu156Gln and Asn549Asp of PK, could be proposed with excitation energy shifts which would enable the observation of two separate tautomer bands. These mutants were synthesized in the Tittmann group, followed by time-resolved UV-vis measurements. Unfortunately, the calculated effects could not be confirmed, either leaving the spectral properties unchanged (Glu156Gln) or leading to the complete disappearance of the band (Asn549Asp). The reason for this could be invalid assigned protonation states of the mutant candidates or effects on the reaction profile arising from them. However, the methodology of the perturbative approach is still promising for calculations of

excitation energy shifts in enzymes. This approach could be applied for the efficient screening of a variety of mutant candidates, thereby identifying the most promising ones which can then be subject of higher level methods.

The next problem considered the spectral characteristics of the 1,2-dihydroxyethyl-ThDP (DHEThDP) system, being a key intermediate in the enzymatic pathway of transketolase (TK). UV-vis measurements from the Tittmann group revealed an uncommon broad absorption band in the spectral range of 2.25–3.54 eV (350–550 nm) linked to this intermediate in *E. coli* TK.[40] Inspection of the X-ray structure showed the possibility for the proton transfer from the substituent to the nearby His473 residue, so that both an enol or enolate state of DHEThDP would be conceivable. Additionally, the pyrimidine ring was observed in its APH$^+$ state. These systems were studied in the active site of the enzyme, employing a QM/MM cluster model. Calculated vertical excitation energies have shown that the low-lying excited states of enol and enolate-DHEThDP are placed at the high and low-energy limit of the measured band, respectively. The excited states have charge-transfer character with the thiazolium-substituent moiety acting as electron donor, whereas the APH$^+$ ring is the electron acceptor. The introduction of a negative charge within the electron donor group for enolate-DHEThDP causes the strong red shift of about 1.1 eV with respect to the enol species. As an attempt to connect these species, the potential of the hydrogen movement was calculated in a constrained environment. The analysis revealed an anharmonic stretch vibration associated with the enol state of DHEThDP, thereby missing the observation of the enolate species. However, the simulated spectra according to this simple reaction path clearly showed that an enolate species is required for the explanation of the measured band. Otherwise, a much sharper absorption band at about 3.26 eV (380 nm) has to be expected.

The investigations of the pre-equilibrium states of ThDP and the key intermediates in the PK and TK enzymes revealed new information of their spectral signatures. This amounted to the revision of assigned intermediates of ThDP-catalyzed enzyme reactions in absorption spectra. Based on model compound calculations including a continuum representation of the enzyme environment, the low-lying excited states were analyzed according to their excitation characters. Thereby, structural motifs of the intermediates were related to expected excitation energy ranges. This approach led to a classification of the excited states into three groups. Firstly, equivalent charge-transfer excitations were collected compared to AP and IP-ThDP, where the electron is transferred from the pyrimidine to the thiazolium ring. Depending on the structural motifs of the bonded

substituent, the excitation energies are accordingly modulated. Coplanar substituents towards the thiazolium ring, thereby extending the π-system of the electron acceptor moiety, lead generally to large red shifts of the absorption band with respect to the unsubstituted cofactors. In contrast to this, a tetrahedrally coordinated carbon atom bonded to the C2 atom of IP-ThDP retains almost the electronic structure of the thiazolium ring. Therefore, the IP band in the spectral region of 4.00–4.13 eV (300–310 nm) should be observed. Low-energy bands (<3.5 eV; 354 nm) are not expected for the latter excited states. The second excitation class has considered the charge-transfer excited states with the π-system of the substituent acting as electron donor and the thiazolium ring being the electron acceptor. These excitations can be related to the spectral signatures of unsubstituted ThDP by interpreting the π-system of the substituent as analogue to the one of the pyrimidine ring. As a consequence, the ThDP bands are shifted according to the electronic structure of the substituent. However, the effect was observed to be rather small for neutral substituents, resulting in excitation energies all above 3.5 eV (354 nm). A significant red shift was only observed if a negative charge is introduced to the substituent, leading to charge-transfer states below 3.5 eV (354 nm). The last excitation class collects local $\pi \rightarrow \pi^*$ transitions for the enamine-ThDP intermediates. It was shown that the band maximum correlates with the degree of conjugation of the thiazolium-substituent moiety. Extending the π-system leads to lower excitation energies, so that such intermediates could be observed in the low-energy regime of absorption spectra. Considering the presented results, a guideline for the interpretation of absorption spectra was established to ThDP-catalyzed enzyme reactions. Structural motifs combined with the excited state characters can be employed to distinguish spectral regions for proposed reaction intermediates.

X-ray structural analysis combined with kinetic measurements of the human transketolase (TK) enzyme from the Tittmann group revealed a fundamental functionality of ThDP-dependent enzymes.[40] Positive cooperativity was observed for substrate binding. This effect was associated with a hydrogen bonded network of glutamic acid residues connecting the two active sites. This was proposed to act as a proton wire for the communication. Additionally, the network contains a glutamine bridge with an associated water molecule. The latter was observed in two conformations so that it could act as proton relay for its transfer from one subunit to the other. MD simulations were carried out to investigate the stability of the hydrogen bond network and the flexibility of the

glutamine bridge.

The MD simulations have shown that the hydrogen bonded network between the glutamate residues is stable during the whole simulation time at ambient conditions. These results fulfills the requirement for an efficient proton transfer. Furthermore, both conformations of the glutamine bridge with the associated water molecule were confirmed in the simulations. A third conformation was also found which deactivates the bridge. This movement was, however, recovered in the progression of the system. The confomational change of the glutamine residues were shown to correlate with the oscillation of the shuttling water through the symmetry axis. Only one of the partially occupied sites observed in the crystal structure was found at the same time. Moreover, the analysis of the water oscillation set the time scale to the nanosecond regime. This is much faster than the catalyzed reaction time, required for an efficient communication. Altogether, the results confirmed the proposal for a proton wire as communication channel between the two active sites of the human TK enzyme, whereby the shuttling water in the glutamine bridge could act as a proton carrier.

The computation of excitation spectra can be a powerful tool in the field of analytical chemistry. This was demonstrated within this work for the spectral signatures of ThDP and its derivatives, leading to the assignments and identifications of enzymatic reaction intermediates. In order to provide highly accurate excitation energies, robust quantum chemical methods have to be applied. The EOM-CCSD method represents such an approach. However, the steep scaling of the computation time with the system size allows only the treatment of small to medium sized molecules. Incremental approaches can be employed to reduce the effort, thereby increasing the applicability of this high-level method. For this reason, the incremenatal EOM-CCSD method, originally introduced by Mata and Stoll,[38] was refined within this work.

In the original incremental EOM-CCSD method orbital charge centroids were employed to construct the groups. Here, it was shown that this criterion suffers from a meaningful description of the delocalized NTOs. A new criterion was proposed in order to identify the important LMOs for the excitation. The density overlap between the CIS excitation space and the LMOs was found to fulfill this requirement. The LMO-LMO exchange interaction was then employed to construct spatially close groups for the two-body (2B) calculations. Based on these criteria, four group construction algorithms were benchmarked to 44 excitations of 16 organic chromophores. The best performance

was obtained for the 2B approach, allowing a larger number of orbitals in the main fragment. This leads to a better description of the excited states in the increment calculations. The error of -0.01 ± 0.12 eV as well as the mean absolute deviation of 0.07 eV with respect to the full EOM-CCSD results are only a fraction of the intrinsic accuracy of the conventional method. The benchmark calculations also helped to identify some issues in the procedure. The fragment calculations break the symmetry of the excited states which can lead to an inbalanced description. Furthermore, excitations with large doubles contributions are poorly decribed, because of the CIS reference state. However, the incremental EOM-CCSD method is intended for application to single electron excitations for large, in most cases unsymmetric molecules. These states were excellently described by the implemented version.

In the benchmark calculations another strategy was also tested in order to identify LMOs having a minor impact on the excited states. Their exclusions from the incremental approach would then lead to savings in the computation time. In the first step, the attempt was made to independently find such orbitals prior to the group construction. For this purpose, the density overlap between the excitation space and the LMOs was considered as criterion. It was shown that this approach leads to a systematic overestimation of the excitation energies, most likely arising from the lack of correlation effects. The independently excluded LMOs may have an indirect effect to spatially close orbitals which have to be included to describe the excitation properly. This problem could be circumvented employing a modified exclusion procedure. Hereby, the groups were initially built. Afterwards, whole groups were rejected from the incremental approach if all LMOs in a group showed a negligible density overlap to the excitation space. Application of this criterion to the excitations of the capped histidine molecule demonstrated a successful performance. Several groups were excluded with virtually no loss of accuracy.

The success of the exclusion criterion for whole groups could be confirmed in the application of the incremental EOM-CCSD method to excited states of ThDP and its derivatives. However, these calculations revealed another problem in the method. For charge-transfer excitations the CIS reference was found to be a strong limiting aspect. The occupied NTOs were frequently observed to be delocalized to the electron acceptor unit. This resulted in significantly overestimated charge-transfer excitation energies. A possible solution would be to employ TD-DFT/TDA reference states. These were shown to qualitatively describe such excited states correct. A subsequent incremental EOM-CCSD approach could then result in highly accurate excitation energies.

Besides the exclusion criterion, the development of the incremental EOM-CCSD method has been focused so far on the robustness of the approach, recovering the conventional calculations. Therein, computation time is naturally saved for the individual increment calculations due to a smaller occupied orbital space. However, the full efficiency will be only achieved through the restriction of the virtual space. The local EOM-CCSD approach by Korona and Werner already provides the methodology.[94] Another promising approach to restrict the virtual space are the recently proposed PNO based methods of Hättig and coworkers.[99–101] The incremental EOM-CCSD method should be coupled to either of these strategies. The resulting method is then expected to give accurate excitation energies at a favourable computational cost.

A
List of Abbreviations

AcThDP	2-acetyl-thiamin diphosphate
AP	4'-aminopyrimidine
APH$^+$	4'-aminopyrimidinium
Asn	asparagine
Asp	aspartic acid, aspartate
B. breve	*Bifidobacterium breve*
CC	coupled cluster
CD	circular dichroism
CIS	configuration interaction with singles
CMO	canonical molecular orbitals
COSMO	conductor-like screening model
DFT	density functional theory
DHEThDP	1,2-dihydroxyethyl-thiamin diphosphate
E. coli	*Escherichia coli*
EOM-CCSD	equation-of-motion coupled cluster singles and doubles
FWHM	full width at half maximum
GAFF	generalized Amber force field
Gln	glutamine
Glu	glutamic acid, glutamate
HF	Hartree-Fock
His	histidine
IP	1',4'-iminopyrimidine

LMO	localized molecular orbitals
LR-CC2	linear response second-order approximate coupled cluster
MD	molecular dynamics
MM	molecular mechanics
NPT	isothermal-isobaric ensemble
NTO	natural transition orbitals
NVT	canonical ensemble
PDB	protein data bank
PDC	pyruvate decarboxylase
PK	phosphoketolase
PNO	pair natural orbital
PT	perturbation theory
QM	quantum mechanics
RMSD	root-mean-square deviation
TD-DFT	time-dependent density functional theory
TDA	Tamm-Dancoff approximation
ThDP	thiamin diphosphate
TK	transketolase
UV-vis	ultraviolet-visible
Zm	*Zymomonas mobilis*

B

Structural Data

B.1. Incremental EOM-CCSD Benchmark Structures

The B3LYP/6-31G* optimized geometries of the benchmark systems for the incremental EOM-CCSD method (Chapter 3) are provided in the xyz-format (Å).[61;62;85] The retained names of the molecules combined with the number of atoms and the total system charge are placed on top of each coordinate set. All systems were computed with spin multiplicty $2S + 1 = 1$.

Acetone 12 0

```
O   12.1376716969    1.1977520390    5.2357660934
C   12.1066751956    1.8852973090    4.2334949149
C   11.0473485967    1.6866243552    3.1595655369
C   13.1260091712    2.9879654187    3.9883435287
H   13.8326529772    3.0283687539    4.8200510189
H   10.3792149437    0.8715832014    3.4458436938
H   13.6695692596    2.8091899285    3.0515807866
H   12.6244962816    3.9588567751    3.8844085621
H   10.4652066353    2.6063070937    3.0167085844
H   11.5171322422    1.4551851255    2.1948282803
```

Acrolein 8 0

```
O   -3.1422524053   -11.6494037374   -3.7910808596
C   -4.2642493112   -11.7496336279   -4.2479752740
C   -4.5618157016   -12.0889488518   -5.6531133163
```

```
C    -5.8235106177   -12.1822744235    -6.0904099543
H    -5.1632383066   -11.5854022734    -3.6103619961
H    -3.7057247999   -12.2563382123    -6.3032054466
H    -6.6650275969   -12.0105431844    -5.4208646437
H    -6.0652972608   -12.4305056893    -7.1199025094
```

Adenine 15 0

```
C    -0.0653266441    -0.0560922608    -0.0201824681
N     0.0610847008     0.0102086977     1.3535324667
C     1.4125229543     0.0124692768     1.6276439292
C     2.0230299651    -0.0514691721     0.3693639084
N     1.0841215654    -0.0953940053    -0.6490475014
N     2.0246482916     0.0601062294     2.8175154873
C     3.3540511410     0.0346241881     2.6766448317
N     4.0825805174    -0.0277854139     1.5490384254
C     3.4341787792    -0.0666961143     0.3715195663
N     4.1607649231    -0.0899704999    -0.7768575825
H     3.9301194221     0.0656588518     3.5996848319
H    -0.6802365222     0.0509148854     2.0385506423
H     5.1488559325    -0.2822792146    -0.6982535605
H     3.6913764496    -0.3145008629    -1.6417834690
H    -1.0372794757    -0.0728235854    -0.4967775077
```

