

# Computation of Electronic Excitation Spectra of Large Biomolecular Systems

Johann Jakob Seibert

Dissertation





# Computation of Electronic Excitation Spectra of Large Biomolecular Systems

Dissertation  
zur  
Erlangung des Doktorgrades (Dr. rer. nat.)  
der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der  
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von  
**Johann Jakob Seibert**  
aus Brunsbüttel

–2020–



---

Dekan:	Prof. Dr. Johannes Beck
Erster Gutachter:	Prof. Dr. Stefan Grimme
Zweiter Gutachter:	Prof. Dr. Thomas Bredow
Fachnaher Gutachter:	Prof. Dr. Sigurd Höger
Fachfremder Gutachter:	Prof. Dr. Rainer Manthey
Erscheinungsjahr:	2020
Tag der Disputation:	08.10.2020

---

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn







# Statement of Authorship

I, Johann Jakob Seibert, hereby declare that I am the sole author of this dissertation. The ideas and work of others, whether published or unpublished, have been fully acknowledged and referenced.

Mettmann, October 29, 2020





# Publications

Parts of this thesis have been published in peer-reviewed journals:

1. J. Seibert, C. Bannwarth and S. Grimme “Biomolecular structure information from high-speed quantum mechanical electronic spectra calculation”, *J. Am. Chem. Soc.* **2017**, *139*, 11682.
2. J. Seibert, J. Pisarek, S. Schmitz, C. Bannwarth and S. Grimme “Extension of the element parameter set for ultra-fast excitation spectra calculation (sTDA-xTB)”, *Mol. Phys.* **2019**, *150*, 1104.
3. J. Seibert, B. Champagne, S. Grimme and M. de Wergifosse “Dynamic structural effects on the second-harmonic generation of short tryptophan-rich peptides and gramicidin A”, *J. Phys. Chem. B* **2020**, *124*, 2568.

Further publications:

1. C. Bannwarth, S. Grimme and J. Seibert "Electronic Circular Dichroism of [16]Helicene With Simplified TD-DFT: Beyond the Single Structure Approach", *Chirality* **2016**, *28*, 365.
2. S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, F. Neese, "Fully Automated Quantum-Chemistry-Based Computation of Spin-Spin-Coupled Nuclear Magnetic Resonance Spectra", *Angew. Chem. Int. Ed.* **2017**, *56*, 14763.
3. J. D. Queen, M. Bursch, J. Seibert, L. R. Maurer, B. D. Ellis, J. C. Fettinger, S. Grimme and P. P. Power "Isolation and Computational Studies of a Series of Terphenyl Substituted Diplumbynes with Ligand Dependent Lead-Lead Multiple-Bonding Character", *J. Am. Chem. Soc.* **2019**, *141*, 14370.
4. M. de Wergifosse, J. Seibert, B. Champagne and S. Grimme "Are Fully-Conjugated Expanded Indenofluorenes Analogues and Diindeno[n]thiophene Derivatives Diradicals? A Simplified (Spin-Flip) Time-Dependent Density Functional Theory [(SF-)sTD-DFT] Study", *J. Phys. Chem. A* **2019**, *123*, 9828.
5. S. Schmitz, K. Ostermeir, A. Hansen, A. H. Göller, J. Seibert and S. Grimme "Quantum chemical calculation of molecular and periodic peptide and protein structures", *J. Phys. Chem. B* **2020**, DOI:10.1021/acs.jpcc.0c00549.

Presentations:

- Poster: "Simplified linear response treatment for the electronic spectra of huge molecular systems", *Symposium for Theoretical Chemistry*, **2016**, Bochum, Germany.
- Poster: "Electronic excitation spectra calculation of biomolecular systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *Excited States Simulations: Bridging Scales workshop*, **2016**, Marseille, France.
- Poster: "Electronic excitation spectra calculation of biomolecular systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *Computational Molecular Science*, **2017**, Warwick, United Kingdom.
- Talk: "xTB – A robust and versatile tight-binding package", *Theoretical Chemistry Colloquia*, **2017**, Bochum, Germany.

- Poster: "Thermochromism of Ferrocene – a TD-DFT study", *International Conference on Density-Functional Theory and its Applications*, **2017**, Tällberg, Sweden.
- Poster: "Electronic excitation spectra calculation of large systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *World Association of Theoretical and Computational Chemists*, **2017**, Munich, Germany.
- Poster: "Electronic excitation spectra calculation of large systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *European Summerschool of Quantum Chemistry*, **2017**, Palermo, Italy.
- Talk: "Electronic excitation spectra calculation of large systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *ACS Meeting*, **2018**, New Orleans, USA.
- Poster: "Electronic excitation spectra calculation of large systems with a tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)", *Winter School in Theoretical Chemistry: Machine Learning*, **2018**, Helsinki, Finland.
- Talk: "From circular dichroism for proteins to non-linear optical properties – an update on sTD(A)-xTB", *Theoretical Chemistry Colloquia*, **2019**, Mühlheim an der Ruhr, Germany.
- Talk: "Structure-property relationship for the first hyperpolarizability of short tryptophan-rich peptides", **2019**, Namur, Belgium.



# Abstract

This thesis presents the theoretical investigation of electronic excitation spectra and nonlinear responses of large biomolecular systems (e.g., proteins or deoxyribonucleic acid fragments), with a particular focus on the simulation of (electronic) circular dichroism (CD) spectra and first hyperpolarizabilities.

In the past, a fully quantum mechanical (QM) treatment of biomolecular structures for optical property calculations was computationally unfeasible. Due to their sizes and complexity, simulations on biomolecules were limited to calculations of subsystems in fragmentation approaches or with some system-specific adjustments. Modern developments regarding ultra-fast excited states calculations enable to characterize systems beyond the nowadays limit of thousand atoms. One prominent development in this regard is the extended tight-binding (xTB) based method combined with the simplified time-dependent density functional theory (sTD-DFT), including both flavors: sTDA-xTB and sTD-DFT-xTB. The applicability or transferability of these methods to systems up to several thousand atoms is by far non-trivial and therefore thoroughly investigated in this work.

For this purpose, the first part of this thesis describes a comprehensive benchmarking on the sTDA-xTB method, regarding the computation of electronic excitation spectra of large biomolecules. This is the first time that relevant biomolecular systems are treated within a full QM approach for the calculation of the excited state properties. Therefore, it is especially remarkable that the computed CD spectra are in excellent agreement with the experiment. The highly efficient sTDA-xTB also enables the computation of spectra averaged along structures from molecular dynamics (MD) simulations, which are often needed to cover non-equilibrium structure and conformational effects. The investigations confirm also the all-round applicability of the sTDA-xTB method, since almost no limitation to a certain spectral range or type of chromophore is observed.

The original publication of the sTDA-xTB method only provided a set of parameters for the most important elements. Some metals, naturally occurring in proteins, were missing from this set. However, such metal-containing proteins play essential roles in many biological mechanisms and are of great interest for spectroscopic studies. This parameterization gap is closed in this thesis, and the missing parameters for the 4d and 5d metals, as well as for the 4p, 5p, and 6p element blocks, are obtained. Comparisons to theory and experiment show that

## *Abstract*

sTDA-xTB provides similar good results as for the elements in the original publication with an average deviation of excitation energies of 0.3–0.5 eV.

For the simulation of electronic excitation spectra, reasonable three-dimensional structures are as important as the calculation of the excited states itself. Therefore, it is of particular interest to apply an accurate and efficient structure method in tandem with the excited state method. The xTB variants for geometries, frequencies, and non-covalent interactions (GFN) are ideally suited in this regard and comprehensively tested in combination with the sTDA-xTB method in this work. Just as CD spectroscopy, second-harmonic imaging microscopy (SHIM) is widely used in medical research. The underlying physical effect for this imaging technique is the second-harmonic generation phenomenon, a scattering process in which the optical frequency of incident photons is doubled. Theoretical models can access this non-linear optical (NLO) phenomenon via the first hyperpolarizability. The second part of this thesis deals with the application of the sTD-DFT-xTB method to relevant biomolecular systems, i.e., a set of tryptophan-rich oligopeptides for SHIM applications. For the first time, a structure-property analysis of the first hyperpolarizability can be conducted, because of the recent development of efficient conformational sampling and a computationally efficient implementation to evaluate first hyperpolarizabilities at the sTD-DFT level of theory. Furthermore, the comparison to commonly-used higher levels of theory shows that the sTD-DFT-xTB method is capable of providing equally reliable second-harmonic generation values at  $10^{-3}$ – $10^{-5}$  of the higher levels computational cost.

In summary, for the first time a full QM treatment of the simulation of CD spectra and NLO properties of large biomolecular systems is reported. For this purpose, all parts of an efficient computational workflow are comprehensively tested and successfully elaborated. An automatic black-box approach for the calculation of optical properties of large biomolecules is desirable and the findings of this work show that xTB-based methods are excellent candidates in this regard.

# Contents

<b>Publications</b>	<b>ix</b>
<b>Abstract</b>	<b>xiii</b>
<b>I. Introduction and Theoretical Background</b>	<b>1</b>
<b>1. Introduction</b>	<b>3</b>
1.1. Electronic Circular Dichroism (CD) . . . . .	4
1.2. Second harmonic imaging technique (SHIM) . . . . .	7
1.3. Objectives . . . . .	8
<b>2. Theoretical Background</b>	<b>11</b>
2.1. Methodological Overview . . . . .	11
2.2. Wave function theory (WFT) . . . . .	12
2.3. Density functional theory (DFT) . . . . .	15
2.4. Linear Response Theory . . . . .	18
2.5. Extended tight-binding (xTB) . . . . .	19
<b>II. Development and Application of the sTDA-xTB method</b>	<b>23</b>
<b>3. Biomolecular structure information from high-speed quantum mechanical electronic spectra calculation</b>	<b>27</b>
3.1. Introduction . . . . .	28
3.2. Results and Discussion . . . . .	29
3.3. Conclusion . . . . .	33
3.4. Technical Details . . . . .	34
<b>4. Extension of the element parameter set for ultra-fast excitation spectra calculation (sTDA-xTB)</b>	<b>37</b>
4.1. Introduction . . . . .	38
4.2. Theory . . . . .	39
4.2.1. The simplified Tamm-Dancoff approximation to TD-DFT .	39
4.2.2. The extended tight-binding composite scheme sTDA-xTB	41

## Contents

4.2.3. Extended AO basis set . . . . .	42
4.3. Parameterization and technical details . . . . .	43
4.3.1. Reference data collection . . . . .	43
4.3.2. Parameterization . . . . .	44
4.4. Results and discussion . . . . .	46
4.4.1. Charges and bond orders from the VTB Hamiltonian . . . . .	46
4.4.2. Excitation energies for the fit sets . . . . .	47
4.4.3. UV and CD spectra in comparison to theory . . . . .	48
4.4.4. UV and CD spectra in comparison to experiment . . . . .	51
4.5. Conclusions . . . . .	58
<b>5. Benchmarking of structural approaches for an efficient quantum-mechanical protein spectra calculation</b>	<b>61</b>
5.1. Introduction . . . . .	61
5.2. Computational Details . . . . .	62
5.3. Results and Discussion . . . . .	64
5.3.1. Model peptides . . . . .	64
5.3.2. Proteins . . . . .	68
5.4. Conclusion . . . . .	74
<b>III. Application of the sTD-DFT-xTB method</b>	<b>77</b>
<b>6. Dynamic structural effects on the second-harmonic generation of short tryptophan-rich peptides and gramicidin A</b>	<b>81</b>
6.1. Introduction . . . . .	82
6.2. Computational Details . . . . .	84
6.3. Results and Discussion . . . . .	89
6.3.1. Conformers . . . . .	89
6.3.2. Molecular dynamics sampling . . . . .	95
6.3.3. Comparison with extrapolated experimental values . . . . .	97
6.4. Conclusion . . . . .	99
<b>IV. Final Summary and Conclusion</b>	<b>103</b>
<b>Bibliography</b>	<b>111</b>
<b>V. Appendix</b>	<b>125</b>
<b>A1.Supporting Information to Chapter 3</b>	<b>126</b>
A1.1.Computational details . . . . .	126



A1.2. Rotatory strengths . . . . .	127
A1.3. Photoactive yellow protein . . . . .	128
A1.4. DNA . . . . .	131
A1.5. 1L2Y Protein . . . . .	131
<b>A2. Supporting Information to Chapter 4</b>	<b>133</b>
A2.1. Additional technical details . . . . .	133
A2.2. Parameters . . . . .	135
A2.2.1. VTB: element-specific parameters . . . . .	135
A2.2.2. XTB: element-specific parameters . . . . .	137
<b>A3. Supporting Information to Chapter 6</b>	<b>152</b>
A3.1. TDHF computed first hyperpolarizabilities . . . . .	152
A3.2. Unit-sphere representation . . . . .	153
A3.3. Molecular dipole and $\beta$ vectors . . . . .	155
A3.4. Molecular dynamics sampling . . . . .	157
A3.5. Comparison of sTD-DFT-xTB with TDHF . . . . .	158
A3.6. Quantitative comparison with experimental values . . . . .	159
<b>D. Acronyms</b>	<b>161</b>



**Part I.**

**Introduction and Theoretical  
Background**



# 1. Introduction

Theoretical models have become an essential part in natural sciences, because of the development of efficient theoretical methods and the rise of computational resources.<sup>1</sup> Theoretical methods – e.g., based on quantum mechanical (QM) theory – are utilized in natural sciences in tandem with the experiment to justify experimental findings, i.e., to resolve the reactivity or spectroscopic properties of molecules.<sup>2–4</sup> Furthermore, theoretical models are employed for pure prediction purposes.<sup>2,4–8</sup>

The three-dimensional structure, or geometry, of molecules is directly connected to the chemical and physical properties of the respective compound.<sup>9</sup> Therefore, the elucidation of this molecular geometry remains an important task. In experiments, X-ray crystallography of molecular crystals is commonly used to determine the geometry of, e.g., organic and biomolecular compounds. However, the crystallization of compounds is often challenging, and additionally, the geometrical conformation in the solid and the solution or gas phase can diverge drastically.<sup>10</sup> Thus, a different approach for the determination of the molecular structure must be applied.

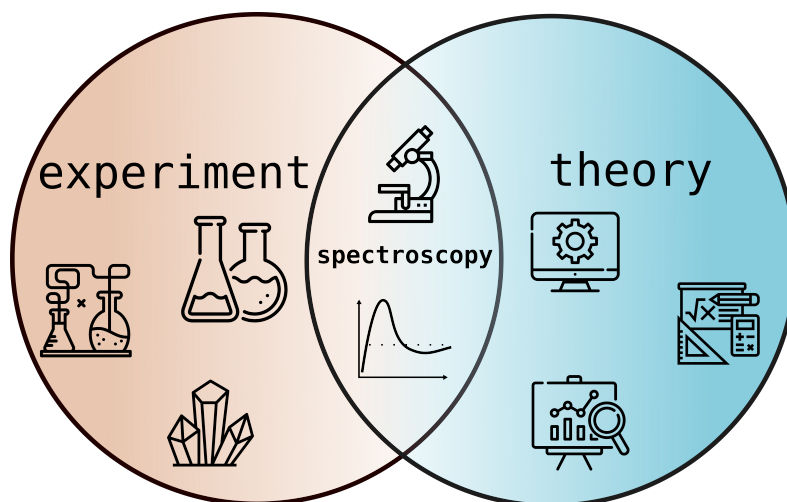


Figure 1.1.: Schematic depiction of the overlap of theoretical models with the experiment.

The combination of experimental measurements and theoretical simulations of chemical systems in solution is one strategy that relies on a connectable metric

## 1. Introduction

between the simulation and the experiments. It must also be possible to evaluate the accuracy of the applied level of theory for the reproduction of data obtained from experiment. In particular, this approach aims at reproducing the experimental conditions in the simulations as close as possible and thus proposing reasonable structures. Spectroscopy is one viable connection between the simulation and the experiment (cf. Figure 1.1). The combined approach of theory and experiment comprises the following steps: first, a molecular structure of the desired system is modeled based on information from the experiment. Second, from this structure, multiple spectroscopic properties are computed, and compared to the measurements from the experiment. Third, if the data does not match, the theoretical model structure is revised. This iterative procedure can reveal structures in solution, which are otherwise inaccessible in a direct way for experimental techniques.

### 1.1. Electronic Circular Dichroism (CD)

Circular dichroism (CD) is a spectroscopic method that is especially susceptible to the three-dimensional arrangement of a compound and thus very well suited for a property, linking theory and experiment.<sup>11,12</sup> In CD spectroscopy, a probe is irradiated by circularly polarized ultraviolet and visible (UV-Vis) light and the absorbance is measured. Then, the difference between the absorbance ( $\epsilon$ ) of left- and right circularly polarized light is calculated by:

$$\Delta\epsilon = \epsilon_L - \epsilon_R$$

A difference in absorbance ( $\Delta\epsilon$ ) is observed if a molecule cannot be superposed on its mirror image by any combination of rotations and translations. Molecules with this property are called chiral and their enantiomers yield inverted spectra. Therefore, CD spectroscopy is widely used for the assignment of the absolute configuration of molecules.<sup>11–18</sup>

Another very prominent application of CD is the structural characterization of polypeptides or proteins. Local excitations in the peptide bonds dominate the absorption in the far ultraviolet (UV) region (160–240 nm), and the respective CD is highly receptive to their configuration, i.e., the protein secondary structure.<sup>19–22</sup> The secondary structure is defined as the three-dimensional arrangement of the peptide backbone. The most naturally occurring secondary structures are  $\alpha$ -helices and  $\beta$ -sheets.<sup>23</sup> The different secondary structure motifs have characteristic CD signals (cf. Figure 1.2), allowing for *in vitro* studies of protein folding.<sup>24–26</sup> Moreover, mutations of a wild type protein – which results in changes in the tertiary structure – alter the CD spectrum in the near UV region (250–300 nm).<sup>27</sup>

## 1.1. Electronic Circular Dichroism (CD)

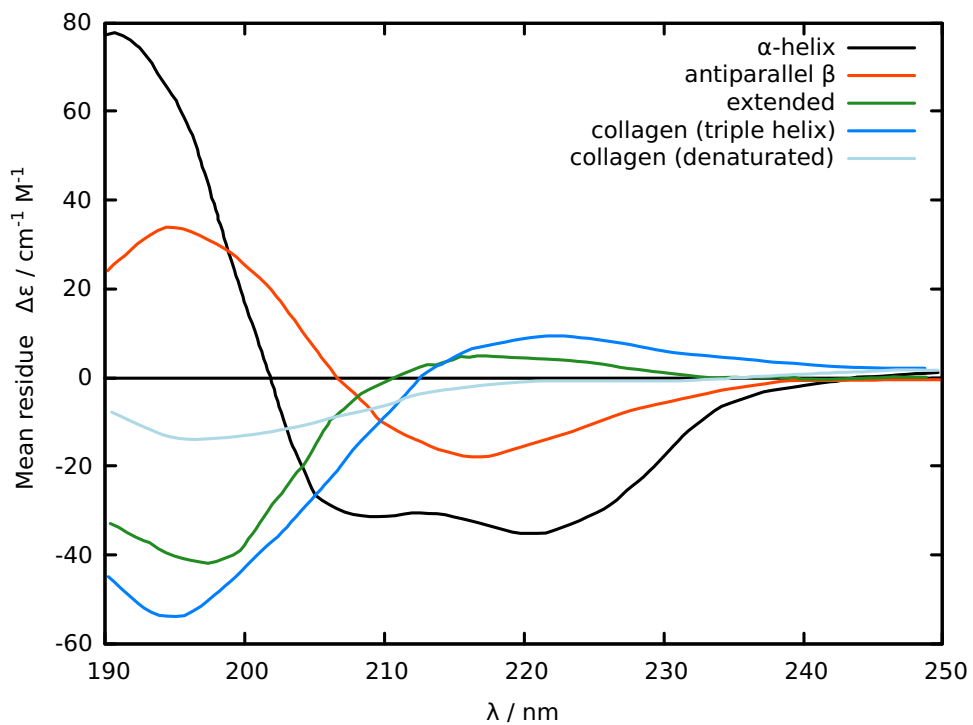


Figure 1.2.: CD spectra of representative secondary structure motifs. Replotted from Ref. 28.