Anthracene 24 0

```
C     0.0964618702     0.0000000000    -0.0228946995
C     0.0964616169     0.0000000000     1.4228948137
C     1.3526071585     0.0000000000     2.1076652636
C     2.5343420867     0.0000000000     1.4134467407
C     2.5343425622     0.0000000000    -0.0134462632
C     1.3526078237     0.0000000000    -0.7076655182
H     1.3509005087     0.0000000000     3.1955397361
H     3.4811539161     0.0000000000     1.9473652692
H     3.4811546783     0.0000000000    -0.5473641786
H     1.3509016103     0.0000000000    -1.7955402370
C    -1.1279813949     0.0000000000    -0.7038915366
C    -1.1279814263     0.0000000000     2.1038908012
C    -2.3525031018     0.0000000000    -0.0228790907
H    -1.1280145115     0.0000000000     3.1926335364
```

```
H   -1.1280144356    0.0000000000   -1.7926345158
C   -2.3525027980    0.0000000000    1.4228790883
C   -3.6085821453    0.0000000000   -0.7076356660
C   -3.6085814115    0.0000000000    2.1076352603
C   -4.7903336167    0.0000000000    1.4134252202
C   -4.7903340229    0.0000000000   -0.0134244658
H   -3.6068941290    0.0000000000   -1.7955184352
H   -5.7371362435    0.0000000000   -0.5473604241
H   -5.7371356611    0.0000000000    1.9473615281
H   -3.6068929334    0.0000000000    3.1955177729
```

Benzene 12 0

```
C    1.3916961401    0.0000000000    0.0000000000
C    0.6958480700   -1.2052442117    0.0000000000
C   -0.6958480700   -1.2052442117    0.0000000000
C   -1.3916961401    0.0000000000    0.0000000000
C   -0.6958480700    1.2052442117    0.0000000000
C    0.6958480700    1.2052442117    0.0000000000
H    2.4736443205    0.0000000000    0.0000000000
H    1.2368221603   -2.1422388215    0.0000000000
H   -1.2368221603   -2.1422388215    0.0000000000
H   -2.4736443205    0.0000000000    0.0000000000
H   -1.2368221603    2.1422388215    0.0000000000
H    1.2368221603    2.1422388215    0.0000000000
```

p-Benzoquinone 12 0

```
C   -0.0398976954    0.0000000000   -0.0230526041
C   -0.0041662753    0.0000000000    1.4635509800
C    1.1595171875    0.0000000000    2.1354233754
C    2.4647584620    0.0000000000    1.4230473353
C    2.4290358985    0.0000000000   -0.0635528258
C    1.2653514152    0.0000000000   -0.7354233898
H   -0.9704355437    0.0000000000    1.9609330316
H    1.2119733771    0.0000000000    3.2209390533
H    3.3953134717    0.0000000000   -0.5609375362
H    1.2128956435    0.0000000000   -1.8209306046
O    3.5258459348    0.0000000000    2.0357715150
O   -1.1009638759    0.0000000000   -0.6357683301
```

Cytosine 13 0

N	0.0243550485	-0.0237774549	0.0161398652
C	0.0193664778	-0.0697786100	1.3346301954
C	1.2080515393	-0.0798678629	2.1502904694
C	2.3909616506	-0.0306939799	1.4813033221
N	2.4000025135	0.0239733509	0.1264007903
C	1.2124865293	0.0330126853	-0.6712537631
N	-1.2013174721	-0.0751517723	1.9483650258
O	1.3326570610	0.0828053665	-1.8843794077
H	1.1668803380	-0.1101306342	3.2322141231
H	3.3559540590	-0.0306743969	1.9791582311
H	3.2655692830	0.0613145667	-0.3964538194
H	-1.9992933697	-0.2239590970	1.3450892737
H	-1.2828436582	-0.3801011613	2.9065586940

Furan 9 0

O	0.0003360480	0.0007900031	0.0000000000
C	1.3644566251	-0.0027312336	0.0000000000
C	1.8450677189	1.2709547054	0.0000000000
C	0.6940975108	2.1303535543	0.0000000000
C	-0.3906229584	1.3077041973	0.0000000000
H	1.8289657230	-0.9773617735	0.0000000000
H	-1.4570441683	1.4764668831	0.0000000000
H	2.8849710847	1.5676864343	0.0000000000
H	0.6831764160	3.2117132297	0.0000000000

Guanine 16 0

C	-0.5494222392	-0.0519198008	0.7250081315
C	0.8202693495	-0.2055684213	0.9456607231
N	-1.2145098904	0.1585855127	-0.4411353054
C	-0.4111685407	0.1888913923	-1.4783257017
N	1.0883581229	-0.3908866483	2.2886788155
H	-0.2684004706	-0.4590075262	3.9268989562
C	-0.0843110074	-0.3490813710	2.8661710130

```
N   -1.1190345607    -0.1447184627     1.9686469114
H   -2.1071494414    -0.0764936982     2.1677339324
C    1.7129336517    -0.1685912536    -0.1855654776
O    2.9225042333    -0.2930086622    -0.2625385058
N    0.9502764005     0.0466888444    -1.3886873579
H    1.5122014519     0.0062768503    -2.2319494408
N   -0.9415094640     0.3196842199    -2.7454480137
H   -1.9184391207     0.5860834301    -2.7310964001
H   -0.3925983707     0.8330655946    -3.4240522801
```

All-*E*-Hexatriene 14 0

```
C    0.2329743013    -0.1552556523     0.0288491394
C   -0.0171265063     0.0276065607     1.3458408043
C    1.5299837152     0.0391980498    -0.5912231576
C    1.7834466331    -0.1418226340    -1.8983652981
C   -1.3135554420    -0.1674775730     1.9668037698
C   -1.5671464606     0.0164668639     3.2735219730
H    2.7736045860     0.0160458604    -2.3154663337
H    1.0052963609    -0.4549455261    -2.5914697726
H    2.3396293752     0.3534688878     0.0688725895
H   -0.5812029122    -0.4709683333    -0.6256391723
H    0.7975019183     0.3430628376     1.9999078668
H   -2.1231872951    -0.4835296786     1.3077204690
H   -2.5570902536    -0.1415920110     3.6909211936
H   -0.7889510202     0.3305773480     3.9661739289
```

Capped Histidine 29 0

```
C   -1.8894496752     0.6539802553     3.6459549493
C   -3.1337681038    -0.2702826411     3.7317940574
O   -4.2498739096     0.0975893361     3.3682698403
N   -1.0423199955     0.2509699330     2.5160434096
C   -0.1377679954    -0.7624999844     2.5940077328
O   -0.0072394641    -1.4554132299     3.6066459601
C    0.7192348228    -0.9919948468     1.3598252205
N   -2.8565961127    -1.5041453296     4.2275840125
C   -3.8519101715    -2.5600756543     4.2813813147
H   -1.2830437284     0.5234625198     4.5478824825
H   -1.1673391669     0.7385613927     1.6396466151
```

```
H    1.7714751463   -0.8747450833    1.6385617802
H    0.5845039745   -2.0238716423    1.0204459011
H    0.4934203431   -0.3107023472    0.5334019238
H   -3.7173562255   -3.2889876233    3.4706960268
H   -3.7936278685   -3.0887277066    5.2390923889
H   -4.8366704942   -2.1011821478    4.1789966299
H   -1.8717561432   -1.7356183421    4.3468678958
H   -1.3994654354    2.7073551074    3.2463800908
H   -3.0099194670    2.2066566607    2.6697950626
H   -5.0577484802    2.1654649373    4.8054897284
H   -4.9754460159    3.4752493522    7.0092408358
H   -2.6246694995    4.3417676806    7.4971894020
N   -4.1776194764    3.3264916071    6.4084119680
C   -2.9019484201    3.7692326735    6.6216986095
C   -2.2954568658    2.1288858200    3.4970225011
N   -2.0970106236    3.4152659460    5.6440749485
C   -2.8841977738    2.7106861782    4.7501089008
C   -4.1792781742    2.6438961790    5.2091188113
```

Naphthalene 18 0

```
C    0.0375381952    0.0000000000   -0.0171788141
C    0.0375382643    0.0000000000    1.4171788187
C    1.2828533008    0.0000000000    2.1030763210
C    2.4722864597    0.0000000000    1.4088070730
C    2.4722866671    0.0000000000   -0.0088071120
C    1.2828535370    0.0000000000   -0.7030766129
H    1.2810248623    0.0000000000    3.1910815347
H    3.4172201977    0.0000000000    1.9461674405
H    3.4172206892    0.0000000000   -0.5461669256
H    1.2810253728    0.0000000000   -1.7910818230
C   -1.2077962570    0.0000000000   -0.7030346895
C   -1.2077957756    0.0000000000    2.1030341770
C   -2.3972514129    0.0000000000   -0.0087314110
C   -2.3972513889    0.0000000000    1.4087321087
H   -1.2060177402    0.0000000000    3.1910413446
H   -3.3421085182    0.0000000000    1.9462505009
H   -3.3421084208    0.0000000000   -0.5462500128
H   -1.2060180326    0.0000000000   -1.7910419181
```

Pyridine 11 0

```
C     0.1335936311     0.0001851871    -0.0612283407
C     0.0740858457     0.0002125948     1.3343330918
C     1.2718994595     0.0001495238     2.0496701442
C     2.4727016337     0.0000634770     1.3393679166
C     2.4190608658     0.0000445191    -0.0564184412
N     1.2778058864     0.0001078286    -0.7573327337
H    -0.7828160902     0.0002285647    -0.6503745617
H    -0.8862605784     0.0002819155     1.8424394964
H     1.2696479933     0.0001661668     3.1368415903
H     3.4309077130     0.0000129479     1.8515191508
H     3.3379766402    -0.0000257254    -0.6416643129
```

Pyrrole 10 0

```
N    -0.0594108394     0.0000000000    -0.0273041531
C    -0.0053822837     0.0000000000     1.3477070472
C     1.3237467072     0.0000000000     1.7144074994
C     2.0972167469     0.0000000000     0.5164336361
C     1.2159652688     0.0000000000    -0.5439793869
H     1.3856289389     0.0000000000    -1.6113784185
H     3.1766324304     0.0000000000     0.4417668526
H     1.6997140951     0.0000000000     2.7290134773
H    -0.9082735458     0.0000000000     1.9417037100
H    -0.9063495183     0.0000000000    -0.5741852640
```

Thymine 15 0

```
C    -0.0112159298     0.0000000000    -0.0361560271
C     0.0231972495     0.0000000000     1.4323173641
C     1.2300214356     0.0000000000     2.0433433309
N     2.4203816559     0.0000000000     1.3439243429
C     2.5044413991     0.0000000000    -0.0444948550
N     1.2567857793     0.0000000000    -0.6499359690
C    -1.2883884937     0.0000000000     2.1636251788
H     1.3339161879     0.0000000000     3.1241334191
H     3.3085511148     0.0000000000     1.8261305375
H     1.2611656880     0.0000000000    -1.6641082596
H    -1.1374763678     0.0000000000     3.2479586981
H    -1.8856034580     0.8786347013     1.8932634006
```

```
H   -1.8856034580   -0.8786347013    1.8932634006
O   -1.0271712968    0.0000000000   -0.7152886941
O    3.5618424941    0.0000000000   -0.6489748678
```

Uracil 12 0

```
H    1.3841070792    0.0000000000   -2.0907615463
C    0.8012015821    0.0000000000   -1.1796619407
C    1.3819216752    0.0000000000    0.0400153134
H    2.4582730059    0.0000000000    0.1782454967
C   -0.6548595422    0.0000000000   -1.2937411802
O   -1.2983319121    0.0000000000   -2.3300122828
N   -1.3148303266    0.0000000000   -0.0428159469
H   -2.3285783542    0.0000000000   -0.0735485569
C   -0.7559799040    0.0000000000    1.2248692443
O   -1.3955012805    0.0000000000    2.2604559546
N    0.6400129551    0.0000000000    1.1991051284
H    1.0825649700    0.0000000000    2.1078503694
```

B.2. Structures of ThDP-Intermediates

The B3LYP-D3(BJ)/def2-SVP optimized geometries of the reaction intermediates of
ThDP-dependent enzymes (Chapter 7) are provided in the xyz-format (Å). [61;62;111–113]
The molecule set includes the unsubstituted cofactors (AP and IP-ThDP) as well as the
16 proposed intermediates, employed for the analysis of their spectral signatures. The
notation of these intermediates follows the labeling in Figure 7.4 and are placed on top
of each coordinate set together with the number of atoms and the total system charge.
All systems were computed with spin multiplicity $2S + 1 = 1$.

AP-ThDP 35 +1

```
C   74.81581371497340   65.28641523088885    0.37487896683925
C   75.37461148128450   64.83829378032097   -0.85049130340504
C   75.75903185290026   65.84967934748477   -1.73377077274633
N   75.66921083038345   67.14512936784469   -1.45573697447464
C   75.18257469586204   67.47015403552167   -0.24683188774510
N   74.73958098695259   66.58557309258730    0.65694728817087
```

```
N    74.40080696033100    64.41308161390498     1.36337174405066
C    75.13922617259504    68.91978313381998     0.12518850161967
H    73.92840912420863    63.56258927272683     1.07129108304045
H    73.92787808058586    64.88744583605035     2.13041679238792
H    76.16985438154346    65.58477822111323    -2.71844536079559
H    75.39159166150915    69.54339122838950    -0.74067302498645
H    75.86519418963000    69.11569737604214     0.93158524279258
H    74.14756756368001    69.18706675460894     0.52042410484939
C    75.66477523774620    63.39643239283060    -1.15507323328715
H    75.64127447938030    63.20587891180855    -2.23354749809049
H    74.95775653040036    62.69659527402634    -0.68909847553549
N    77.03339645027204    63.01929689882697    -0.67336940525323
C    77.35057790945477    63.12787034943060     0.61382655402589
S    78.97872151535393    62.79093069864729     0.90573431963954
C    79.26050668923007    62.48384065520763    -0.79153330662780
C    78.09317508189773    62.64890864811269    -1.50168907097166
C    80.62340001579963    62.13268110076045    -1.31312005244387
C    81.70453270167505    61.98721302714716    -0.22737764027205
O    82.97544406826286    61.79704458023119    -0.77306184909739
C    77.92030182854866    62.47041750043314    -2.97506100445878
H    77.12738567063953    61.74280264521242    -3.20785677300054
H    78.85097459128879    62.09617786139108    -3.41835715603002
H    77.67242466198516    63.42005675788411    -3.47612386889645
H    80.95243142490011    62.91478161527083    -2.01878262077993
H    80.55489848727898    61.19601764651887    -1.89436999881107
H    81.41682873790461    61.18044869301921     0.47767518243315
H    81.75306597965584    62.92277936533359     0.35739945905035
H    83.06707017570925    60.88496424908062    -1.07924335035482
H    76.61973085912378    63.43158290042253     1.36880991714069
```

IP-ThDP 35 +1

```
C    75.10294676220362    65.17710495693154     0.35633831694944
C    75.52165990522461    64.84570793418170    -1.01726015072871
C    75.75801901084979    65.86208704762511    -1.88454711229754
N    75.62942259503181    67.16220533232325    -1.47159862355108
C    75.26097647223307    67.45235036353763    -0.16842404371173
N    75.00395237793690    66.52964934077328     0.69456470751454
N    74.84086807490664    64.23286529712757     1.19327456681827
C    75.19969604060657    68.90532475156698     0.19337978902986
H    75.79809949963179    67.92081143384505    -2.12340768665009
```

```
H    74.53905409031805    64.63855128335739     2.08433406684275
H    76.04140917344024    65.71203481422384    -2.92850825335508
H    74.58129492214347    69.47244213665887    -0.52064322279590
H    76.21143313865383    69.34537381714571     0.18572248157164
H    74.77664203419498    69.00693289081057     1.19913176470909
C    75.59974133501760    63.40054861375838    -1.41017139230761
H    75.63362279708970    63.26622283319759    -2.49726423288603
H    74.72453962042424    62.86793933523052    -1.01569507395015
N    76.79685200722398    62.72107058292912    -0.83690231596757
C    76.79059743169923    62.25262883794508     0.40602932597783
S    78.29176096533605    61.61464886917013     0.85327126262071
C    78.96981755923088    61.99599036605240    -0.71819628782679
C    78.01781250741594    62.59028227738582    -1.50765526506695
C    80.40277285674424    61.70112874604916    -1.03104383516743
C    81.37712639845428    62.04451308913477     0.11460599557782
O    82.70678898378239    61.99068990264245    -0.31202978129867
C    78.17046258721881    63.04232920607668    -2.92274394043176
H    77.47318446012900    62.51434596443770    -3.59170795336667
H    79.18576726735232    62.82952918636344    -3.27690318434420
H    77.99219827858192    64.12414877574300    -3.02560312573646
H    80.70713521501307    62.29244717705718    -1.90674589516894
H    80.52077029364280    60.63853219906071    -1.31165607887948
H    81.18369383607053    61.38405517190242     0.98422501046371
H    81.18422084172710    63.07650221069112     0.45335173044218
H    83.00690990873210    61.07221306569495    -0.33145522868180
H    75.91509775173834    62.41668318936863     1.04772166565279
```

ThDP-Intermediate A 41 +1

```
C    75.18795680106966    65.04403529257229     0.24921273191077
C    75.38603715806455    64.65352731548784    -1.10504651863405
C    75.60047116051557    65.70686257010087    -1.99626793210261
N    75.71854002691309    66.97781867611828    -1.62761883514922
C    75.63432975510575    67.23157383858424    -0.30904887428215
N    75.34885237644534    66.31738068559062     0.62252673605837
N    74.86401984486534    64.14657913594162     1.22590770394421
C    75.84959006344232    68.64345722410002     0.14861825450488
H    74.32166857631962    63.32550689416580     0.97485069028508
H    74.62452239608254    64.57522588850802     2.11578907304035
H    75.67917540411059    65.49902103150713    -3.07215234122919
H    76.24431900882647    69.25694585906255    -0.67040207402985
```

```
H    76.53079396164421    68.66664994003801     1.01218577925954
H    74.89274984705885    69.07373295184996     0.48906973508176
C    75.36466289529577    63.22466930453692    -1.56390127909319
H    75.41685023238166    63.16849541746413    -2.65957678584588
H    74.48331638743166    62.66967308026418    -1.23398536180518
N    76.55583917538711    62.45261300737876    -1.04934454402439
C    76.49414090183288    61.50785440815198    -0.09236635079855
S    78.05395760248007    60.89201103438781     0.25597367472499
C    78.79454347985948    61.91553218128233    -0.92132921303798
C    77.83805655425967    62.70420348937827    -1.53187864646892
C    80.27144092031693    61.87496694020541    -1.14814512790990
C    81.12658800884149    62.20729711911233     0.09998115333978
O    82.47698494608812    62.31858905140169    -0.23673623925447
C    78.10436018726837    63.71341118424230    -2.59698749480270
H    77.38232093743993    63.64221466699181    -3.42198190302591
H    79.10325518155140    63.55584545882462    -3.02024963480091
H    78.05741464052434    64.73507746652187    -2.18958365052058
H    80.54023835168699    62.59789915958405    -1.92896909734811
H    80.56096115766221    60.87865083446657    -1.52644647031557
H    80.94039697349363    61.45630273797445     0.89447242868092
H    80.80481766567729    63.18200589452216     0.50373043633955
H    82.87620412355338    61.44101480179808    -0.30416676308682
C    75.26924868774093    60.95227104459114     0.58681156423655
O    74.15864499572196    61.39062476725334     0.32677353798825
C    75.48852773467992    59.86412362538459     1.56231917640740
C    74.43980473578863    59.25087234101589     2.13478120900387
H    76.50701772219672    59.55439785948067     1.81141596187737
H    74.58041610763266    58.44203305349116     2.85582970949358
H    73.41899431274265    59.55060076666580     1.87968758138890
```

ThDP-Intermediate B 40 +1

```
C    75.14585282616557    64.98263052135790     0.22629171462115
C    75.37913037526253    64.63042099364149    -1.13087746000262
C    75.62279347178118    65.70866945555365    -1.98494087088977
N    75.73963374621890    66.96655597146602    -1.57301634046209
C    75.62568494729507    67.18104648641457    -0.25037245223635
N    75.30623842274714    66.23941680700665     0.64441038819997
N    74.78496965835714    64.05153963512657     1.16912257097864
C    75.84429217902253    68.57437824096512     0.25849185143062
H    74.16967663533448    63.29939129769561     0.87333753049059
```

```
H    74.52561472458038    64.47271787926604     2.05843445615345
H    75.72621129631248    65.53412389929794    -3.06440989091589
H    76.24576541246553    69.21474180924453    -0.53609265584590
H    76.52255925413678    68.56296675773335     1.12488671432968
H    74.88841137110737    68.99591100690735     0.61203958605938
C    75.37470889190743    63.21338956846391    -1.62489923121373
H    75.44088814571332    63.18178479103991    -2.72035113006234
H    74.49525652849347    62.63858506743054    -1.32337765775207
N    76.56503015719110    62.44736741104843    -1.10652628254880
C    76.49934474470176    61.53783664697079    -0.11895570195993
S    78.05439796065680    60.93970262461266     0.26950185176337
C    78.80359996333345    61.92528661616613    -0.93760420995114
C    77.85111205879500    62.69103401931047    -1.58202246235956
C    80.28143927252511    61.87656497116796    -1.15508829760635
C    81.13021679178951    62.19852457486424     0.09919922781125
O    82.48303541265231    62.30424706922164    -0.22870164343520
C    78.11886470929858    63.67662768347132    -2.66795248827871
H    77.40377363939318    63.57965425589921    -3.49669346988057
H    79.12252283024279    63.51851752606925    -3.07953475679978
H    78.06083220584445    64.70670849406089    -2.28305806166831
H    80.55893420766061    62.60077714718364    -1.93177795291333
H    80.56677315734066    60.87967563837901    -1.53561288949972
H    80.93572667394949    61.44488175454634     0.88942216745894
H    80.81105112186296    63.17311395445613     0.50556866756107
H    82.87798005911959    61.42488363843261    -0.29832054534605
C    75.27404428443155    60.99914612608138     0.57512389897650
O    74.16100092770360    61.26673388080616     0.17091120730358
C    75.52001249103213    60.14061919455422     1.78662101408230
H    74.56045802424673    59.74755231026191     2.14421691789370
H    76.19579385720799    59.29986844460520     1.56212611871256
H    75.98711856211899    60.74611182921947     2.58165756780153
```

ThDP-Intermediate C 40 +1

```
C    75.52583522077310    65.26125654335632     0.23572958005958
C    75.41872181268825    64.74748456077097    -1.08727589310815
C    75.32290870141377    65.71118901607365    -2.09227721168939
N    75.33253346534416    67.02081806168519    -1.87381786383864
C    75.45981389000313    67.41134761204353    -0.59412046229634
N    75.56351926060186    66.58194862776870     0.44840432059303
N    75.64339256502994    64.46123665563483     1.33577401057140
```

```
C    75.47306550507186    68.88332487729579    -0.31184702396613
H    75.20073987909927    63.54826153649800     1.33629436375414
H    75.60743996770209    64.96930590476214     2.21508032020685
H    75.23200365053474    65.38940691709350    -3.14008333649416
H    75.27325188531961    69.44855374251898    -1.23006543713266
H    76.45049712124955    69.18243645931224     0.10107999432609
H    74.71960300550928    69.12739417763639     0.45199254059684
C    75.38031444160957    63.28940678313162    -1.44286138978589
H    75.36226737649581    63.17445025702075    -2.53748618153379
H    74.49316375169049    62.78807152062208    -1.04359138612686
N    76.55374732731849    62.49071827894439    -0.95973811889001
C    76.42003485797234    61.38137263230921    -0.21929504328188
S    77.92471693251819    60.62897857003597     0.05519269212372
C    78.75601400324803    61.82209910058052    -0.89861250594287
C    77.86805155597138    62.77124617482470    -1.35221280628620
C    80.24104064065941    61.75179929298711    -1.04934428076275
C    81.01917390858527    62.09019889450166     0.24747148166838
O    82.39320216310969    62.16066237686364     0.00985072975639
C    78.20354795699514    63.96976143548972    -2.17713429191240
H    77.53414574583815    64.07640743802475    -3.04140878957338
H    79.22868273004043    63.88393921310791    -2.55676097507231
H    78.12786585601270    64.89216567559490    -1.58275684345826
H    80.56663120597295    62.45563123622224    -1.82662425218647
H    80.53353100550575    60.74519820077592    -1.39396238474760
H    80.75574111191786    61.36192301004257     1.04101088757739
H    80.69543778962294    63.08081943250200     0.60823450369337
H    82.76765434744631    61.27046341379948    -0.02532038895089
C    75.18209120464914    60.80117171380021     0.33174659797213
O    74.36098898875680    61.74001115957174     0.88049641035064
C    74.94705134004826    59.47948972084280     0.30935894168226
H    74.04842612440409    59.05856290119325     0.76809862550449
H    75.63722575935738    58.79341991862146    -0.18114134383609
H    73.53084594391218    61.34168395613862     1.18596921043642
```

ThDP-Intermediate D 39 0

```
C    77.65540369632527    62.63018110631371    -1.43140336340998
C    76.32522204624853    62.49009418435808    -0.91981321395106
C    76.14857867672178    61.49940017779382     0.03475535360017
N    77.13199345624869    60.73457888242855     0.51596811295761
C    78.35577031782007    60.98451553986853     0.02621515177703
```

```
N    78.65120512055766    61.88365928288417    -0.91347836696198
N    77.95771568820575    63.51217508926436    -2.40359639374493
C    79.49286861105030    60.17363454916161     0.58562399898471
H    77.21867760551358    63.85560805553671    -3.06547283046355
H    78.91137647114330    63.45346405503496    -2.74510514667510
H    75.14525188979137    61.31955895025182     0.44651245212151
H    80.02864629637854    59.65635020701390    -0.22614835048147
H    79.11101981721833    59.44014475655027     1.30694411680430
H    80.22310287353697    60.83402733166117     1.08177737932470
C    75.18478699513429    63.34702222716447    -1.38369651392285
H    74.25719508633652    63.06441994194754    -0.86436775989003
H    75.03827377459761    63.26792843694221    -2.46892814907701
N    75.41345798092252    64.79997785946470    -1.12258306297416
C    75.44788324692207    65.70727167816797    -2.11145380459664
S    75.70572764259462    67.28035105127196    -1.47948395055645
C    75.69217028899094    66.68676412530829     0.16420820163989
C    75.54260323077476    65.32732180001003     0.17071111937564
C    75.89133991775898    67.60833800411531     1.32893793690354
C    77.34060246322601    67.64458249606044     1.85086148492635
O    77.46385645249745    68.42886733652513     3.01195785031679
C    75.48965889189579    64.46231553177776     1.38739793737493
H    74.63843688545396    63.76533512669862     1.35838626669926
H    75.38479030557288    65.09114089204840     2.28087381704464
H    76.40132492452308    63.85617035793131     1.49409961949296
H    75.23840003415327    67.30632632597310     2.16296038502795
H    75.58139206729618    68.62966506264878     1.04909052524684
H    78.01808632035666    67.97596462129218     1.03865816176986
H    77.65169133956091    66.62365986852406     2.12517746880430
H    77.37673266274317    69.35931532609476     2.76629163616674
C    75.36221145362370    65.42195227857087    -3.58392583012860
O    75.92507627678872    64.34653953022891    -3.98543694975018
C    74.64352728813863    66.35006797640922    -4.29913714738531
H    74.48772472132821    66.19265326480898    -5.36858207868406
H    74.18352118204869    67.21685971189282    -3.82381106370632
```