CD spectroscopy requires only low sample concentrations<sup>29,30</sup> and is, therefore, a widely used technique in medical research.<sup>31,32</sup> In particular, CD plays an important role in studies about prion diseases.<sup>29,33</sup> Prions are incorrectly folded proteins that can transfer their misfolding onto other instances of prion proteins.<sup>34–36</sup> Many lethal and transmissible neurodegenerative diseases in mammals result from such a cascaded misfolding of prions.<sup>34,36–39</sup> This folding process can be monitored by CD spectroscopy since the natural state of prions comprises  $\alpha$ -helices that change into beta-sheet or random coil structures.<sup>33,35</sup> Figure 1.3 depicts the globular part of the prion protein structure.

## 1. Introduction

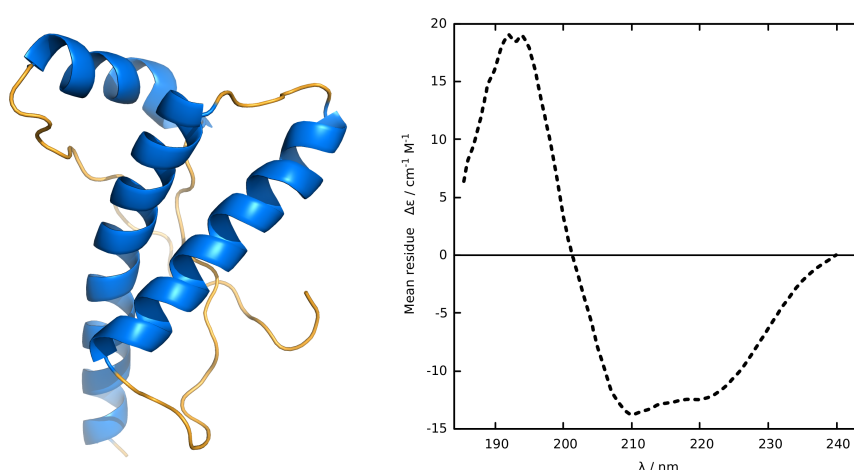


Figure 1.3.: Left: Cartoon representation of the globular domain of the human prion protein (protein database (PDB) code: 1QLX).<sup>40</sup> The blue parts correspond to an  $\alpha$ -helical secondary structure and the orange parts to unstructured loop regions. Right: Experimental CD spectrum of the human prion protein in its natural state. Replotted from Ref. 41.

Schellman pointed out three essential aspects for the sufficient theoretical description of CD spectra.<sup>42</sup> First, macroscopic CD spectra must have a connection to QM derived properties. Second, efficient electronic structure methods must be developed to compute the QM properties sufficiently. Third, CD is highly sensitive to the molecular structure, and therefore, the structural ensemble (all populated conformations of the molecule) in solution must be reproduced as accurate as possible. The first aspect was accomplished in 1975 for CD with the derivation of the rotatory strength via electric and magnetic transition dipole moments.<sup>42,43</sup> The second aim is solved for small molecules (less than 20 atoms) by applying high-level coupled-cluster theory.<sup>11,13,14</sup> For medium sized molecules (less than 100 atoms), time-dependent (TD) - density functional theory (DFT) established as a method for computing CD spectra with reasonable accuracy.<sup>11,15,16,18,44</sup> However, large molecules or aggregates – containing more than 100 atoms – pose still a difficulty for electronic structure methods and are only treatable in a fragmentation approach in the past.<sup>45–47</sup> The third aspect is also not challenging for small and rigid molecules. However, the treatment of many chemical systems remains non-trivial due to the complexity of generating a full ensemble.<sup>48</sup> This is especially problematic, when chromophores comprise flexible molecular geometries, which is of particular importance for the simulation of protein CD spectra.<sup>49</sup>



## 1.2. Second harmonic imaging technique (SHIM)

Another important spectroscopic property, used in medical research is the second harmonic generation (SHG)<sup>50,51</sup>, on which the biomedical imaging technique SHIM bases.<sup>52–54</sup> SHG is a nonlinear optical (NLO) effect where a photon is scattered in a material with twice the energy of the two incident photons.<sup>50</sup> The SHG phenomenon is schematically depicted in Figure 1.4.

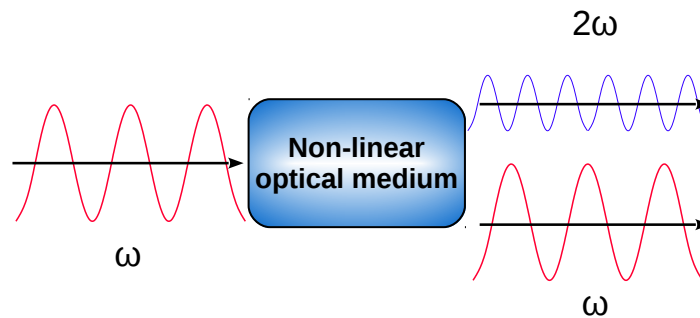


Figure 1.4.: Schematic view of the SHG conversion of an incident photon beam with frequency  $\omega$  in a non-linear optical medium.

SHIM was developed to enhance the contrast when imaging non-centrosymmetric molecular arrangements.<sup>55–58</sup> Structures like the collagen triple helix<sup>59–61</sup>, microtubules, or  $\alpha$ -helix-rich myosins<sup>55,57</sup> – considered as endogenous probes –, are responsible for the captured SHG signals. Biomolecular arrangements that are unable to exhibit SHG can be made visible in SHIM via exogenous labeling by the use of dyes.<sup>55</sup> Since the scattering process of SHG is non-invasive, SHIM is widely used for the analysis of cancer progression and tumor characterization.<sup>62–66</sup> Figure 1.5 depicts two ovarian biopsies.<sup>66</sup> The collagen structures are made visible by SHIM and differ significantly between the healthy and malignant probe.

## 1. Introduction

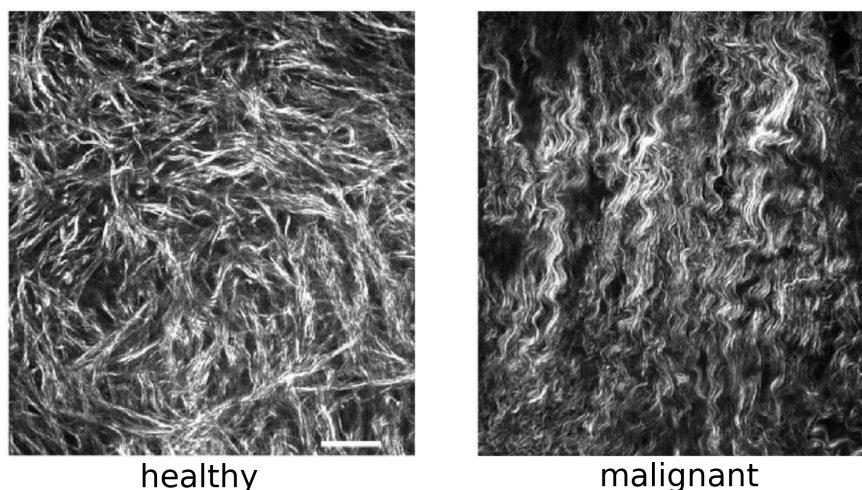


Figure 1.5.: SHG images from normal (left) and cancerous (right) ovarian biopsies. The scale bar refers to a distance of 25  $\mu\text{m}$  (from Nadiarnykh, O. *et al.* **2010** *BMC Cancer* 10:94., open access).

SHIM bases on a macroscopic phenomenon and is, therefore, inaccessible within a direct approach for theoretical methods. The underlying SHG, however, can be accessed via the first hyperpolarizability, which is a molecular property. Theoretical methods can then be used for the prediction and interpretation of the first hyperpolarizabilities. This is very powerful in an interdisciplinary approach for the synthesis of exogenous dyes, and the genetic engineering of endogenous biotags.<sup>67</sup> Small organic molecules up to 50 atoms pose no difficulty for computational methods in terms of an accurate prediction of hyperpolarizabilities.<sup>68–70</sup> Large systems – with more than 1000 atoms (e.g., proteins) – still challenge theoretical methods and are considered as unachievable with current methods.<sup>71</sup>

### 1.3. Objectives

The objective of this work is to test and study the applicability of newly developed, efficient computational methods for the treatment of large, biomolecular systems. A particular focus is set on spectroscopic properties like CD and first hyperpolarizabilities.

Experimental conditions or investigated systems are often simplified or approximated to limit the complexity of the theoretical studies. This implies the risk of massive errors in the simulations. The aim of the present work is closing this gap between real systems in their natural state and model systems, applied in the theoretical calculations. The recently developed methods based on simplified TD-DFT<sup>72,73</sup> and extended tight-binding ( $\times\text{TB}$ )<sup>74–76</sup> show huge potential for re-

ducing approximations of experimental systems or conditions. These methods are sufficiently tested for molecules smaller than 500 atoms, but still, the transferability to large biomolecular systems is non-trivial and is investigated in this work thoroughly.

The upcoming chapter gives a brief overview of well-established QM methods that are relevant for the present work. Part II deals with the application of the sTDA-xTB method for CD and UV-Vis spectra. Part III is devoted to the application of the sTD-DFT-xTB method for NLO properties, e.g., the molecular first hyperpolarizability. Part IV draws conclusions and provides perspectives about the possible advances for the discussed research fields.