ThDP-Intermediate E 50 +1

```
C   -12.50111771428253    -6.24725571702190    22.76670534547087
C   -12.87020811359376    -7.40980045220744    21.94227391693232
C   -11.92620379509785    -7.89753927983277    21.09374950998032
N   -10.66844626736485    -7.35442890203126    21.05508578930075
```

```
C    -10.31291075723676    -6.36348721347267    21.94894508228739
N    -11.16785316465360    -5.85391732942531    22.77122889537025
N    -13.38108607244682    -5.60848409308315    23.46777319148455
C     -8.88356187731085    -5.91360408866030    21.93902460382010
H     -9.97911134955801    -7.71677897791468    20.40484278279341
H    -12.95352269788827    -4.82619539795470    23.96987048679635
H    -12.11309781503737    -8.73290465148862    20.41568278998923
H     -8.21708925444607    -6.72908705340904    22.26695466820976
H     -8.56969695269278    -5.60351126568033    20.92890980666366
H     -8.77312442989009    -5.06756831327636    22.62645160960784
C    -14.25326072711912    -8.00642299493118    21.99022402806407
H    -14.32487623400529    -8.90084420123533    21.36015816655602
H    -15.01489303997626    -7.29306192720079    21.66066538859911
N    -14.63139409642132    -8.40961082805683    23.37455740985303
C    -15.46262260725756    -7.71371563079154    24.15076646636030
S    -15.60253775960452    -8.42251190389968    25.69715588216244
C    -14.50677769003161    -9.71078485747574    25.27297190161581
C    -14.07007144504879    -9.54408174830901    23.98629126777864
C    -14.22614871948224   -10.81265452737386    26.24612758969053
C    -14.03202623043929   -10.35483370104559    27.69348324325036
O    -13.87516574766003   -11.51822897520959    28.45780859052765
C    -13.15514070240156   -10.44922482210133    23.23132753478386
H    -12.83462154967163   -11.27779265711948    23.87319511348364
H    -13.66105788545879   -10.88416853550969    22.35392499674807
H    -12.25848007901193    -9.91822327678192    22.88223930420012
H    -15.06042967061726   -11.53675620506190    26.23499130250185
H    -13.33166373746023   -11.37010524002238    25.93204933896807
H    -13.15503319105010    -9.67783109237033    27.75435060134316
H    -14.91683586615649    -9.76264862619775    28.01572328202876
H    -13.68700733656456   -11.27887617950059    29.37431317731906
C    -16.37771226701622    -6.54292749515529    23.75645804587933
O    -15.86711433432328    -5.78607533086012    22.71598825849684
C    -17.62331217054131    -7.26924077706341    23.15643193892345
O    -18.77483230711331    -6.94195857287921    23.74299635099926
O    -17.51615523296007    -8.05422570680511    22.25171382069333
C    -16.66061904224582    -5.67303524802809    25.01449605109707
C    -16.97000287354343    -4.21390520041018    24.63075998157179
O    -16.07039060663718    -3.40779924764615    24.65473336033663
C    -18.38117085726467    -3.86242965293459    24.23299438704289
H    -14.91905799874470    -5.53247293727593    23.00140138842002
```

```
H   -19.49331173595734    -7.41543768774875    23.28391344171835
H   -17.46408383454613    -6.11083590295367    25.62399648881212
H   -15.73901970592781    -5.61455547755634    25.60796422331735
H   -19.11545171056311    -4.30498255155710    24.92240878709922
H   -18.49430674729632    -2.77111891775765    24.19170853429881
H   -18.58410799838106    -4.28523762771537    23.23475987675201
```

ThDP-Intermediate F 50 +1

```
C   -13.39060802925408    -9.77633288352404    23.63264916129179
C   -14.35857976991903    -8.69897605320841    23.37118141804763
C   -15.28433739385424    -8.43382512100626    24.32724268289420
N   -15.27653745611535    -9.12247837243274    25.51488106680266
C   -14.27602613776218   -10.03529010278161    25.78236379054297
N   -13.37492114014861   -10.33655345758388    24.90785829964332
N   -12.57802368753531   -10.18147742183169    22.71367541722837
C   -14.26443991644412   -10.64702633103191    27.15057866085569
H   -15.98326482037262    -8.92826159725994    26.21620201354606
H   -11.97406734430872   -10.92256523484058    23.07770806074732
H   -16.06154100988032    -7.67474004202441    24.21052510307153
H   -13.98183852269169    -9.89213441222520    27.90370136506117
H   -15.25610155641069   -11.04331778325687    27.42209620285785
H   -13.52989240770191   -11.45973898832865    27.17396957539608
C   -14.34762407049756    -7.95905061239007    22.06155521357213
H   -15.06887762742647    -7.13144645353339    22.04894123748137
H   -14.59222827900477    -8.64324273169628    21.24458398780901
N   -13.00609344230908    -7.40103796845655    21.75486300814093
C   -12.25375788216292    -7.83806021328467    20.74327831992171
S   -10.73159413195505    -7.08115348449354    20.72646459041288
C   -11.13769214629754    -6.12045659648912    22.12231152176614
C   -12.40281174770635    -6.42201458775895    22.55355153517183
C   -10.13977950367804    -5.19067266474761    22.73406160976324
C    -9.20325190680271    -5.89757810756315    23.71987797309946
O    -8.36262382628742    -4.91498206079243    24.25826634350994
C   -13.11793725432762    -5.79139533523548    23.70390163856746
H   -12.55963582746804    -4.91153474513038    24.04568830205923
H   -14.12575946437929    -5.45575850349700    23.41452334393262
H   -13.21846641902176    -6.48270617473438    24.55314632918749
H    -9.53059364465440    -4.70184042105048    21.95783279417186
H   -10.66189644008965    -4.38548654434441    23.27287986503642
H    -9.81462111969452    -6.40737629911056    24.49374150397811
```

```
H    -8.64359289761706    -6.69118665149781   23.18223163145671
H    -7.73702251891526    -5.32490491518379   24.86898422804688
C   -12.64576507995502    -8.82494035169311   19.63378577938460
O   -13.29592860254465    -9.95027815711819   20.15242284778662
C   -13.72483051256409    -8.10829602459294   18.76792776853386
O   -14.31182657856832    -7.11052508563568   19.11070340541739
O   -13.95371643500814    -8.76533857194120   17.63713156212136
C   -11.41598520034326    -9.11284139933848   18.78991878260966
C   -10.44671593809327    -9.98346662341700   19.13682486150293
O   -10.53691754377976   -10.66817568033466   20.30191701426083
C    -9.21233379320082   -10.21965416118869   18.32048383351561
H   -12.91069644164951   -10.15999794585522   21.06024626039783
H   -14.67685185496346    -8.31374343118477   17.16560116562570
H   -11.32327031183883    -8.59568957791855   17.83576390745078
H    -9.77969954710189   -11.26048465599293   20.40075471695755
H    -9.21278697953290    -9.62182169639314   17.40119503044870
H    -8.31738228617880    -9.94858861561202   18.90920442642291
H    -9.11985855198287   -11.28361214945724   18.04163784249156
```

ThDP-Intermediate G 49 0

```
C   -13.50379293473438    -9.90747182129967   23.66369238853806
C   -14.42764719861020    -8.79408461456532   23.36617241321537
C   -15.34892173719090    -8.45990305917280   24.29664894089528
N   -15.37043985292715    -9.10311607334439   25.51931992841226
C   -14.37018106817751   -10.00347336941459   25.83823652312215
N   -13.47567129213004   -10.37595511503452   24.99107524112285
N   -12.76814148530465   -10.39399414994336   22.73689867419290
C   -14.35786382017723   -10.51392109334545   27.25147967818914
H   -16.05638592775721    -8.84312110482785   26.21672429321903
H   -12.14993219092038   -11.12595326539861   23.09578465546012
H   -16.08548980628326    -7.66654962583170   24.15148887056236
H   -14.07891307466374    -9.70784283774739   27.95138570935890
H   -15.34831766602045   -10.89526313762590   27.55005194132450
H   -13.61963545644676   -11.31993365238471   27.33062528315148
C   -14.33790754958959    -8.09731097261113   22.04138792737900
H   -15.07407822982115    -7.28385165348023   21.95965871806194
H   -14.51361263973977    -8.82385096857614   21.23720753592572
N   -13.00012740001283    -7.52481994221416   21.76371463811351
C   -12.33090648671986    -7.80089447000724   20.63892479935896
S   -10.93702691861956    -6.84090601597856   20.46080498848687
```

```
C    -11.22633378243860     -6.09375540223579     22.02260910076800
C    -12.36971861183427     -6.57708717338751     22.59122056395679
C    -10.20396911982089     -5.17861552980350     22.61934153505941
C     -9.11363752000200     -5.96432670198376     23.34986686197205
O     -8.20846437669675     -5.03942527293937     23.90917331870533
C    -12.91975556241558     -6.18840287123502     23.92788889643194
H    -12.36070565563034     -5.32505006473244     24.31122257366613
H    -13.98153534639452     -5.90312517848541     23.87476613439141
H    -12.82962721627092     -7.00500378020093     24.66035002053401
H     -9.72874055387985     -4.56275323121500     21.83807262785811
H    -10.67087129841538     -4.47873696427951     23.33102547585096
H     -9.59192133462261     -6.60192618883416     24.12190568329172
H     -8.62352405983435     -6.65067114470871     22.63089867970264
H     -7.47233040635388     -5.52867519339472     24.29673975043178
C    -12.69906935332828     -8.72538038946318     19.47292823293149
O    -13.36645857377805     -9.88408321340308     19.96589434568091
C    -13.75344041434558     -7.99045432075762     18.62712700773283
O    -14.32145104948521     -6.97441207391866     18.96386832516377
O    -14.01927499607533     -8.61876399241116     17.47881949441971
C    -11.42586412845034     -8.96337210898841     18.71511997443973
C    -10.44349312780630     -9.63377134238501     19.45639183466197
O    -10.60375109775348     -9.91164916321288     20.67747492838268
C     -9.10310079901421     -9.96478724625216     18.81548468274921
H    -12.79273214781579    -10.23061299545377     20.69055385882204
H    -14.71705398651284     -8.10321180691819     17.03819507905815
H    -11.39613648732563     -8.83304618190732     17.63431169430799
H     -9.05379676673707     -9.70665628955328     17.74618702739860
H     -8.31314497782697     -9.41403174189465     19.35291825738661
H     -8.88600996618648    -11.03756783724820     18.94844316919599
```

ThDP-Intermediate H 46 0

```
C     70.80809186590241     66.70037746472021      0.65402712578975
C     70.95098846862797     67.56399320729005     -0.47096277460334
C     70.04070028169977     68.61401163827034     -0.52012274076744
N     69.13705524782000     68.88367563854621      0.42711593678379
C     69.15241873901088     68.06589156659550      1.49009850824309
N     69.92788452058414     66.99147970120212      1.62928927315030
N     71.53471465553093     65.56980039446741      0.80952737792461
C     68.19363236038735     68.37139139416928      2.60671839671300
H     72.01279747445336     65.13414850643470      0.02138431698340
```

```
H    71.28535192275790    64.99649034795009     1.60628023242333
H    70.05349636887200    69.29029226797303    -1.38642374700203
H    68.24471229338774    69.43849487132046     2.87279969692676
H    68.41745518748118    67.75141131455294     3.48384419210759
H    67.16030556361720    68.16861682155471     2.27664954408227
C    72.01708584165357    67.39336143147827    -1.52856416087465
H    71.82689817292376    68.11578673422859    -2.33682724999896
H    71.98218137998168    66.39863287917741    -1.97988892520825
N    73.39308071176359    67.59640872764011    -1.03902299427497
C    74.31019048918455    66.56491280590605    -0.80027531234264
S    75.81349898362403    67.25975732877205    -0.15515225944006
C    75.10762012534984    68.88392619492990    -0.14916099815135
C    73.84811004904182    68.86949445612706    -0.64653196337380
C    75.87720073605884    70.01657725629148     0.45036383770695
C    75.78534564999894    70.05508426426792     1.98448750117284
O    76.44740408455669    71.16428409482144     2.54837824120838
C    72.98107437731898    70.08071602568698    -0.79438758003054
H    72.51760601048484    70.13747415379765    -1.79112346099428
H    73.58531322389302    70.98576359120987    -0.65443858996929
H    72.16589044576158    70.10060951141921    -0.05552218570992
H    75.50928723560671    70.97689131827869     0.05763612788671
H    76.94076288238506    69.94828766688380     0.15231265867409
H    76.15267385596677    69.09398879668288     2.39703650941029
H    74.72680112325205    70.14034557980467     2.27759793948446
H    77.38917309861237    71.09306630681618     2.34329433627874
C    74.18371345778351    65.20872740056444    -0.97475156095660
O    73.00675962408482    64.69524097837191    -1.51378412831509
C    75.16868217532483    64.27291625084248    -0.48572664492595
C    75.14660272605656    62.92457608889487    -0.60697181581953
O    74.08388926901594    62.32206412893559    -1.28221995459129
C    76.12016343822764    61.97729293545464     0.01150459405177
H    73.14626940871436    63.74018192719252    -1.66459047868155
H    75.99493171878858    64.68842361358085     0.09411933390012
H    74.37881744417021    61.50873667524233    -1.71339727397756
H    76.59373285163718    61.32359010802283    -0.74791104160677
H    76.92188993201299    62.51994296491051     0.53118306724449
H    75.61777852663143    61.31373666871934     0.74011809346914
```

ThDP-Intermediate I 52 0

```
C    72.40214283185695    66.17160120294272     2.60592540563586
```

```
C    71.82617399546385    66.90301083187988     1.52513300293339
C    70.87330251314235    67.84303198445080     1.89716733689448
N    70.53650528833313    68.12615909410502     3.16126857892357
C    71.19532842052041    67.44023916096343     4.10410557728384
N    72.09042312875356    66.47551335431625     3.87508282699973
N    73.29055047978558    65.16507813956261     2.41409740725746
C    70.90643953239011    67.77325158753074     5.54309864417005
H    73.37206784979416    64.68552998138743     1.52224952999818
H    73.58779813513544    64.67226474730798     3.24749424686270
H    70.35497355434973    68.41510499324876     1.11658770205654
H    70.13993808623054    68.55531884672338     5.60835746518886
H    71.82647358844733    68.11809438058593     6.04305534719479
H    70.56731215049354    66.87461714784449     6.08282580372803
C    72.24903226115485    66.72589481712998     0.08718332703176
H    71.58507796535861    67.31810712887479    -0.55862977257328
H    72.16701652741479    65.68713479622127    -0.24308286243277
N    73.64067564336428    67.15384305223152    -0.15697712221645
C    74.71557698322373    66.29222800707487    -0.34347963450497
S    76.22838763360774    67.19930276478591    -0.19609118137796
C    75.34717210786846    68.72043453754345    -0.03389905931731
C    74.00954592579058    68.50617372774184    -0.02519736161742
C    76.06443859441917    69.99851382920883     0.26711887783018
C    76.19445312555239    70.22254811206156     1.78426063591826
O    76.66784271565398    71.50681818210306     2.11515982629276
C    72.98173212771309    69.58842527470777     0.07215122065723
H    72.13528154741292    69.40758564935069    -0.60565153099326
H    73.43424950806346    70.54791546131342    -0.21026993167258
H    72.57489981438731    69.69245550135052     1.08861115327186
H    75.51522682899977    70.85541976643263    -0.15567514827973
H    77.06233679576485    70.00755963964406    -0.20429211904792
H    76.81473304490184    69.41586085066334     2.22184141894881
H    75.19619706028938    70.12884296117568     2.23952311566367
H    77.58863429392756    71.57721202590913     1.83026817962683
C    74.74660135752526    64.93029355406854    -0.56896019601584
O    73.58380866143933    64.18572246378284    -0.36815499328354
C    75.93562456364823    64.21463812475399    -0.91465559233197
C    76.07991458503223    62.85850841586313    -0.84528908031864
C    77.22253960835845    62.09946141561451    -1.33867316550079
C    77.48407823035630    60.80847317056239    -0.83002345855170
C    78.58096660581494    60.06354909425330    -1.25653218682886
```

```
C    79.43899044594620    60.59514817122668    -2.22311865409300
C    79.20028138577622    61.86346236017746    -2.76278858185244
C    78.10149674463615    62.59872786561881    -2.32570091753273
Cl   80.81422846617770    59.66359003057824    -2.76668672675008
H    73.60646361744790    63.42882599172572    -0.97482112399214
H    76.76413525481944    64.81866549805977    -1.30060165562992
H    75.30597359055500    62.27370589506447    -0.33500856400021
H    76.81752967532061    60.39023711005474    -0.07046140037932
H    78.78183362969715    59.07351553340463    -0.84392714385408
H    79.86859288198796    62.25866745030266    -3.53029895902858
H    77.90377563589462    63.57191531651370    -2.77993250639133
```

ThDP-Intermediate J 53 +1

```
C    78.39212718334640    62.31664170541672    -0.68400992021924
C    77.84688474939948    61.45248456925612     0.37568445666451
C    76.49348810796049    61.39108190220279     0.49358257084207
N    75.68035570431822    62.06128715855555    -0.38223718493624
C    76.22442562660707    62.75889145401840    -1.44343763850979
N    77.50053646582749    62.86059249446729    -1.60509200464254
N    79.65344135084129    62.58686986521131    -0.75090278808371
C    75.26377044808174    63.40083536698561    -2.39650412065377
H    74.67367407568476    62.03451219884190    -0.25863247852942
H    79.83336510118653    63.21601281982408    -1.53822697038565
H    75.98854853755549    60.82503216149804     1.27970344454053
H    74.59383868013170    64.09442741064490    -1.86088377067337
H    74.63807896398656    62.64484371947210    -2.89960262954537
H    75.83108906733270    63.95821634805020    -3.15078943674049
C    78.74059556803704    60.76688321285510     1.37453390602776
H    78.14099990527976    60.30656375355441     2.16982661961961
H    79.43958735831490    61.47788697218986     1.83093375886233
N    79.58803279965784    59.67620357920467     0.80315898352261
C    80.90698122293331    59.77265815304611     0.61367409226170
S    81.55138356891747    58.30133314641102     0.02278474703465
C    79.99356493976582    57.52175296713107     0.08413081296996
C    79.03912615788441    58.41210843348565     0.49351841916097
C    79.86454563973707    56.09116411138126    -0.33230331976311
C    80.23724225733946    55.83469597026475    -1.81020092718674
O    79.99922950120920    54.50883206573348    -2.18329215789085
C    77.57744809381147    58.14608511812436     0.63643512587535
H    77.22857548813984    58.30380420196041     1.66951524547744
```

H	77.36301556946734	57.10552868824595	0.36715934310881
H	76.98396993003230	58.79063822248312	−0.02634416854152
H	78.83396196416109	55.74534319498250	−0.17076300614572
H	80.50566914128392	55.46486136570075	0.31172417634107
H	81.28653303904878	56.14763107562747	−1.99034847144404
H	79.60713995370978	56.47059124257532	−2.45184525885650
H	80.68235966310114	53.93933114511220	−1.80451466904528
C	81.88495693097316	60.94861766382281	0.83342433776453
O	81.28293756682056	62.10785753162197	1.29258612670590
C	82.62999884498491	61.10518968416532	−0.47279196937861
C	83.90301325186526	60.72878434932044	−0.68322441022610
C	84.57583730010649	60.71861516712690	−1.98691701726343
C	83.85067823313591	60.79955217689124	−3.19303861844604
C	84.49265996362823	60.77243123783831	−4.42448166442773
C	85.88740651794471	60.66009307524370	−4.47113869693865
C	86.63383385680878	60.57040443528168	−3.29223755926987
C	85.97520059221334	60.59383765941663	−2.06311805419452
Cl	86.68250451500074	60.61682146754472	−6.01442623320681
H	80.73261440535278	62.45570292178484	0.52474371416402
H	82.58608018770575	60.59307316819122	1.61127843995763
H	82.01850509584766	61.49793559301028	−1.28925663442339
H	84.49940273924437	60.37497620800254	0.16739636342560
H	82.76096045295743	60.86735091378170	−3.17009579118531
H	83.92678167601288	60.82776758862403	−5.35524478035968
H	87.71958508454195	60.47455401333288	−3.34792608673160
H	86.55968596076245	60.51844635048440	−1.14238824648194

ThDP-Intermediate K 51 +1

C	77.53262023253610	66.06387388079872	−4.22595430758022
C	76.68769354506281	64.95806620205148	−4.53006972587925
C	77.04344538920343	64.23385811624929	−5.67030455338975
N	78.11544027164770	64.48757572216788	−6.41016040592835
C	78.89193649400669	65.50967078697391	−6.00304257252746
N	78.62999818038926	66.29759378971428	−4.96080063028712
N	77.30638834712536	66.90105548121170	−3.17891109223427
C	80.12450377314296	65.80168158301532	−6.80405454679775
H	76.35767704727222	67.03112170174832	−2.83044262046952
H	77.92674079019920	67.70369908182329	−3.13883589738325
H	76.41624802814815	63.38828762622664	−5.98821407707952
H	80.54815148664170	64.87251034070776	−7.20892029471374

```
H    80.86251474838373    66.34265850693811    -6.19741293767277
H    79.85507510746046    66.43904625748112    -7.66368309294970
C    75.50740641075451    64.55305828720535    -3.69815660460497
H    75.00345746786380    63.68642407351214    -4.14865573084568
H    74.77353317926195    65.35490185687117    -3.57485268647510
N    75.90067283951606    64.13694775017177    -2.30523388028518
C    75.47349388133912    64.76108339952218    -1.19775885538139
S    76.14004674066140    64.04971263419169     0.20588048121545
C    77.00663746441028    62.87558451996773    -0.72719226117516
C    76.76049305886565    63.06133656235338    -2.07149089053142
C    77.94534617338638    61.91813717709589    -0.06686663314110
C    79.26391629269058    62.59772105525395     0.38592190641424
O    80.21950309255246    61.66780548772848     0.79577552043334
C    77.31335396357949    62.22167545592494    -3.17559988387465
H    76.54009056891908    61.95655504498539    -3.90879831839413
H    77.72193079765283    61.29142940673970    -2.76350496666395
H    78.11917131252528    62.74207910061856    -3.71275596912362
H    78.20526190571458    61.11317068841172    -0.76825442193117
H    77.45796285651990    61.44084293076162     0.80053694247548
H    79.03361568471627    63.35281434703580     1.16474299264796
H    79.69131290280347    63.14501346163669    -0.46987893508183
H    79.98617677474172    61.31872168756660     1.66663572768678
C    74.63411907237193    66.01308270126563    -1.08819096897079
O    74.63555079801667    66.82373913801908    -2.01350975579845
C    73.95682772737985    66.20076724572191     0.18178192452426
C    73.51741620812234    67.44252077206110     0.54222138835182
C    72.92653135551654    67.83715636096207     1.79818174423325
C    72.75372818734886    69.21571293130673     2.06140304838125
C    72.25476264930946    69.66132070233165     3.27902768741260
C    71.90346818144781    68.72241201598835     4.25837281477681
C    72.03997404227260    67.34403986516243     4.01569321080410
C    72.54910762683791    66.91170694965763     2.80103189753598
Cl   71.29458588149147    69.26276241573140     5.78178716819956
H    73.86450883658890    65.34933992023129     0.85915490206369
H    73.68431936523029    68.24131931377063    -0.18928704041185
H    73.03444544631284    69.94496140840376     1.29757687254951
H    72.13250666172597    70.72610298689612     3.48334107512495
H    71.74306452524795    66.63234780647844     4.78720890184271
H    72.65033862508447    65.83895246134918     2.62330935090941
```

ThDP-Intermediate L 55 +1

```
C    84.27460947530005    44.42306263903343    160.17605021542988
C    83.41694943187349    44.44377384580420    158.98121110556059
C    82.08540672075694    44.63563740913027    159.17079414427965
N    81.57200994006864    44.81439195410211    160.42876876039287
C    82.40601116055433    44.82604416670861    161.52747581622634
N    83.68181277701923    44.66231526276369    161.41238158663609
N    85.54230138737871    44.17454115799153    160.09262968270667
C    81.76169940361747    45.03133922219721    162.86461919169869
H    80.57240230810423    44.93462717587460    160.55526857854443
H    85.95912625993058    44.19006030986581    161.02756174822602
H    81.36440964518989    44.64451667728368    158.35062974280328
H    81.27246064614111    46.01828256045238    162.92014588243396
H    80.99192414196509    44.26451382133780    163.05019800884125
H    82.53068074267097    44.96376195302729    163.64259541622815
C    83.96030298203601    44.15108294932366    157.60768742615551
H    83.15725257445577    44.16977953261404    156.85998198502350
H    84.43686817749082    43.16524461888974    157.57666069013453
N    84.99273173867643    45.12427131536062    157.15673529116592
C    86.29907756975527    44.84353828019445    157.07321923980311
S    87.18328248904386    46.19065043834491    156.51041011460251
C    85.73833747274738    47.14407891697978    156.34909993007622
C    84.64872443608927    46.42788874518643    156.76927479697952
C    85.78565685064052    48.56401731258968    155.88069051515592
C    85.72362122741630    49.58310957848232    157.02527251286651
O    85.53280922862454    50.83684269101240    156.43006542915961
C    83.24019885241549    46.92279186175788    156.81287717223367
H    83.17463619812493    47.88534753067911    156.29128201157067
H    82.54382713432304    46.22965280672719    156.31713547323005
H    82.89338694285642    47.07078069590121    157.84714554557740
H    86.68960526605139    48.74877054642054    155.28110611703963
H    84.92996689277281    48.76111990042681    155.21491587082619
H    84.89132795139571    49.31142321911867    157.70839738179120
H    86.65694250939600    49.52201967250756    157.62022329092616
H    85.65345631024164    51.53192690406171    157.08997641753814
C    87.01371478567560    43.49491831376765    157.26809019310949
O    86.32365377285018    42.62767194865921    158.10760919233590
C    88.44759076114192    43.73302185958028    157.71761362673320
C    88.75103541297770    44.48999902148478    158.78661251607554
C    90.07839595782092    44.72205862219474    159.35576814771349
```

```
C   90.20129224080156   45.56870402058246   160.47562021120322
N   91.35483242543965   45.86980332222335   161.06584226988306
C   92.46884033859746   45.33039931109906   160.56613051989117
C   92.47294342694181   44.46485219357977   159.46186464133072
C   91.26334824195247   44.15739142639958   158.84875434723864
C   87.01796353473797   42.81517782339536   155.86220972175218
O   85.77095616989691   42.50147999921615   155.48202061822457
O   87.99854410083783   42.60754613395165   155.20577110895795
H   86.13862073432371   43.13514628777889   158.96213556514601
H   89.20313491595691   43.17713014903117   157.16034793198997
H   87.93680523116554   45.00144272781264   159.31322263588183
H   89.29966781561002   46.02368816579427   160.90663086032336
H   93.40700178153367   45.59916957504553   161.06466222269862
H   93.40970282016946   44.04281315653787   159.09209584610807
H   91.23806192864370   43.49024053512935   157.98377626227185
H   85.82720475779978   42.05490673458505   154.61815146926648
```