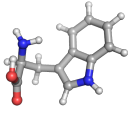

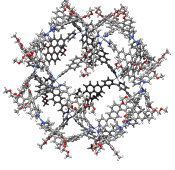
## 2. Theoretical Background

### 2.1. Methodological Overview

Spectroscopic properties are commonly measured in solution and at ambient temperature (298 K). Concentrations of several  $\text{mmol L}^{-1}$  are typical for such experiments. A macroscopic simulation is inaccessible for theoretical models because the consideration of a droplet with a volume of 1 mL requires the simulation of  $10^{17}$  molecules. Instead, a nanoscopic approach is chosen, where non-interacting solute molecules are simulated as a single molecule or a set of conformations. The solvent is, in most cases, either neglected or approximated within a continuum approach. This way, the electrostatic environment of the solution is accounted for. Quantum mechanical methods are then applied to compute the electronic structure of the solute molecule. Finding a suitable computational and theoretical method for the treatment of the target systems is the task of computational chemistry. A compromise between versatility, accuracy, and speed is inevitable when choosing a level of theory. The result of this compromise is often a multilevel scheme, in which many theoretical levels are applied to solve specific tasks. The general classes of theoretical methods are the *ab-initio* wave function theory (WFT), *first-principles* density functional theory (DFT), semiempirical quantum mechanics (SQM), and classical mechanics methods (like force fields (FFs)). The latter applies an atomistic treatment in contrast to the electron-based approaches of the remaining methods. There are also coarse-grained, and continuum mechanics available as theoretical methods, but they are not considered in this thesis. Table 2.1 gives a general overview of the typically-used level of theories for different sized systems.

## 2. Theoretical Background

Table 2.1.: Overview of the multilevel scheme in quantum chemistry.

# atoms	small <100	medium 100–500	large >500
			
conformational sampling	SQM	SQM	FF
optimization/frequencies	DFT	SQM	SQM/FF
energies	WFT/DFT	DFT	low-cost DFT/SQM
excited states	WFT/DFT	DFT/SQM	SQM

The computation of large biomolecular systems and their excited state properties is the aim of this thesis. Due to the large molecule size (typically larger than 1000 atoms), SQM is the most chosen level of theory in the following chapters. WFT and DFT are used in this thesis as reference methods for systems of feasible size, and provide the theoretical basis for SQM. Thus, in the following, the main principles of these theories are explained.

### 2.2. Wave function theory (WFT)

The time-independent, non-relativistic Schrödinger equation (SE)<sup>77</sup> is a common origin for wave function (WF) based, quantum-mechanical theory for molecules:

$$\hat{H}\psi = E\psi, \quad (2.1)$$

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}, \quad (2.2)$$

with the WF  $\psi$  and the Hamiltonian  $\hat{H}$ . The latter can be separated into kinetic ( $\hat{T}$ ) and potential ( $\hat{V}$ ) energy operators of the particles e and n (electrons and nuclei, respectively). Due to the much slower speed of the nuclei in comparison to the electrons, they are de facto immobile. The separation of the electron and nuclei-dependent terms of the Hamiltonian because of the speed difference of the particles is called Born–Oppenheimer (BO) approximation.<sup>78</sup> The electronic SE reads as follows:

$$\hat{H}_e\psi_e = E_e\psi_e, \quad (2.3)$$

$$\hat{H}_e = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee}. \quad (2.4)$$

## 2.2. Wave function theory (WFT)

The electronic Hamiltonian and its different parts can be expressed as sums over the contributing particles (atomic units are used throughout this thesis):

$$\hat{H}_e = -\frac{1}{2} \sum_i^N \hat{\nabla}_i^2 - \sum_i^N \sum_A^M \frac{Z_A}{r_{iA}} + \sum_i^{N-1} \sum_{j>i}^N \frac{1}{r_{ij}}, \quad (2.5)$$

with the indices  $i, j$  for electrons, and  $A$  for nuclei.  $N, M$  are the number of electrons and nuclei, respectively. The kinetic energy is computed as the second derivative ( $\hat{\nabla}_i^2$ ) of the wave function with respect to the atomic positions. The second term yields the potential or Coulomb energy between nucleus  $A$  and electron  $i$ . Both terms combine into the one-electron operator  $\hat{h}_i$ . The third term represents the Coulomb potential for the electron pair  $ij$ , which is designated as the two-electron operator  $\hat{V}_{ee}$ . The electronic SE can thus be formulated as:

$$\hat{H}_e \psi_e = \left[ \sum_i^N \hat{h}_i + \sum_i^{N-1} \sum_{j>i}^N \frac{1}{r_{ij}} \right] \psi_e = E \psi_e, \quad (2.6)$$

$$\hat{h}_i = -\frac{1}{2} \sum_i^N \hat{\nabla}_i^2 - \sum_A^M \frac{Z_A}{r_{iA}}. \quad (2.7)$$

The expectation value of the electronic Hamiltonian and the normalized, electronic WF yield the electronic energy (in Dirac notation):

$$E_e = \langle \psi_e | \hat{H}_e | \psi_e \rangle. \quad (2.8)$$

The Coulomb energy between all nuclei pairs is added as a constant to yield the total energy of the system:

$$\begin{aligned} E_{\text{tot}} &= E_e + \sum_A^{M-1} \sum_{B>A}^M \frac{Z_A Z_B}{r_{AB}} \\ &= E_e + E_{\text{nn}}. \end{aligned} \quad (2.9)$$

Since in the following only electronic terms are considered, the index  $e$  is neglected. Furthermore, the Dirac notation for all integrals is used throughout. After detailing the Hamiltonian, the wave function is now discussed. A reasonable way of expressing an electronic wave function is the Slater determinant<sup>79</sup>  $\Phi_0$ . Here, the so-called molecular orbital (MO) represents independent one-electron

## 2. Theoretical Background

wave functions

$$\Phi_0(\underline{1}, \underline{2}, \dots, \underline{N}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{1\sigma}(\underline{1}) & \psi_{2\tau}(\underline{1}) & \cdots & \psi_{N\nu}(\underline{1}) \\ \psi_{1\sigma}(\underline{2}) & \psi_{2\tau}(\underline{2}) & \cdots & \psi_{N\nu}(\underline{2}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1\sigma}(\underline{N}) & \psi_{2\tau}(\underline{N}) & \cdots & \psi_{N\nu}(\underline{N}) \end{vmatrix}, \quad (2.10)$$

where the determinant is normalized by  $\frac{1}{\sqrt{N!}}$ . It contains  $N$  spin orbitals  $\psi_i(j)$  that are occupied by  $N$  electrons. Using a Slater determinant ensures the antisymmetry of the WF with respect to an exchange of two electrons (Pauli principle<sup>80</sup>).

**Hartree–Fock (HF)** The HF approximation<sup>81</sup> is the most established approach to transform the WF-based SE into computationally solvable equations. Although this theory lacks some crucial terms for an accurate computation of the electronic energy, it is the foundation for many more elaborated WF theories. It is based on the pseudo-eigenvalue equation:

$$\hat{f}_i \phi_i = \epsilon_i \phi_i, \quad (2.11)$$

$$\hat{f}_i = \hat{h}_i + \hat{v}_i^{\text{HF}}, \quad (2.12)$$

$$\hat{v}_i^{\text{HF}} = \sum_j^N \hat{J}_{ij} - \hat{K}_{ij}. \quad (2.13)$$

The Fock operator  $\hat{f}_i$  connects to the MOs  $\phi_i$  as eigenfunctions and to the orbital energies  $\epsilon_i$  as eigenvalues. The operators  $\hat{J}_{ij}$  and  $\hat{K}_{ij}$  are the Coulomb and exchange operators, respectively, and are defined as:

$$\hat{J}_{ij} |\phi_i(1)\rangle = \langle \phi_j(2) | \frac{1}{r_{ij}} | \phi_j(2)\rangle | \phi_i(1)\rangle, \quad (2.14)$$

$$\hat{K}_{ij} |\phi_i(1)\rangle = \langle \phi_j(2) | \frac{1}{r_{ij}} | \phi_i(2)\rangle | \phi_j(1)\rangle. \quad (2.15)$$

The operator  $\hat{K}$  arises from the antisymmetry of the wave function and incorporates interactions of electrons with the same spin. Operator  $\hat{J}$  represents the mean Coulomb interaction of all electrons. This so-called mean-field approach is the key feature of the HF approximation. The exact electron pair Coulomb repulsion is substituted by an approximated, average field, created by all electrons. This way, the many-body problem of the two-electron term is circumvented.

The expansion of the MOs in a linear combination of atomic orbitals (LCAO) is another essential scheme. This way, the prior integrodifferential equations can be reformulated in a matrix form (Roothaan equations<sup>82</sup>). Thus, standard mathe-



### 2.3. Density functional theory (DFT)

matical computation schemes can be applied to solve these equations.

$$\phi_i = \sum_{\mu} C_{\mu i} \varphi_{\mu} \quad (2.16)$$

$$\mathbf{FC} = \epsilon \mathbf{CS} \quad (2.17)$$

$\mathbf{F}$  denotes the Fock matrix that contains elements of the form  $\langle \phi_i | \hat{f} | \phi_j \rangle$ . The MO coefficients are contained in matrix  $\mathbf{C}$ . To ensure orthogonal orbitals, the overlap matrix  $\mathbf{S}$  is introduced. Since the Fock matrix depends on the coefficients – contained in  $\mathbf{C}$  –, the equation must be solved iteratively. This is achieved by a linear variation of the MO coefficients (self-consistent field (SCF) approach). The final energy that is based on the converged, self-consistent coefficients is then computed by:

$$E_{\text{HF}} = \sum_i^N \langle \phi_i | \hat{h}_i | \phi_i \rangle + \sum_i^{N-1} \sum_{i>j}^N \left( \langle \phi_i | \hat{J}_{ij} | \phi_i \rangle - \langle \phi_i | \hat{K}_{ij} | \phi_i \rangle \right). \quad (2.18)$$

**Correlation** This theoretical approach properly describes the correlated motion of two electrons of parallel spin that is due to the Pauli principle<sup>80</sup> (Fermi correlation). The correlated movement, related to the acting Coulomb repulsion between the electrons (Coulomb correlation), is neglected, since a single electron in HF theory only perceives the average field of the remaining electrons. Even at the basis set limit of atomic orbitals (HF limit), the HF theory is unable to give the exact solution to the BO approximated, non-relativistic SE of a multi electron system. Thus, the (electron) correlation energy defines as the difference between the exact energy and the energy obtained at the HF limit\*:

$$E_c = E_{\text{exact}} - E_{\text{HF}}. \quad (2.19)$$

### 2.3. Density functional theory (DFT)

Hohenberg and Kohn proofed<sup>9</sup> that the electron density  $\rho$  of a system completely determines the electronic energy of the ground state. DFT is based on this direct connection. An unknown density functional  $E[\rho]$  exists that relates the exact energy to the electron density. In analogy to equation 2.4, the functional can be split into further, contribution-specific functionals:

$$E[\rho] = T_e[\rho] + V_{\text{en}}[\rho] + V_{\text{ee}}[\rho], \quad (2.20)$$

---

\*The HF limit for the definition of the correlation energy is considered as the lowest possible unrestricted HF energy.