ThDP-Intermediate M 51 0

```
C   92.01174608379159   43.73380235803417   159.25165746923795
C   93.03635208709777   44.64699081265633   159.63574728817642
C   92.58468334543873   45.91089471956926   160.00579147624779
N   91.30309661446491   46.28254648781159   160.04079576495857
C   90.41583336895170   45.34057343616571   159.68580315552575
N   90.71802945377193   44.10197517619010   159.29925523467023
N   92.26423208668929   42.46706938294935   158.83955574464045
C   88.96192406172233   45.72736950803422   159.69745213134357
H   93.16798415306661   42.21365265620119   158.44793851397645
H   91.45387894048885   41.95678944582072   158.50797908236964
H   93.31976685637986   46.67491526668753   160.29226302758778
H   88.78542918813426   46.53526961827223   158.96913369438241
H   88.67727257372121   46.12426845828990   160.68459243807570
H   88.33330537349204   44.86582420881817   159.43918791434359
C   94.51011796451564   44.31693390107908   159.61919088651553
H   95.08204035709664   45.23589156328212   159.81632075647528
H   94.81721923444941   43.95797802637526   158.63462155445052
N   94.93813216179728   43.29279385745291   160.59308092553223
C   95.44562753128710   42.04560804958209   160.24483384508920
S   95.92568247513057   41.18225745139222   161.71159386937555
C   95.57416653560317   42.59587645461843   162.71286695535420
C   95.04130010952839   43.59407508350928   161.96532888181147
```

```
C    95.82756474643254    42.55581791449953    164.18989526932199
C    94.59700292293336    42.20194927947697    165.02619607791326
O    94.98565865283605    42.22880358145530    166.38336228424487
C    94.62178762911017    44.93068449657246    162.49217186711786
H    93.53020779098483    45.06749412606635    162.45224708055287
H    95.07357282505311    45.75622381968022    161.92171569232298
H    94.94384802569256    45.03474756619328    163.53618405145056
H    96.61926121588243    41.82552584432034    164.41925098346451
H    96.20554095140758    43.53058420797412    164.54003347013841
H    94.21803083933162    41.20698576539206    164.71519685627146
H    93.79000185953618    42.93172409313941    164.81164748067783
H    94.23291724936394    41.96203877406725    166.92540995353630
C    95.57986221139525    41.45151580106754    159.00308124955714
O    94.96750967459243    42.05213203079133    157.90249289144958
C    96.27205473714805    40.22390079086993    158.76185937789430
C    96.21272441913398    39.53233135908724    157.58483370539537
C    97.01936873512516    38.38454455017182    157.20239291914922
C    96.72093746085702    37.69561321152535    156.00253722552580
N    97.39321565051154    36.64501274551956    155.54411614917422
C    98.43075517827087    36.20128250539899    156.25981193488190
C    98.82764914475031    36.79588935292424    157.46356911237126
C    98.12144529019706    37.89814659975637    157.93633385179240
H    95.52906028667091    41.89899842296043    157.12608497330751
H    96.92043423754600    39.86720118438549    159.56862850065127
H    95.47619083151613    39.84468507657077    156.83534872028693
H    95.87440385810670    38.03758457884308    155.39224001816675
H    98.97494971543985    35.33823574400168    155.85870812054711
H    99.68505447510327    36.40206707266217    158.01477506149575
H    98.43149682845171    38.38954458183444    158.86165451119945
```

ThDP-Intermediate N 52 +1

```
C    84.29487722088253    44.61131974296374    160.05322069580805
C    83.42630509905514    44.38410933829205    158.88740427764643
C    82.08491047950707    44.47294356642686    159.08272999179576
N    81.57333995097670    44.81389860185693    160.30878904211957
C    82.41899428677424    45.16022158814629    161.34293105660396
N    83.70242108081840    45.08743615362195    161.22078547062003
N    85.56709264236530    44.38407382112109    160.00520886437994
C    81.77708232762446    45.61905298230540    162.61833084297302
H    80.56864543324323    44.85268724049086    160.44606305425580
```

```
H   85.98627646426117   44.60638076257597   160.91238795073414
H   81.35626856939285   44.28673618197713   158.29075638628697
H   81.03004940711988   46.40672361298974   162.42980702270140
H   81.26261719063211   44.77711242709807   163.11261567400234
H   82.55247556812691   46.00104030844536   163.29252769894430
C   83.99303328825060   44.02096975355103   157.53766356778715
H   83.19285092514308   43.92613756339661   156.79182851366431
H   84.54919663014030   43.07688731394519   157.58453358494776
N   84.95166901789004   45.04809924885478   157.04467115129793
C   86.27764608824874   44.86672458007657   156.99358313898011
S   87.07261225032281   46.29780981985257   156.51548434534686
C   85.56537611954764   47.15182872776290   156.34493369437254
C   84.52146014585105   46.33665268578282   156.68753513815318
C   85.52987639514915   48.60189895828844   155.97971498110198
C   85.41572423264820   49.51171069507940   157.20943509314000
O   85.28424168191401   50.82222902953906   156.73530910978099
C   83.07735041994469   46.71813674664546   156.68117615539578
H   82.95109989823418   47.65905014943377   156.13139672031758
H   82.45319473197100   45.95865990818870   156.18690036851962
H   82.68923250088844   46.86512036804877   157.70046181590598
H   86.42657474031999   48.88458032576892   155.40773320892359
H   84.66845723608822   48.80217587351584   155.32309074308620
H   84.54088268812647   49.19072795467073   157.81512106794565
H   86.31185415218255   49.37095377655979   157.84709792794175
H   85.32466890362360   51.44040697073940   157.47627788305527
C   87.08226038977467   43.57893758435177   157.22266196571550
O   86.47370995263127   42.72622909649724   158.13024693041095
C   88.52816111781115   43.93082746347198   157.55913630674914
C   88.91261478512121   44.13774991474045   158.82992844728699
C   90.24840443792853   44.42762369485492   159.34800594500649
C   90.38572906961990   44.70739006530788   160.72241993593542
N   91.54226783350872   44.96657924066670   161.32432786903343
C   92.64944798778248   44.95472602291213   160.57886002049875
C   92.64121958684291   44.69282431027702   159.20130962202185
C   91.42580041986503   44.42839029731641   158.57839227550767
H   86.23262991389421   43.26766909072554   158.94208264400388
H   89.24014944788375   43.96652119859110   156.72848463728030
H   88.14197569583847   44.06951763281297   159.60513970940633
H   89.49024988191614   44.71636655315424   161.35799748278390
H   93.59197763252941   45.16201540635418   161.09764354068935
```

```
H   93.57594506758282   44.68895002962749   158.63678609833235
H   91.39267947028097   44.19935723127519   157.51029030489653
H   87.07855254192289   43.06865138905046   156.23919202590446
```

ThDP-Intermediate O $52 +1$

```
C   83.30895275793763   43.74520232327485   161.41173764533698
C   83.48768668853498   43.74774044655690   160.00570855354803
C   82.37871735912265   43.35235909436427   159.25573841953863
N   81.19864826202144   43.05788140034463   159.78251985238251
C   81.09746233724940   43.17149274855905   161.12069261993503
N   82.10456891729362   43.48639453774123   161.93974444069269
N   84.31372601003991   44.07297310870263   162.27850157190943
C   79.74770759749359   42.91344632300062   161.71789554227180
H   85.27970097294613   43.90657868994455   162.01264999978227
H   84.09200006349695   43.91898736667868   163.25796437911413
H   82.46079573069734   43.29002573953433   158.16097768194484
H   79.48554195819116   41.84988719793299   161.58957132748048
H   78.98482842407728   43.48974615450800   161.17419784180183
H   79.73333522422089   43.16932437937400   162.78460153916640
C   84.74144857634288   44.21040660738723   159.33599484097027
H   84.60113347002353   44.22842551406163   158.24426427675471
H   85.59982338900971   43.56432544472838   159.55331849144116
N   85.16109894257093   45.59056062902193   159.74790397796991
C   86.43807383767708   45.87074904310743   160.04284393829440
S   86.65843767880632   47.52134183312479   160.38576313654011
C   84.97760748580865   47.85567251962292   160.09922080482957
C   84.30697020168316   46.70003741531515   159.77267804081117
C   84.44230044985045   49.23673110091154   160.30516470795291
C   84.02479129627795   49.49884750302726   161.75936524651047
O   83.64630209482716   50.84441230785222   161.82708205373473
C   82.84984506386917   46.61223685025654   159.45435260833284
H   82.28487953619947   46.14197966207577   160.27210373872327
H   82.66140029326367   46.02939605537378   158.54264442016188
H   82.45049638198209   47.62246695396924   159.30117172493587
H   85.19811486792905   49.98770794922289   160.02615745171957
H   83.57690106644259   49.41140941730155   159.64813950858837
H   84.87880236751958   49.26084477659171   162.42771621992571
H   83.20400027923368   48.80409325907680   162.03273215103704
H   83.33726828813752   51.05257508347376   162.71810264353692
C   87.56355474727361   44.92381765626565   160.07631789550345
```

```
O    87.36072223754132    43.90265868934168    160.96446324814559
C    88.64671773268168    45.11022586663977    159.29986848982560
C    89.85575258109741    44.21075061349425    159.24996446042621
C    91.04825235138549    44.88165756999592    158.60316370256865
C    91.33910879263111    44.66544504604083    157.24749887146163
N    92.34723940840651    45.24812542930713    156.60125337613803
C    93.13435101974969    46.08187148160017    157.28449412190690
C    92.94159911632131    46.37323045051807    158.63995005294748
C    91.87777787190183    45.76434942305313    159.30648256279247
H    88.08437810268592    43.25895201718650    160.91752775498870
H    88.63398416936450    45.96795616358016    158.62308177350349
H    90.12271266960315    43.86680799472061    160.26684329262795
H    89.60281896593767    43.29933776231006    158.67627531370525
H    90.71660149449754    43.97682915381490    156.66025316754906
H    93.95819559528805    46.54447179448072    156.72967726972078
H    93.61534559325055    47.05829274400251    159.15899170521308
H    91.69578067960475    45.97336470762804    160.36510054330009
```

ThDP-Intermediate P 50 +1

```
C    84.04170282411984    43.41395792984189    160.90856578882241
C    83.27174159580834    43.94714073283748    159.84069913867725
C    81.88628914000806    43.88528352544486    160.02555140089271
N    81.30316865507528    43.42835234078426    161.12609760703401
C    82.12104637423656    43.01721164735570    162.11400336545640
N    83.45291287480768    42.98887941216510    162.03356301364838
N    85.40811690456083    43.33555372068477    160.85940842435147
C    81.47655291295111    42.54277761541874    163.38265125819879
H    85.84540186570348    43.15171565721810    159.96029313299181
H    85.80823927501218    42.81995734577418    161.63898349558144
H    81.21852014001431    44.21898954575921    159.21954859397192
H    80.72547105099741    41.77189828706433    163.15219183861140
H    80.93341940346248    43.37582029669799    163.85962758244480
H    82.22655257955812    42.14581593559985    164.07750007845314
C    83.87895001300717    44.52654032284244    158.59544458263736
H    83.10280767257792    44.73509335511240    157.84849732966947
H    84.63362837303687    43.88424152223086    158.13122838888410
N    84.56790118343788    45.83962066513168    158.86182870318092
C    85.90018454486577    45.98758832363171    158.94059525439903
S    86.31173556620379    47.60478125152478    159.31933634639898
C    84.65381886506285    48.08114160199515    159.38108468803478
```

```
C    83.83929987951763    47.00228165071720    159.11855178890880
C    84.28836665013152    49.47685831795045    159.77289728230150
C    84.08122620176783    49.61871067497334    161.28750357355042
O    83.93023227307704    50.98385309915179    161.54450731826009
C    82.34928872673787    47.03147013409173    159.10567004141538
H    81.93263450155229    46.46885131116892    159.95336816620423
H    81.93396659052968    46.60411962839394    158.18054593816112
H    82.00108412890097    48.06799675603649    159.18315428945701
H    85.08251094463854    50.17886762181230    159.47431034484572
H    83.37764224731104    49.79875259034115    159.24535230712519
H    84.95733541147942    49.18102225463331    161.81049529405880
H    83.19781904936262    49.02077558057393    161.59305422091370
H    83.71412389838240    51.12111752853351    162.47571566406404
C    86.97150838210663    44.95486095847025    158.67903623136812
O    86.68251093632927    43.88152554822921    158.15909190661691
C    88.31598569428127    45.34289935975232    159.08062163004163
C    89.40188470504988    44.60567241936051    158.72161384772187
C    90.78052541096974    44.89404013427411    159.05753750058514
C    91.80824413524300    44.13537640984261    158.45094043418459
N    93.10188336662114    44.33194632968353    158.67657261362919
C    93.45263917490728    45.29500872919893    159.53247461405277
C    92.52476067761089    46.10829711227029    160.20357025488190
C    91.17406106402034    45.90657254063323    159.96130184095909
H    88.44402847639108    46.25357356880389    159.66773410864056
H    89.21914337528840    43.71887225314602    158.10457107328571
H    91.54472820076930    43.33625598770867    157.74591820045569
H    94.52635207133348    45.43378542598585    159.70523862940058
H    92.87031976314107    46.87621496216584    160.89881228779251
H    90.42955124804062    46.51723307698130    160.47613258477566
```

Bibliography

[1] G. F. Combs Jr, J. P. McClung, *The Vitamins: Fundamental Aspects in Nutrition and Health*, 5$^{\text{th}}$ Edition, Academic press, **2017**.

[2] L.-J. Baker, J. A. Dorocke, R. A. Harris, D. E. Timm, The Crystal Structure of Yeast Thiamin Pyrophosphokinase, *Structure* **2001**, *9*, 539–546.

[3] R. L. Schowen, Thiamin-Dependent Enzymes, *Comprehensive Biological Catalysis* **1998**, *2*, 217–266.

[4] R. Breslow, Rapid Deuterium Exchange in Thiazolium Salts, *J. Am. Chem. Soc.* **1957**, *79*, 1762–1763.

[5] R. Kluger, K. Tittmann, Thiamin Diphosphate Catalysis: Enzymic and Nonenzymic Covalent Intermediates, *Chem. Rev.* **2008**, *108*, 1797–1833.

[6] J. Swings, J. de Ley, The Biology of Zymomonas, *Bacteriol. Rev.* **1977**, *41*, 1–41.

[7] E. C. Heath, J. Hurwitz, B. L. Horecker, A. Ginsburg, Pentose Fermentation by Lactobacillus Plantarum: I. The Cleavage of Xylulose 5-Phosphate by Phosphoketolase, *J. Biol. Chem.* **1958**, *231*, 1001–1029.

[8] M. Schramm, V. Klybas, E. Racker, Phosphorolytic Cleavage of Fructose-6-phosphate by Fructose-6-phosphate Phosphoketolase from Acetobacter xylinum, *J. Biol. Chem.* **1958**, *233*, 1283–1288.

[9] G. Schenk, R. G. Duggleby, P. F. Nixon, Properties and Functions of the Thiamin Diphosphate Dependent Enzyme Transketolase, *Int. J. Biochem. Cell Biol.* **1998**, *30*, 1297–1318.

[10] A. Flechner, U. Dressen, P. Westhoff, K. Henze, C. Schnarrenberger, W. Martin, Molecular Characterization of Transketolase (EC 2.2.1.1) Active in the Calvin Cycle of Spinach Chloroplasts, *Plant Mol. Biol.* **1996**, *32*, 475–484.