## 2. Theoretical Background

$$V_{ee}[\rho] = J[\rho] + K[\rho], \quad (2.21)$$

with the functionals for the kinetic energy  $T_e[\rho]$ , the Coulomb interaction of electrons and nuclei  $V_{en}[\rho]$ , and the Coulomb interaction of electron pairs  $V_{ee}[\rho]$ . Similar to equation 2.13, the latter can be segmented into functionals for Coulomb ( $J[\rho]$ ) and Exchange ( $K[\rho]$ ) interactions. The described density functional theory is, in principle, orbital free. This implies a massive benefit over standard WFT, since the density only depends on three variables (e.g., the coordinates in space), the wave function, however, on  $3N$  variables ( $N$ : number of electrons). The classical formulations for the Coulomb terms  $J[\rho]$  and  $V_{en}[\rho]$  are:<sup>83,84</sup>

$$V_{en}[\rho] = \sum_A^M \int \frac{Z_A(R_A)\rho r}{|R_A - r|} dr, \quad (2.22)$$

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(r)\rho(r')}{|r - r'|} dr dr'. \quad (2.23)$$

Thomas, Fermi and Dirac were the first to derive – based on the uniform electron gas (UEG) – the approximations for  $T_e[\rho]$  (Thomas<sup>85</sup> and Fermi<sup>86</sup>) and  $K[\rho]$  (Dirac<sup>87</sup>):

$$T_e[\rho] = \frac{3}{10} (3\pi^2)^{\frac{2}{3}} \int \rho(r)^{\frac{5}{3}} dr, \quad (2.24)$$

$$K[\rho] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(r)^{\frac{4}{3}} dr. \quad (2.25)$$

This approach refers to the so-called Thomas–Fermi–Dirac (TFD) model. The kinetic energy of the electrons is computed within this approximation very inaccurately. Since this is a large part of the total energy, the TFD model is unable to calculate bonding in molecules sufficiently.

**Kohn–Sham (KS) DFT** Kohn and Sham<sup>88</sup> proposed an improvement on the orbital-free TFD model. They introduced a reference system of non-interaction electrons that has the same electron density as the real system. Then, the kinetic energy is computed of these KS-orbitals. Since this DFT ansatz is the only commonly used version of DFT, the term KS-DFT will be denoted in the following as DFT. This approach recovers almost 99% of the kinetic energy, but the introduced orbitals extend the variables from three to  $3N$  (as in WFT). In KS theory, the missing kinetic energy, the correlation energy and exchange energy are collected in the exchange–correlation functional  $E_{XC}$ , which is usually split in an exchange  $E_X$  and a correlation  $E_C$  part:

$$E_{XC}[\rho] = E_X[\rho] + E_C[\rho]. \quad (2.26)$$

### 2.3. Density functional theory (DFT)

Except for this term, which has to be approximated, the remaining terms are computed analytically, similar to HF theory. The central equation of KS-DFT is formulated as

$$\hat{f}_i^{\text{KS}}[\rho]\phi_i = \left[ \hat{h}_i[\rho] + \sum_j^N (\hat{J}_{ij} + v_{\text{XC}}[\rho]) \right] \phi_i = \epsilon_i \phi_i, \quad (2.27)$$

$$v_{\text{XC}}[\rho] = \frac{\delta E_{\text{XC}}[\rho]}{\delta \rho}, \quad (2.28)$$

with  $\hat{f}_i^{\text{KS}}$  as the KS-operator (similar to the Fock operator). The exchange-correlation potential  $v_{\text{XC}}[\rho]$  substitutes the exchange operator from HF theory. The iterative solving of this equation yields the electronic energy and the final set of KS orbitals.

**Jacob's ladder** There are many approaches to approximate the exchange-correlation functional  $E_{\text{XC}}$ . These density functional approximations (DFAs) are typically categorized by the information that enters the equations for the computation of the exchange or correlation energy. Perdew and Schmidt established the picture of the mythological *Jacob's ladder* for the different levels of DFAs.<sup>89</sup> Within this metaphor, the accuracy increases from the 'Hartree hell' to the 'heaven of chemical accuracy'. The local spin density approximation (LSDA) occupies the lowest rung of the ladder. The exchange-correlation terms of this class of DFAs depend only on the local electron density  $\rho[\vec{r}]$ . DFAs within the generalized gradient approximation (GGA) take additionally the first derivative of the electron density  $\vec{\nabla}\rho[\vec{r}]$  into account (second rung). On the third rung, higher order derivatives of the electron density or the kinetic energy density  $\tau$  are considered and denoted as meta-GGAs. The step to the next rung is considered as a big improvement in terms of accuracy.<sup>90,91</sup> In these so-called hybrid functionals<sup>92</sup>, the exchange energy is partly substituted by the analytically derived Fock exchange (inclusion of occupied orbitals). The last rung of DFAs – called double-hybrid – considers additionally information about virtual orbitals,<sup>93</sup> e.g., by means of a modified Møller-Plesset perturbation theory.<sup>94</sup>

## 2.4. Linear Response Theory

The absorption of electromagnetic radiation in the UV-Vis energy range provokes transitions from the electronic ground state to an excited state. The previously described HF and DFT methods provide orbital coefficients that correspond to a wave function of the electronic ground state. For the computation of the excited state, the ground state wave function must be perturbed by a TD external potential. In the case of UV-Vis absorptions, the potential is small and thus solvable within a linear response treatment.<sup>95</sup> The excitation energies and transition dipoles are obtained by solving the TD-DFT non-Hermitian eigenvalue problem:<sup>96,97</sup>

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \omega & 0 \\ 0 & -\omega \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \quad (2.29)$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are the orbital rotation Hessian matrices, with  $\mathbf{X}$  and  $\mathbf{Y}$  as their eigenfunctions. The excitation energy  $\omega$  is computed as the eigenvalue of this equation. The matrix elements for  $\mathbf{A}$  and  $\mathbf{B}$  are defined – for the spin-restricted case – as:

$$\mathbf{A}_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + 2\langle ij|ab\rangle - a_x\langle ia|jb\rangle + (1 - a_x)\langle ij|f_{XC}|ab\rangle, \quad (2.30)$$

$$\mathbf{B}_{ia,jb} = 2\langle ib|aj\rangle - a_x\langle ia|bj\rangle + (1 - a_x)\langle ib|f_{XC}|aj\rangle. \quad (2.31)$$

The DFT ground state calculation yields the orbital energies of the virtual and occupied orbitals ( $\epsilon_a$  and  $\epsilon_i$ , respectively).  $a_x$  denotes the amount of Fock exchange in the applied functional. The response of the density functional, within the adiabatic approximation, is given by the term  $\langle ia|f_{XC}|jb\rangle$  and  $\langle ia|f_{XC}|bj\rangle$ , where  $f_{XC}$  reads as follows:

$$f_{XC}(r_1, r_2) = \frac{\delta^2 E_{XC}[\rho]}{\delta\rho(r_1)\delta\rho(r_2)}. \quad (2.32)$$

For real orbitals, the non-Hermitian linear response time-dependent eigenvalue problem in equation 2.29 can be transformed into a Hermitian one:<sup>96,98</sup>

$$(\mathbf{A} - \mathbf{B})^{\frac{1}{2}}(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})^{\frac{1}{2}}\mathbf{Z} = \omega^2\mathbf{Z}, \quad (2.33)$$

while  $\mathbf{Z}$  is given by:

$$\mathbf{Z} = \sqrt{\omega}(\mathbf{A} - \mathbf{B})^{-\frac{1}{2}}(\mathbf{X} + \mathbf{Y}). \quad (2.34)$$

**Tamm–Dancoff approximation (TDA)** Hirata and Head-Gordon applied the TDA<sup>99</sup> to TD-DFT, where the contribution of the **B** matrix is neglected.<sup>100</sup> Instead of solving two eigenvalue problems as in hybrid TD-DFT, only one eigenvalue problem is solved.

$$\mathbf{A}\mathbf{t} = \omega_{\text{TDA}}\mathbf{t} \quad (2.35)$$

To emphasize that the eigenfunctions and eigenvalues are different to the ones in equation 2.29, **t** replaces the solution vector **X**, and the index TDA is added to  $\omega$ . In case of  $\alpha_x = 1$ , TDA equals the configuration interaction singles (CIS) approach.<sup>97</sup> TDA excitation energies are only slightly larger than the respective ones obtained from TD-DFT, but at a much lower cost.<sup>100</sup>

## 2.5. Extended tight-binding (xTB)

Semiempirical quantum mechanical methods were introduced in the early days of quantum chemistry because computational resources were limited, and even small molecules were inaccessible with standard HF theory or DFT. To have applicable methods back then, the computationally demanding integrals of the *ab-initio* theories were approximated, and empirical parameters were introduced to recover most of the accuracy. SQMs have not lost their importance in modern quantum chemistry. Even with the existing computational resources, the *ab-initio* simulation of large supra- or biomolecules is unfeasible. Thus SQM or FF methods are required to overcome the limitation of computationally demanding *ab-initio* models.