[11] J. Yi, N. Nemeria, A. McNally, F. Jordan, R. S. Machado, J. R. Guest, Effect of Substitutions in the Thiamin Diphosphate-Magnesium Fold on the Activation of the Pyruvate Dehydrogenase Complex from *Escherichia coli* by Cofactors and Substrate, *J. Biol. Chem.* **1996**, *271*, 33192–33200.

[12] R. A. W. Frank, C. W. M. Kay, J. Hirst, B. F. Luisi, Off-Pathway, Oxygen-Dependent Thiamine Radical in the Krebs Cycle, *J. Am. Chem. Soc.* **2008**, *130*, 1662–1668.

[13] Y. Lindqvist, G. Schneider, Thiamin Diphosphate Dependent Enzymes: Transketolase, Pyruvate Oxidase and Pyruvate Decarboxylase, *Curr. Opin. Struct. Biol.* **1993**, *3*, 896–901.

[14] Y. A. Muller, Y. Lindqvist, W. Furey, G. E. Schulz, F. Jordan, G. Schneider, A Thiamin Diphosphate Binding Fold Revealed by Comparison of the Crystal Structures of Transketolase, Pyruvate Oxidase and Pyruvate Decarboxylase, *Structure* **1993**, *1*, 95–103.

[15] A. Kaplun, E. Binshtein, M. Vyazmensky, A. Steinmetz, Z. Barak, D. Chipman, K. Tittmann, B. Shaanan, Glyoxylate Carboligase Lacks the Canonical Active Site Glutamate of Thiamin-Dependent Enzymes, *Nature Chem. Biol.* **2008**, *4*, 113–118.

[16] D. Kern, G. Kern, H. Neef, K. Tittmann, G. Killenberg, M. Jabs, C. Wikner, G. Schneider, G. Hübner, How Thiamine Diphosphate is Activated in Enzymes, *Science* **1997**, *275*, 67–70.

[17] F. Jordan, N. Nemeria, S. Zhang, Y. Yan, P. Arjunan, W. Furey, Dual Catalytic Apparatus of the Thiamin Diphosphate Coenzyme: Acid-Base via the 1',4'-

Iminopyrimidine Tautomer along with Its Electrophilic Role, *J. Am. Chem. Soc.* **2003**, *125*, 12732–12738.

[18] D. Meyer, P. Neumann, R. Ficner, K. Tittmann, Observation of a Stable Carbene at the Active Site of a Thiamin Enzyme, *Nature Chem. Biol.* **2013**, *9*, 488–490.

[19] G. Schneider, Y. Lindqvist, Crystallography and Mutagenesis of Transketolase: Mechanistic Implications for Enzymatic Thiamin Catalysis, *Biochim. Biophys. Acta* **1998**, *1385*, 387–398.

[20] F. Jordan, Current Mechanistic Understanding of Thiamin Diphosphate-Dependent Enzymatic Reactions, *Nat. Prod. Res.* **2003**, *20*, 184–201.

[21] G. Wille, D. Meyer, A. Steinmetz, E. Hinze, R. Golbik, K. Tittmann, The Catalytic Cycle of a Thiamin Diphosphate Enzyme Examined by Cryocrystallography, *Nature Chem. Biol.* **2006**, *2*, 324–328.

[22] K. Tittmann, R. Golbik, K. Uhlemann, L. Khailova, G. Schneider, M. Patel, F. Jordan, D. M. Chipman, R. G. Duggleby, G. Hübner, NMR Analysis of Covalent Intermediates in Thiamin Diphosphate Enzymes, *Biochemistry* **2003**, *42*, 7885–7891.

[23] N. S. Nemeria, S. Chakraborty, A. Balakrishnan, F. Jordan, Reaction Mechanisms of Thiamin Diphosphate Enzymes: Defining States of Ionization and Tautomerization of the Cofactor at Individual Steps, *FEBS J.* **2009**, *276*, 2432–2446.

[24] H. Patel, N. S. Nemeria, F. H. Andrews, M. J. McLeish, F. Jordan, Identification of Charge Transfer Transitions Related to Thiamin-Bound Intermediates on Enzymes Provides a Plethora of Signatures Useful in Mechanistic Studies, *Biochemistry* **2014**, *53*, 2145–2152.

[25] M. Merski, C. A. Townsend, Observation of an Acryloyl-Thiamin Diphosphate Adduct in the First Step of Clavulanic Acid Biosynthesis, *J. Am. Chem. Soc.* **2007**, *129*, 15750–15751.

[26] S. Chakraborty, N. S. Nemeria, A. Balakrishnan, G. S. Brandt, M. M. Kneen, A. Yep, M. J. McLeish, G. L. Kenyon, G. A. Petsko, D. Ringe, F. Jordan, Detection

and Time Course of Formation of Major Thiamin Diphosphate-Bound Covalent Intermediates Derived from a Chromophoric Substrate Analogue on Benzoylformate Decarboxylase, *Biochemistry* **2009**, *48*, 981–994.

[27] F. Jordan, N. S. Nemeria, Progress in the Experimental Observation of Thiamin Diphosphate-Bound Intermediates on Enzymes and Mechanistic Information Derived from these Observations, *Bioorg. Chem.* **2014**, *57*, 251–262.

[28] S. R. Martin, P. M. Baley, Absorption and Circular Diochroism Spectroscopy in *Calcium-Binding Protein Protocols: Volume 2: Methods and Techniques. Methods in Molecular Biology* (Ed. H. J. Vogel), Vol. 173, , Springer, Totowa, NJ, **2002**, pp. 43–55.

[29] N. Nemeria, A. Baykal, E. Joseph, S. Zhang, Y. Yan, W. Furey, F. Jordan, Tetrahedral Intermediates in Thiamin Diphosphate-Dependent Decarboxylations Exist as a 1',4'-Imino Tautomeric Form of the Coenzyme, Unlike the Michaelis Complex or the Free Coenzyme, *Biochemistry* **2004**, *43*, 6565–6575.

[30] A. T. Baykal, L. Kakalis, F. Jordan, Electronic and Nuclear Magnetic Resonance Spectroscopic Features of the 1',4'-Iminopyrimidine Tautomeric Form of Thiamin Diphosphate, a Novel Intermediate on Enzymes Requiring this Coenzyme, *Biochemistry* **2006**, *45*, 7522–7528.

[31] M. Schreiber, M. R. Silva-Junior, S. P. A. Sauer, W. Thiel, Benchmarks for Electronically Excited States: CASPT2, CC2, CCSD, and CC3, *J. Chem. Phys.* **2008**, *128*, 134110.

[32] K. Sneskov, O. Christiansen, Excited State Coupled Cluster Methods, *WIREs Comput. Mol. Sci.* **2012**, *2*, 566–584.

[33] H. Stoll, The Correlation Energy of Crystalline Silicon, *Chem. Phys. Lett.* **1992**, *191*, 548–552.

[34] B. Paulus, P. Fulde, H. Stoll, Electron Correlations for Ground-State Properties of Group-IV Semiconductors, *Phys. Rev. B* **1995**, *51*, 10572–10578.

[35] K. Doll, M. Dolg, P. Fulde, H. Stoll, Correlation Effects in Ionic Crystals: The Cohesive Energy of MgO, *Phys. Rev. B* **1995**, *52*, 4842–4848.

[36] K. Doll, M. Dolg, P. Fulde, H. Stoll, Quantum Chemical Approach to Cohesive Properties of NiO, *Phys. Rev. B* **1997**, *56*, 10282–10288.

[37] K. Doll, H. Stoll, Cohesive Properties of Alkali Aalides, *Phys. Rev. B* **1997**, *56*, 10121–10127.

[38] R. A. Mata, H. Stoll, An Incremental Correlation Approach to Excited State Energies Based on Natural Transition/Localized Orbitals, *J. Chem. Phys.* **2011**, *134*, 034122.

[39] J. F. Stanton, R. J. Bartlett, The Equation of Motion Coupled-Cluster Method. A Systematic Biorthogonal Approach to Molecular Excitation Energies, Transition Probabilities, and Excited State Properties, *J. Chem. Phys.* **1993**, *98*, 7029–7039.

[40] K. Tittmann, internal communication.

[41] S. Wille, *QM/MM Studies of ThDP-Dependent Enzyme Catalysis*, Master's Thesis, Georg-August-University Göttingen, Germany, **2017**.

[42] F. Jensen, *Introduction to Computational Chemistry*, 2$^{\text{nd}}$ Edition, John Wiley & Sons, Ltd, **2007**.

[43] M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln, *Ann. d. Phys.* **1927**, *389*, 457–487.

[44] A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, 1$^{\text{st}}$ Edition, Dover Publications, Inc., **1996**.

[45] T. Helgaker, P. Jørgensen, J. Olsen, *Modern Electronic-Structure Theory*, 1$^{\text{st}}$ Edition, John Wiley & Sons, **2000**.

[46] J. C. Slater, The Theory of Complex Spectra, *Phys. Rev.* **1929**, *34*, 1293–1322.

[47] E. U. Condon, G. H. Shortley, *The Theory of Atomic Spectra*, Cambridge University Press, **1951**.

[48] A. Dreuw, M. Head-Gordon, Single Reference ab Initio Methods for the Calculation of Excited States of Large Molecules, *Chem. Rev.* **2005**, *105*, 4009–4037.

[49] J. Čížek, On the Correlation Problem in Atomic and Molecular Systems. Calcula-tion of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, *J. Chem. Phys.* **1966**, *45*, 4256–4266.

[50] G. D. Purvis III, R. J. Bartlett, A Full Coupled-Cluster Singles and Doubles Model: The Inclusion of Disconnected Triples., *J. Chem. Phys.* **1982**, *76*, 1910–1918.

[51] O. Christiansen, H. Koch, P. Jørgensen, The Second-Order Approximate Coupled Cluster Singles and Doubles Model CC2, *Chem. Phys. Lett.* **1995**, *243*, 409–418.

[52] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, A Fifth-Order Per-turbation Comparison of Electron Correlation Theories, *Chem. Phys. Lett.* **1989**, *157*, 479–483.

[53] J. Schirmer, F. Mertins, Review of Biorthogonal Coupled Cluster Representations for Electronic Excitation, *Theor. Chem. Acc.* **2010**, *125*, 145–172.

[54] H. J. Monkhorst, Calculation of Properties with the Coupled-Cluster Method, *Int. J. Quantum Chem.* **1977**, *12*, 421–432.

[55] J. Olsen, P. Jørgensen, Linear and Nonlinear Response Functions for an Exact State and For an MCSCF State, *J. Chem. Phys.* **1985**, *82*, 3235–3264.

[56] H. Koch, P. Jørgensen, Coupled Cluster Response Functions, *J. Chem. Phys.* **1990**, *93*, 3333–3344.

[57] R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, 1$^{\text{st}}$ Edition, Oxford University Press, **1989**.

[58] P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, *Phys. Rev.* **1964**, *136*, B864.

[59] J. Xia, C. Huang, I. Shin, E. Carter, Can Orbital-Free Density Functional Theory Simulate Molecules?, *J. Chem. Phys.* **2012**, *136*, 084102.

[60] W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Corre-lation Effects, *Phys. Rev.* **1965**, *140*, A1133.

[61] A. D. Becke, Densty-Functional Thermochemistry. III. The Role of Exact Exchange, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[62] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fiels, *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[63] A. D. Becke, Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior, *Phys. Rev. A* **1988**, *38*, 3098–3100.

[64] C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density, *Phys. Rev. B* **1988**, *37*, 785–789.

[65] T. Yanai, D. T. Tew, N. C. Handy, A New Hybrid Exchange-Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP), *Chem. Phys. Lett.* **2004**, *393*, 51–57.

[66] W. Hieringer, A. Görling, Failure of Time-Dependent Density Functional Methods for Excitations in Spatially Separated Systems, *Chem. Phys. Lett.* **2006**, *419*, 557–562.

[67] M. J. G. Peach, P. Benfield, T. Helgaker, D. J. Tozer, Excitation Energies in Density Functional Theory: An Evaluation and a Diagnostic Test, *J. Chem. Phys.* **2008**, *128*, 044118.

[68] C. A. Ullrich, *Time-Dependent Density-Functional Theory: Concepts and Applications*, 1st Edition, Oxford University Press, **2012**.

[69] E. Runge, E. K. M. Gross, Density-Functional Theory for Time-Dependent Systems, *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

[70] R. van Leeuwen, Mapping from Densities to Potentials in Time-Dependent Density-Functional Theory, *Phys. Rev. Lett.* **1999**, *82*, 3863–3866.

[71] M. E. Casida, *Recent Developments and Applications of Modern Density Functional Theory* (ed. D. E. Chong), World Scientific, Vol. 1, **1995**, pp. 155–192.

[72] S. Hirata, M. Head-Gordon, Time-dependent Density Functional Theory within the Tamm–Dancoff Approximation, *Chem. Phys. Lett.* **1999**, *314*, 291–299.

[73] D. Jacquemin, V. Wathelet, E. A. Perpete, C. Adamo, Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules, *J. Chem. Theory Comput.* **2009**, *5*, 2420–2435.

[74] S. S. Leang, F. Zaheriev, M. S. Gordon, Benchmarking the Performance of Time-Dependent Density Functional Methods, *J. Chem. Phys.* **2012**, *136*, 104101.

[75] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd Edition, Pearson Education Limited, **2001**.

[76] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatman III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules, *Comput. Phys. Commun.* **1995**, *91*, 1–41.

[77] P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann. d. Phys.* **1921**, *64*, 253–287.

[78] A. Warshel, M. Levitt, Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme, *J. Mol. Biol.* **1976**, *103*, 227–249.

[79] H. M. Senn, W. Thiel, QM/MM Methods for Biomolecular Systems, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[80] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer, C. Lennartz, QUASI: A General Purpose Implementation of the QM/MM Approach and its Application to Problems in Catalysis, *J. Mol. Struct. THEOCHEM* **2003**, *632*, 1–28.

[81] M. S. Gordon ed., *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*, 1st Edition, John Wiley & Sons, Ltd, **2017**.

[82] R. L. Martin, Natural Transition Orbitals, *J. Chem. Phys.* **2003**, *118*, 4775–4777.