The recently proposed suite of xTB methods is the basis for many calculations in this thesis. In chapter 4, the simplified Tamm–Dancoff approximation (sTDA)-xTB method<sup>76</sup> for the computation of excited state spectra is detailed. For the structure optimization and molecular dynamic simulations, the geometries, frequencies, and non-covalent interactions (GFNn)-xTB (n=1,2)<sup>74,75</sup> methods are used, which are introduced in this section.

**GFN1-xTB** The GFN1-xTB method uses a modified, minimally polarized valence Slater type orbital basis set (STO-nG). Additional *s* functions are assigned to hydrogen atoms to account for hydrogen bonding. *d* functions are placed on heavier main group elements to describe hypervalent structures properly. Table 2.2 shows the applied basis set for the GFN1-xTB method.

## 2. Theoretical Background

Table 2.2.: Description of the Slater-type AO basis sets of GFN1-xTB.  $n$  denotes the principal quantum number of the valence shell of the element.

element	basis functions
H	$ns, (n+1)s$
He	$ns$
Be-F, group 1, Zn, Cd, Hg, Tl-Bi	$nsp$
group 2, 13-18	$nsp, (n+1)d$
transition metals and lanthanides	$(n-1)d, nsp$

The total energy of GFN1-xTB defines as:

$$E_{\text{tot}}^{\text{GFN1-xTB}} = E_{\text{el}} + E_{\text{rep}} + E_{\text{disp}}^{\text{D3(BJ)}} + E_{\text{XB}} + G_{\text{Fermi}}. \quad (2.36)$$

For the computation of the electronic energy  $E_{\text{el}}$ , the GFN1-xTB Fock matrix elements have to be set up:

$$\begin{aligned} F_{\mu\nu}^{\text{GFN1-xTB}} = & H_{\mu\nu}^0 + \frac{1}{2} S_{\mu\nu} \sum_C \sum_{l''} (\gamma_{AC, l''} + \gamma_{BC, l''}) p_{l''}^C \\ & + \frac{1}{2} S_{\mu\nu} (q_A^2 \Gamma_A + q_B^2 \Gamma_B) \end{aligned}$$

with  $\mu \in l(A), \nu \in l'(B)$ . (2.37)

The first order term  $H_{\mu\nu}^0$  contains information about the Hückel constants, the effective atomic energy levels, and the atomic electronegativities. The second order term incorporates the electrostatic contributions  $\gamma$  within a monopole approximation and shell-wise charges  $p_{l''}^C$ . The indices  $l''$  and  $C$  denote the angular momenta and nuclei, respectively. Mulliken<sup>101</sup> partial charges  $q$  and the charge derivative of the chemical hardness *via* the Hubbard parameter  $\Gamma$ <sup>102</sup> determine the third order term. The atomic overlap  $S_{\mu\nu}$  scales the second and third order terms. After a self-consistent computation of the charges, the electronic energy is derived by:

$$E_{\text{el}} = \sum_i n_i \langle \psi_i | H_0 | \psi_i \rangle + \frac{1}{2} \sum_{A,B} \sum_{l(A)} \sum_{l'(B)} p_l^A p_{l'}^B \gamma_{AB, l''} + \frac{1}{3} \sum_A \Gamma_A q_A^3 - T_{\text{el}} S_{\text{el}}. \quad (2.38)$$

The repulsion energy  $E_{\text{rep}}$  of equation 2.36 is computed as an atom pair-based contribution, including the element-specific parameter  $\alpha$ , and the global parameter

## 2.5. Extended tight-binding (xTB)

$k_f$ .  $Z_A^{\text{eff}}$  denotes the effective nuclear charge of atom A.  $E_{\text{rep}}$  defines as:

$$E_{\text{rep}} = \sum_{AB} \frac{Z_A^{\text{eff}} Z_B^{\text{eff}}}{R_{AB}} e^{-(\alpha_a \alpha_b)^{0.5} (R_{AB})^{k_f}}. \quad (2.39)$$

The established D3 correction<sup>103</sup> – applied with a Becke–Johnson (BJ) damping function<sup>104</sup> – is used for the computation of the dispersion energy  $E_{\text{disp}}$ . D3 describes properly the long-range electron correlation effects that are typically missing in most of the approximated  $E_{\text{XC}}$  terms of KS-DFT. Three-body terms are neglected for the GFN1-xTB dispersion. The atom pair-wise term reads as follows:

$$E_{\text{disp}}^{\text{D3(BJ)}} = -\frac{1}{2} \sum_{A \neq B} \left( s_6 \frac{C_6^{\text{AB}}}{R_{AB}^6 + (\alpha_1 R_{AB}^0 + \alpha_2)^6} + s_8 \frac{C_8^{\text{AB}}}{R_{AB}^8 + (\alpha_1 R_{AB}^0 + \alpha_2)^8} \right), \quad (2.40)$$

with the internuclear distance  $R_{AB}$  between atom A and B, and for GFN1-xTB adjusted  $\alpha_1$ ,  $\alpha_2$ , and  $s_8$  damping parameters. The dispersion coefficients  $C_6^{\text{AB}}$  and  $C_8^{\text{AB}}$  are determined element-wise by precomputed hydride model systems with varying coordination numbers. The  $s_6$  parameter is set to 1 by definition<sup>†</sup> and  $R_{AB}^0$  is calculated from the dispersion coefficients.

The halogen bond correction  $E_{\text{XB}}$  bases on a Lennard-Jones potential<sup>105</sup> and defines as:

$$E_{\text{XB}} = \sum_{\text{XB}} f_{\text{damp}}^{\text{AXB}} k_x \frac{1 + \left( \frac{R_{\text{cov,AX}}}{R_{\text{AX}}} \right)^{12} - k_{\text{X2}} \left( \frac{R_{\text{cov,AX}}}{R_{\text{AX}}} \right)^6}{\left( \frac{R_{\text{cov,AX}}}{R_{\text{AX}}} \right)^{12}}, \quad (2.41)$$

with the global parameter  $k_{\text{X2}}$  and the effective covalent distances  $R_{\text{cov,AX}}$ . The geometrical dependence of the halogen bonds is included in  $f_{\text{damp}}^{\text{AXB}}$  that decays for non-linear AXB constellations.

No spin density-dependent terms are present in GFN1-xTB, but to be able to handle static correlation effects, a finite temperature treatment is applied. The last term of equation 2.36 ( $G_{\text{Fermi}}$ ) is the contribution of an electronic free energy at a finite electronic temperature due to Fermi smearing.<sup>106</sup> It ensures a variational solution for fractionally occupied orbitals and reads as follows:

$$G_{\text{Fermi}} = k_B T_{\text{el}} \sum_{\sigma=\alpha,\beta} \sum_i [n_{i\sigma} \ln(n_{i\sigma}) + (1 - n_{i\sigma}) \ln(1 - n_{i\sigma})], \quad (2.42)$$

with  $k_B$  as Boltzmann's constant and  $n_{i\sigma}$  as fractional occupation number of the spin MO.

<sup>†</sup>Except for double-hybrid DFAs.

## 2. Theoretical Background

**GFN2-xTB** The GFN2-xTB method supersedes its precursor (GFN1-xTB) in many cases. The main theoretical advances are the inclusion of multipole electrostatics – up to second-order – and the omission of a specific halogen bond correction and an adjusted basis set for hydrogen bonding. The applied basis functions in GFN2-xTB are shown in table 2.3.

Table 2.3.: Description of the Slater-type AO basis sets of GFN2-xTB.  $n$  denotes the principal quantum number of the valence shell of the element.

element	basis functions
H	$ns$
He	$ns, (n+1)p$
group 1, Be-F, Zn, Cd, Hg-Po	$nsp$
Ne	$nsp, (n+1)d$
group 2, 13-18	$nspd$
transition metals and lanthanides	$nd, (n+1)sp$

The total energy expression of GFN2-xTB is defined as:

$$E_{\text{tot}}^{\text{GFN2-xTB}} = E_{\text{rep}} + E_{\text{disp}} + E_{\text{EHT}} + E_{\text{IES+IXC}} + E_{\text{AES}} + E_{\text{AXC}} + G_{\text{Fermi}}. \quad (2.43)$$

The repulsion energy  $E_{\text{rep}}$  is determined by equation 2.39. The dispersion energy  $E_{\text{disp}}$  is a self-consistent variant of the D4 model<sup>107,108</sup> that accounts for the electronic structure *via* atomic partial charges. The extended Hückel-type energy  $E_{\text{EHT}}$  is computed similar to the first term, and the isotropic electrostatic and exchange-correlation energy  $E_{\text{IES+IXC}}$  similar to the second and third term of equation 2.38. The energy of the anisotropic electrostatic interactions  $E_{\text{AES}}$  is derived from contributions from charge-dipole, charge-quadrupole, and dipole-dipole interactions:

$$E_{\text{AES}} = E_{q\mu} + E_{q\Theta} + E_{\mu\mu}, \quad (2.44)$$

with  $q$ ,  $\mu$ , and  $\Theta$  as cumulative atomic multipole moments (CAMM)<sup>109</sup> up to second order. These CAMMs are also used for the computation of the anisotropic exchange-correlation energy  $E_{\text{AXC}}$ :

$$E_{\text{AXC}} = \sum_{\text{A}} \left( f_{\text{XC}}^{\mu\text{A}} |\mu_{\text{A}}|^2 + f_{\text{XC}}^{\Theta\text{A}} \|\Theta_{\text{A}}\|^2 \right), \quad (2.45)$$

where  $f_{\text{XC}}^{\mu\text{A}}$  and  $f_{\text{XC}}^{\Theta\text{A}}$  are fitted element-specific parameters. The last term ( $G_{\text{Fermi}}$ ) of equation 2.43 is adapted from the GFN1-xTB theory, i.e., equation 2.42.



## **Part II.**

# **Development and Application of the sTDA-xTB method**



## II. Development and Application of the sTDA-xTB method

As outlined in chapter 1, CD spectroscopy is a viable connection between theory and experiment for the structure elucidation of molecules in solution. Therefore, it is of particular interest to develop efficient theoretical models for the simulation of CD spectra. Furthermore, such methods must be tested for their applicability to large systems, since this is by far non-trivial.

Part II deals with the diversification and application of the sTDA-xTB method<sup>76</sup>, which has emerged as a fast and accurate method for the calculation of electronic excitation spectra. In particular, chapter 3 addresses the transferability of the sTDA-xTB method to large biomolecular systems like entire proteins and deoxyribonucleic acid (DNA) fragments. For this purpose, CD and UV-Vis spectra are computed and compared to the experiment for three proteins and one DNA fragment. Additionally, the effect of sampling structures along a molecular dynamics (MD) trajectory on the spectra simulation is investigated. This is the first time that relevant biomolecular systems can be treated within a full QM approach for the calculation of the excited state properties.

The original publication of the sTDA-xTB method<sup>76</sup> only provided an element parameter set for the most important elements (H-Zn,Br,I). Some metals, naturally occurring in proteins (e.g., cadmium), were missing from this first parameter set. Such metal-containing proteins are important for many biological mechanisms and of great interest for spectroscopic studies.<sup>110,111</sup> sTDA-xTB is designed to be a fast, accurate, and versatile method, and therefore, the parameterization for almost the entire periodic table is essential. Chapter 4 is dedicated to close this parameterization gap and obtain the missing parameters for the 4d and 5d metals, as well as for the 4p, 5p, and 6p element blocks. The quality of these new parameters is analyzed, and the applicability of the sTDA-xTB method for large transition metal-containing complexes is shown.

Chapter 3 reports the applicability of the sTDA-xTB method for protein and DNA spectra calculation. However, different level of theories are used to obtain optimized structures (e.g., HF-3c<sup>112</sup>, GFN1-xTB<sup>74</sup>, AMBER ff12SB<sup>113</sup>). HF-3c performed very well for the proteins and DNA fragments in chapter 3 but is computationally not feasible for proteins larger than 1000 atoms. The AMBER FF, on the other hand, is a very fast method for protein structure optimization, due to its atomistic design and special parameterization. The latter, however, limits the versatility of this method and, e.g., metal-containing proteins are not accessible. The recently proposed universal FF GFN-FF<sup>114</sup> is not limited to certain chemical motifs and is able to compute the structural properties of systems with remarkable accuracy. Chapter 5 is devoted to assess the structure methods GFN2-xTB<sup>75</sup> (the successor of GFN1-xTB) and GFN-FF with respect to their performance for providing optimized structures and MD simulations as basis for the sTDA-xTB spectra calculations of proteins.



### 3. Biomolecular structure information from high-speed quantum mechanical electronic spectra calculation

Jakob Seibert\*, Christoph Bannwarth<sup>†</sup> and Stefan Grimme\*

*Received 6th of June 2017, Published online 18th of August 2017*

Reprinted (adapted) with permission from<sup>‡</sup>

Seibert, J; Bannwarth, C.; Grimme, S. *J. Am. Chem. Soc.* **2017**, *139*, 11682–11685.

— Copyright © 2017, American Chemical Society.

DOI <https://doi.org/10.1021/jacs.7b05833>

#### Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

---

\*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

<sup>†</sup>Department of Chemistry, Stanford University Stanford, California 94305, United States

<sup>‡</sup>Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

**Abstract** A fully quantum mechanical (QM) treatment to calculate electronic absorption (UV-Vis) and circular dichroism (CD) spectra of typical biomolecules with thousands of atoms is presented. With our highly efficient sTDA-xTB method, spectra averaged along structures from molecular dynamics (MD) simulations can be computed in a reasonable time frame on standard desktop computers. This way, non-equilibrium structure and conformational as well as purely quantum mechanical effects like charge-transfer or exciton-coupling are included. Different from other contemporary approaches, the entire system is treated quantum mechanically and neither fragmentation nor system-specific adjustment is necessary. Among the systems considered are a large DNA fragment, oligopeptides, and even entire proteins in an implicit solvent. We propose the method in tandem with experimental spectroscopy or X-ray studies for the elucidation of complex (bio)molecular structures including metallo-proteins like myoglobin.

### 3.1. Introduction

Electronic circular dichroism (CD) spectroscopy has become an indispensable tool in biochemistry. Along with nuclear magnetic resonance (NMR) and X-ray diffraction, CD spectroscopy plays an important role in structure elucidation of proteins or DNA<sup>21</sup>. Compared to X-ray diffraction, it is a much less demanding technique both in terms of sample size and preparation time. While this is also true for NMR, the latter does not reveal information about the absolute molecular configuration. Consequently, CD has become an important tool in biochemists' daily routine. The far UV region of protein absorption is dominated by electronic transitions in the amide groups and the corresponding CD spectra are highly sensitive to their configuration, namely the protein secondary structure<sup>49</sup>. Different secondary structure motifs have characteristic CD spectra in this region allowing for in-vitro studies of protein folding, which is important for prion diseases like BSE<sup>34-36</sup>. Furthermore, differences in the tertiary structure between mutant and wild type proteins can be drawn from the near UV region.<sup>27</sup>

In particular due to developments on intense synchrotron radiation sources, protein CD spectra can now be recorded down to about 160 nm wavelengths<sup>24</sup>. However, regarding the theoretical methods, which enable the interpretation of these far UV spectra, there is still much room for improvement. The treatment of molecules with thousands of atoms by ab-initio methods, without requiring special computer hardware, is routinely not possible and currently employed approaches to calculate CD spectra of proteins include specifically trained neural networks and exciton coupling approaches<sup>115-118</sup>. The transferability of these methods to other secondary structures, chromophoric units, or energy windows (e.g., down to 160 nm) has not been demonstrated and a fundamental, purely quantum mechanical

treatment of the entire protein (without fragmentation) is highly desirable. We have recently introduced a semi-empirical quantum mechanical (QM) scheme, sTDA-xTB, which is specifically designed to compute electronic absorption and CD spectra of molecules with more than 1000 atoms and for (almost) arbitrary chemical composition<sup>76</sup>. Whereas the good performance for small systems is established, the applicability of this method to large realistic biomolecular systems still remains to be shown. Here, we demonstrate for the first time the capability of our high-speed method to compute CD spectra of polypeptides, proteins, and DNA fragments in an almost "black-box" manner. We focus here on the excited state calculations which can be combined with existing schemes for structure generation and conformational sampling.

### 3.2. Results and Discussion

First, we show that the method can simulate the characteristic features of isolated secondary structure motives such as a small alpha helix or a short beta sheet. It is well-known that even small oligopeptides with a few amino acids exhibit characteristic CD spectra<sup>119</sup>. We computed for seven different structures of a small peptide the respective CD spectra (Figure A1.3). It is clearly shown that the characteristic features of the protein secondary structure motives correlate with the dihedral angles of the input structures (Ramachandran angle  $\phi$  and  $\psi$ )<sup>120</sup> and hence the sTDA-xTB method is capable to describe such structure sensitive chromophores. In Figure 3.1 the CD spectrum of a peptide containing 20 residues is shown (PDB code 1l2y)<sup>121</sup>. The secondary structure mainly consists of  $\alpha$ -helices. In the experimental CD spectrum, features characteristic for  $\alpha$ -helices are present. The positive band at 190 nm and the negative at 210 nm correspond to  $\pi \rightarrow \pi^*$  transitions of the peptide chromophore and the negative feature at 225 nm is denoted in the literature as the  $n \rightarrow \pi^*$  transition<sup>19</sup>. The simulated spectrum fits very well to the experimental spectrum. All bands are reproduced very well in terms of absolute intensities and excitation energies (all bands are shifted by a case specific constant of typically a few tenths of an eV correcting for systematic errors in the calculations in particular regarding the treatment of solvation and the neglect of geometry relaxation upon excitation). The calculation of electronic spectra for such relatively small model peptides (230 atoms) would be possible in principle with more sophisticated DFT based methods in hours to days of computation time. In this respect our method is quite outstanding as the entire calculation takes only 26 seconds on a conventional desktop computer for a single structure. This allows for the increase of the size of molecules to that of realistic and really interesting systems, which would not be possible otherwise. Myoglobin plays a very important role in intramuscular oxygen transport. Cal-

## II. Development and Application of the sTDA-xTB method

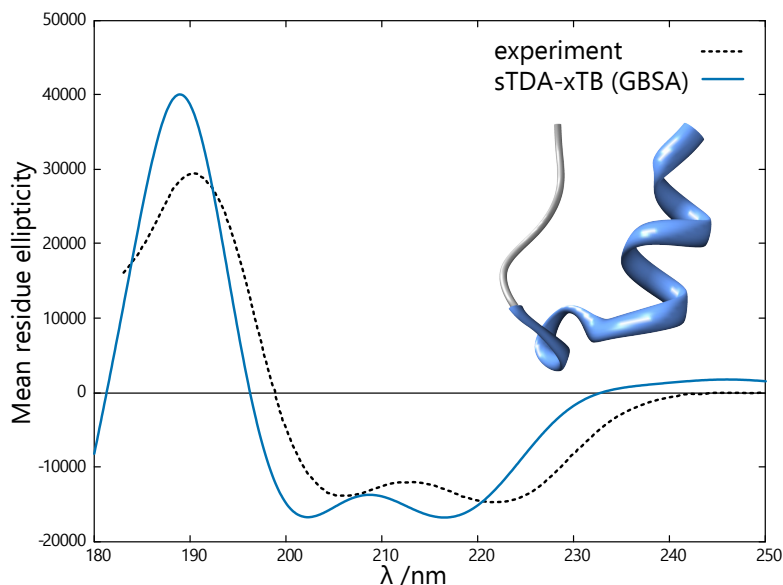


Figure 3.1.: Calculated CD spectra of 1l2y (20 residue protein) (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV. The experimental curves (black dotted line) are taken from Ref.<sup>121</sup> and refer to a solution of 1l2y in water (pH 7 buffer).

culating an electronic spectrum of this well known member of the globin family, containing around 2500 atoms, is far beyond the capabilities of conventional quantum mechanical methods. In Figure 3.2 the CD spectrum computed by sTDA-xTB for myoglobin is shown. Due to the eight alpha helices in myoglobin, the CD spectrum is dominated by the band features described in the last example for the 1L2Y model peptide. The computed spectrum shows all characteristic features and is in a very good agreement with the experimental one. The spectrum of myoglobin is computed on 100 snapshots taken from an MD trajectory. This is necessary in order to eliminate artificial CD transitions on floppy side chains, which interfere with the main transitions of the peptide backbone and which would be present in a single structure approach while they average out by the MD. The proposed theoretical method allows the computation of hundreds of snapshots in a reasonable time frame.