[83] J. Pipek, P. G. Mezey, A Fast Intrinsic Localization Procedure Applicable for ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions, *J. Chem. Phys.* **1989**, *90*, 4916–4926.

[84] W. Kohn, Density Functional Theory for Systems of Very Many Atoms, *Int. J. Quantum Chem.* **1995**, *56*, 229–232.

[85] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-Type Basis Set for Second-Row Elements, *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[86] T. H. Dunning, Jr., Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen, *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[87] R. A. Kendall, T. H. Dunning, Jr., R. J. Harrison, Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions, *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[88] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, Molpro: A General Purpose Quantum Chemistry Program Package, *WIREs Comput. Mol. Sci.* **2012**, *2*, 242–253.

[89] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, MOLPRO, version 2015.1, a package of ab initio programs, **2015**, see http://www.molpro.net.

[90] T. Yoshikawa, M. Kobayashi, A. Fujii, H. Nakai, Novel Approach to Excited-State Calculations of Large Molecules Based on Divide-and-Conquer Method: Application to Photoactive Yellow Protein, *J. Phys. Chem. B* **2013**, *117*, 5565–5573.

[91] S. Saebø, P. Pulay, Local Treatment of Electron Correlation, *Ann. Rev. Phys. Chem.* **1993**, *44*, 213–236.

[92] C. Hampel, H.-J. Werner, Local Treatment of Electron Correlation in Coupled Cluster Theory, *J. Chem. Phys.* **1996**, *104*, 6286–6297.

[93] M. Schütz, H.-J. Werner, Low-Order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD), *J. Chem. Phys.* **2001**, *114*, 661–681.

[94] T. Korona, H.-J. Werner, Local Treatment of Electron Excitations in the EOM-CCSD Method, *J. Chem. Phys.* **2003**, *118*, 3006–3019.

[95] W. Kutzelnigg, Die Lösung des quantenmechanischen Zwei-Elektronenproblems durch unmittelbare Bestimmung der natürlichen Einelektronenfunktionen, *Theor. Chim. Acta* **1963**, *1*, 327–342.

[96] C. Edminston, M. Krauss, Pseudonatural Orbitals as a Basis for the Superposition of Configurations. I. He_2^+, *J. Chem. Phys.* **1966**, *45*, 1833–1839.

[97] W. Meyer, Ionization Energies of Water from PNO-CI Calculations, *Int. J. Quantum Chem.* **1971**, *5*, 341–348.

[98] F. Neese, F. Wennmohs, A. Hansen, Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method, *J. Chem. Phys.* **2009**, *130*, 114108.

[99] B. Helmich, C. Hättig, Local Pair Natural Orbitals for Excited States, *J. Chem. Phys.* **2011**, *135*, 214106.

[100] B. Helmich, C-Hättig, A Pair Natural Orbital Implementation of the Coupled Cluster Model CC2 for Excitation Energies, *J. Chem. Phys.* **2013**, *139*, 084114.

[101] M. S. Frank, C. Hättig, A Pair Natural Orbital Implementation of CCSD Excitation Energies within the Framework of Linear Response Theory, *J. Chem. Phys.* **2018**, *148*, 134102.

[102] M. Head-Gordon, R. J. Rico, M. Oumi, T. J. Lee, A Doubles Correction to Electronic Excited States from Configuration Interaction in the Space of Single Substitutions, *Chem. Phys. Lett.* **1994**, *219*, 21–29.

[103] X.-Y. Pei, K. M. Erixon, B. F. Luisi, F. J. Leeper, Structural Insights into the Prereaction State of Pyruvate Decarboxylase from Zymomonas Mobilis, *Biochemistry* **2010**, *49*, 1727–1736.

[104] D. Meyer, P. Neumann, C. Parthier, R. Friedemann, N. Nemeria, F. Jordan, K. Tittmann, Double Duty for a Conserved Glutamate in Pyruvate Decarboxylase: Evidence of the Participation in Stereoelectronically Controlled Decarboxylation and in Protonation of the Nascent Carbanion/Enamine Intermediate, *Biochemistry* **2010**, *49*, 8197–8212.

[105] S. Lüdtke, P. Neumann, K. M. Erixon, F. Leeper, R. Kluger, R. Ficner, K. Tittmann, Sub-Ångström Resolution Crystallography Reveals Physical Distortions that Enhance Reactivity of a Covalent Enzymatic Intermediate, *Nature Chem.* **2013**, *5*, 762–767.

[106] I. N. Levine, *Quantum Chemistry*, 6th Edition, Pearson Prentice Hall, **2009**.

[107] M. E. Madjet, A. Abdurahman, T. Renger, Intermolecular Coulomb Coupling from ab Initio Electrostatic Potentials: Application to Optical Transitions of Strongly Coupled Pigments in Photosynthetic Antennae and Reaction Centers, *J. Phys. Chem. B* **2006**, *110*, 17268–17281.

[108] J. Adolphs, F. Müh, M. E. Madjet, T. Renger, Calculation of Pigment Transition Energies in the FMO Protein, *Photosynth. Res.* **2008**, *95*, 197–209.

[109] J. Adolphs, F. Müh, M. E. Madjet, M. Schmidt am Busch, T. Renger, Structure-Based Calculations of Optical Spectra of Photosystem I Suggest an Asymmetric Light-Harvesting Process, *J. Am. Chem. Soc.* **2010**, *132*, 3331–3343.

[110] T. Renger, F. Müh, Understanding Photosynthetic Light-Harvesting: A Bottom up Theoretical Approach, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3348–3371.

[111] S. Grimme, S. Ehrlich, H. Krieg, A Consistent and Accurate ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, *J. Chem. Phys.* **2010**, *132*, 154104.

[112] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[113] F. Weigend, R. Ahlrichs, Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[114] J. Almlöf, O. Vahtras, M. W. Feyereisen, Integral Approximations for LCAO-SCF Calculations, *Chem. Phys. Lett.* **1993**, *213*, 514–518.

[115] F. Weigend, Accurate Coulomb-Fitting Basis Sets for H to Rn, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[116] F. Neese, F. Wennmohs, A. Hansen, U. Becker, Efficient, Approximate and Parallel Hartree-Fock and Hybrid DFT Calculations. A 'chain-of-spheres' Algorithm for the Hartree-Fock Exchange, *Chem. Phys.* **2009**, *356*, 98–109.

[117] A. Klamt, G. Schüürmann, COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient, *J. Chem. Soc. Perkin Trans 2* **1993**, (5), 799–805.

[118] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters, *Proteins* **2006**, *65*, 712–725.

[119] N. Nemeria, S. Chakraborty, A. Baykal, L. G. Korotchkina, M. S. Patel, F. Jordan, The 1',4'-Iminopyrimidine Tautomer of Thiamin Diphosphate is Poised for Catalysis in Asymmetric Active Centers on Enzymes, *Proc. Natl. Acad. Sci.* **2007**, *104*, 78–82.

[120] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and Testing of a General Amber Force Field, *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[121] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Comput. Phys.* **1977**, *23*, 327–341.

[122] S. Miyamoto, P. A. Kollman, SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models, *J. Comput. Chem.* **1992**, *13*, 952–962.

[123] F. Neese, The ORCA Program System, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

[124] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarova, A. Kovalenko, P. A. Kollman, AMBER 11, **2010**.

[125] ChemShell, a Computational Chemistry Shell, see `www.chemshell.org`.

[126] Dalton, a Molecular Electronic Structure Program, Release DALTON2013.0, **2013**, see `http://daltonprogram.org/`.

[127] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansik, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. Rybkin, P. Salek, C. C. M. Samson, A. Sánchez de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, H. Ågren, The Dalton Quantum Chemistry Program System, *WIREs Comput. Mol. Sci.* **2013**, *4*, 269–284.

[128] Y. Zhao, D. G. Truhlar, The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 other Functionals, *Theor. Chem. Acc.* **2008**, *1202*, 215–241.

[129] S. P. de Visser, M. G. Quesne, B. Martin, P. Comba, U. Ryde, Computational Modelling of Oxygenation Processes in Enzymes and Biomimetic Model Complexes, *Chem. Commun.* **2014**, *50*, 262–282.

[130] K. J. Gruys, C. J. Halkides, P. A. Frey, Synthesis and Properties of 2-Acetylthiamin Pyrophosphate: An Enzymatic Reaction Intermediate, *Biochemistry* **1987**, *26*, 7575–7585.

[131] W. Shin, J. Plecher, G. Blank, M. Sax, Ring Stacking Interactions between Thiamin and Planar Molecules As Seen in the Crystal Structure of a Thiamin Picrolonate Dihydrate Complex, *J. Am. Chem. Soc.* **1977**, *99*, 3491–3499.

[132] C. Adamo, V. Barone, Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model, *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[133] G. Pescitelli, N. Sreerama, P. Salvadori, K. Nakanishi, N. Berova, R. W. Woody, Inherent Chirality Dominates the Visible/Near-Ultraviolet CD Spectrum of Rhodopsin, *J. Am. Chem. Soc.* **2008**, *130*, 6170–6181.

[134] C. Kiefl, N. Sreerama, R. Haddad, L. Sun, W. Jentzen, Y. Lu, Y. Qiu, J. A. Shelnutt, R. W. Woody, Heme Distortions in Sperm-Whale Carbonmonoxy Myoglobin: Correlations between Rotational Strengths and Heme Distortions in MD-Generated Structures, *J. Am. Chem. Soc.* **2002**, *124*, 3385–3394.

[135] T. Zaunmüller, M. Eichert, H. Richter, G. Unden, Variations in the Energy Metabolism of Biotechnologically Relevant Heterofermentative Lactic Acid Bacteria During Growth on Sugars and Organic Acids, *Appl. Microbiol. Biotechnol.* **2006**, *72*, 421–429.

[136] A. Yevenes, P. A. Frey, Cloning, Expression, Purification, Cofactor Requirements, and Steady State Kinetics of Phosphoketolase-2 from Lactobacillus plantarum, *Bioorg. Chem.* **2008**, *36*, 121–127.

[137] A. Dreuw, J. L. Weismann, M. Head-Gordon, Long-Range Charge-Transfer Excited States in Time-Dependent Density Functional Theory Require Non-Local Exchange, *J. Chem. Phys.* **2003**, *119*, 2943–2946.

[138] J. Zhang, Y. Liu, Computational Studies on the Catalytic Mechanism of Phosphoketolase, *Comput. Theor. Chem.* **2013**, *1025*, 1–7.

[139] H. Eyring, The Activated Complex in Chemical Reactions, *J. Chem. Phys.* **1935**, *3*, 107–115.

[140] K. Tittmann, G. Wille, X-Ray Crystallographic Snapshots of Reaction Intermediates in Pyruvate Oxidase and Transketolase Illustrate Common Themes in Thiamin Catalysis, *J. Mol. Catal. B: Enzym.* **2009**, *61*, 93–99.

[141] P. Asztalos, C. Parthier, R. Golbik, M. Kleinschmidt, G. H. M. S. Weiss, R. Friedemann, G. Wille, K. Tittmann, Strain and Near Attack Conformers in Enzymic Thiamin Catalysis: X-ray Crystallographic Snapshots of Bacterial Transketolase in Covalent Complex with Donor Ketoses Xylulose 5-Phosphate and Fructose 6-Phosphate, and in Noncovalent Complex with Acceptor Aldose Ribose 5-Phosphate, *Biochemistry* **2007**, *46*, 12037–12052.

[142] E. Fiedler, S. Thorell, T. Sandalova, R. Golbik, S. König, G. Schneider, Snapshot of a Key Intermediate in Enzymatic Thiamin Catalysis: Crystal Structure of the α-carbanion of (α,β-Dihydroxyethyl)-Thiamin Diphosphate in the Active Site of Transketolase from *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci.* **2002**, *99*, 591–595.

[143] C. Wikner, L. Meshalkina, U. Nilsson, Y. Lindqvist, M. Sundström, G. Schneider, Analysis of an Invariant Cofactor-Protein Interaction in Thiamin Diphosphate-Dependent Enzymes by Site-Directed Mutagenesis. Glutamic Acid 418 in Transketolase is Essential for Catalysis, *J. Biol. Chem.* **1994**, *269*, 32144–32150.

[144] C. Wikner, U. Nilsson, L. Meshalkina, C. Udekwu, Y. Lindqvist, G. Schneider, Identification of Catalytically Important Residues in Yeast Transketolase, *Biochemistry* **1997**, *36*, 15643–15649.

[145] E. Fiedler, R. Golbik, G. Schneider, K. Tittmann, H. Neef, S. König, G. Hübner, Examination of Donor Substrate Conversion in Yeast Transketolase, *J. Biol. Chem.* **2001**, *276*, 16051–16058.

[146] P. Botschwina, Vibrational Frequencies from Anharmonic Ab Initio/Empirical Potential Energy Functions. III Stretching Vibrations of Hydrogen Cyanide and Acetylenes, *Chem. Phys* **1982**, *68*, 41–63.

[147] C. L. Perrin, J. B. Nielson, "Strong" Hydrogen Bonds in Chemistry and Biology, *Annu. Rev. Phys. Chem.* **1997**, *48*, 511–544.

[148] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, Atomic-Level Characterization of the Structural Dynamics of Proteins, *Science* **2010**, *330*, 341–346.

[149] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, AAAI, Chapter A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, **1996**, pp. 226–231.

[150] F. Plasser, H. Lischka, Analysis of Excitonic and Charge Transfer Interactions from Quantum Chemical Calculations, *J. Chem. Theory Comput.* **2012**, *8*, 2777–2789.

[151] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, V. Barone, Geometries and Properties of Excited States in the Gas Phase and in Solution: Theory and Application of a Time-Dependent Density Functional Theory Polarizable Continuum Model, *J. Chem. Phys.* **2006**, *124*, 094107.

[152] F. Jordan, Z. H. Kudzin, C. B. Rios, Generation and Physical-Properties of Enamines Related to the Key Intermediate in Thiamin Diphosphate Dependent Enzymeatic Pathways, *J. Am. Chem. Soc.* **1987**, *109*, 4415–4416.