The sTDA-xTB method was not specifically fitted to describe proteins and as being a QM approach, it is not limited to protein calculations. In order to show its versatility, a second important structural motive is studied, the DNA. The three main conformations (A-, B- and Z-DNA) are distinguishable in CD spectra resulting from different configurations of the chromophores (the nucleobase pairs) and hence the shape of the helix.<sup>122</sup> Here, a well defined DNA system<sup>123</sup> with recurring sequence is used. In Figure 3.3 the comparison of the experimental and calculated



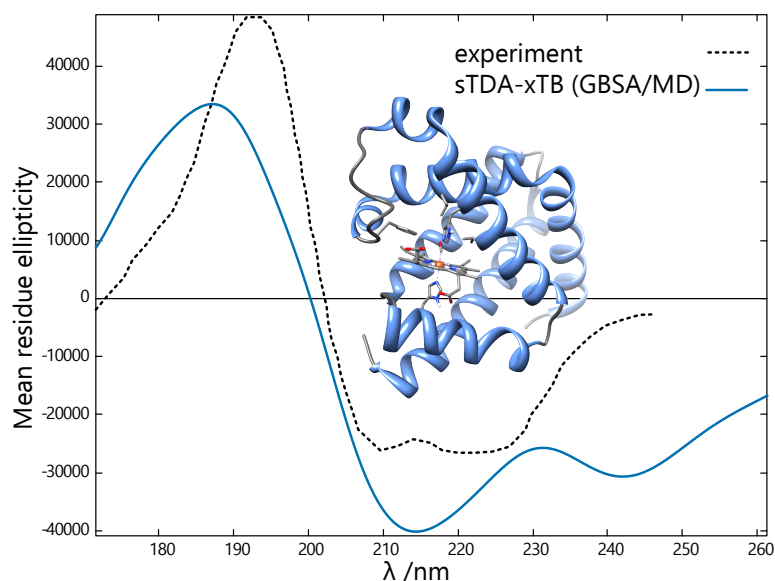


Figure 3.2.: Calculated CD spectra of myoglobin computed (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at  $1/e$  maximum of 0.35 eV and the spectrum is red-shifted by 1 eV. The experimental curves (black dotted line) are taken from Ref.<sup>24</sup> and refer to a solution of myoglobin in water.

CD spectrum of poly[d(A)].poly[d(T)] is given which exhibit remarkable mutual agreement. Despite the somewhat different spectral range in peptides/proteins compared to DNA, the computed spectra are equally well described. Note, that no DNA specific adjustments have been made in our theoretical procedure. We have now shown that our method is capable of calculating spectra in the far UV region (protein backbone, 160-240 nm) and that it is also suitable in the near UV (DNA, < 300 nm). This consistency represents already an achievement because the electronic structure of high- and low-energy excited states can be rather different.<sup>14</sup>

Next, we will show that our approach is also suited to absorption in the visible range thus bridging excitation energy scales occurring in typical bio-molecules. The photoactive yellow protein (PYP) is a small cytosolic photoreceptor.<sup>124</sup> The chromophore of this protein is a deprotonated 4-hydroxy cinnamic acid derivative which is covalently bound to the apoenzyme via a thioester bond to Cysteine 69. The sTDA-xTB computed absorption and CD spectrum in comparison to the experiment is presented in Figs. 4 and S1, respectively, showing again excellent correspondence. The computed vertical excitation energy for the bright  $\pi \rightarrow \pi^*$  transition (3.27 eV) is in line with values from higher level methods like TD-DFT and SAC-CI<sup>45</sup> obtained for fragmented (divide and conquer) model systems. Our approach is significantly faster (computation time of 3.5 h) and

## II. Development and Application of the sTDA-xTB method

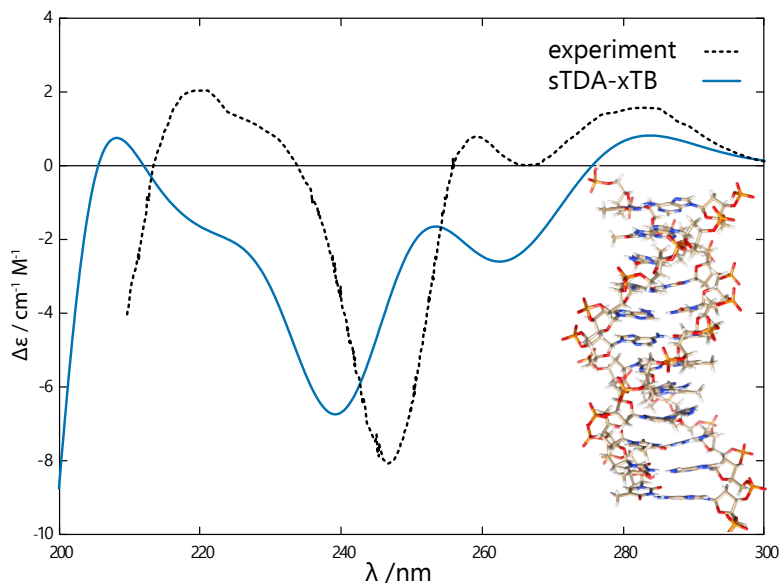


Figure 3.3.: Calculated CD spectrum of poly[d(A)].poly[d(T)] (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.6 eV. The experimental curves (grey solid line) are taken from Ref.<sup>122</sup> and refer to an aqueous 3.5M NaCl solution of poly[d(A)].poly[d(T)].

requires no fragmentation allowing for an almost black box use. Although other semi-empirical methods for excited state calculation exists and were previously compared to sTDA-xTB<sup>76</sup>, the extension to the presented systems is not obvious. Often, such approaches are not suitable, due to the insufficient minimal basis or lack of parametrized elements. The solvation model is another main issue. Other semi-empirical methods have been combined rather with surface segment-based solvation models like PCM<sup>126</sup> or COSMO<sup>127</sup>. These solvation models are state-of-the-art for implicit solvation in a first-principles QM framework. However, they are numerically noisy and too elaborate in combination with a semi-empirical QM method, thus being detrimental in the application to large biomolecules. The sTDA-xTB method employs a GBSA approach<sup>128,129</sup>, which does not suffer from the aforementioned problems. Furthermore, the sTDA-xTB has hardly any limitations regarding the parametrization and employs diffuse basis functions. It even has some advantages over the TD-DFT approaches, which often suffer from self-interaction error related issues. Concerning the PYP system, a discussion of some TD-DFT problems<sup>97,130</sup> and comparison of sTDA-xTB and hybrid sTDA-DFT is presented in the Supporting Information. The computation of spectra, which are dominated by valence or Rydberg transitions over a large spectral range is by far not trivial. Even range-separated hybrid TD-DFT does not behave consistently for differently sized chromophores<sup>131</sup>. The difficulties of

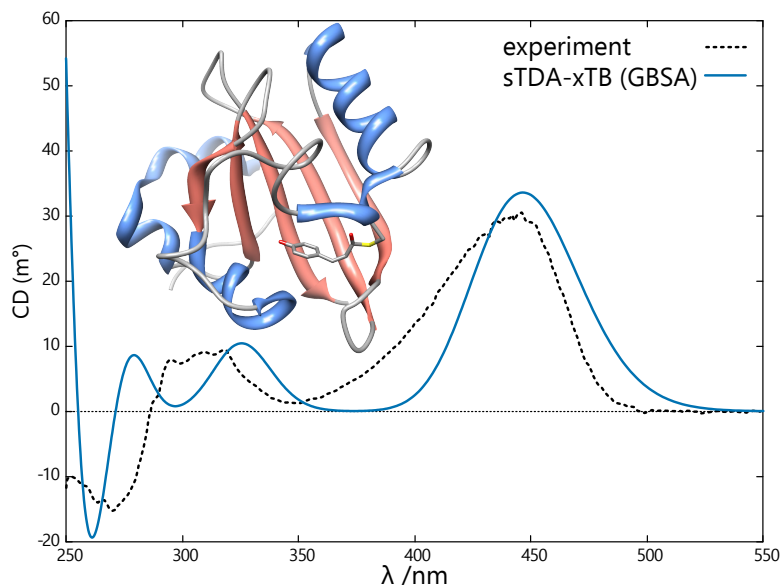


Figure 3.4.: Calculated CD spectrum of the photoactive yellow protein (PYP) (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV. The experimental curves (grey solid line) are taken from Ref.<sup>125</sup> and refer to a solution of PYP in water.

treating flexible systems, as indicated in Figure A1.3, still remains. An evaluation of excited state spectra becomes complex and dependent on the weights of the individual conformations.<sup>132,133</sup> The presented sTDA-xTB method offers a solution for routinely approaching the reliable computation of spectral features for truly flexible biomolecules. Even large conformational ensembles could be treated. In principle, sTDA-xTB may be combined with any underlying QM or MM approach that provides good structures or gives rise to reasonable MD ensembles. We apply the in our opinion best, yet, still affordable and generally applicable semi-empirical QM approach, namely our recently presented GFN-xTB<sup>74</sup> method. It can be used in “affordable” sampling procedures, e.g., around the minimum position in large biomolecular systems. The importance of this sampling is very clearly depicted in Figure A1.5. When going beyond the minimum structure approach, one loses the bias induced by the minimum structure. The combination of using thermodynamically resolved statistical weights of conformers and MD simulation to account for non-minimum effects will be the aim of future work.

### 3.3. Conclusion

The four presented examples show the versatile applicability of our sTDA-xTB method which is not limited to a certain spectral range or type of chromophore.

## II. Development and Application of the sTDA-xTB method

The vacuum-UV to UV range is very important for elucidating structures or conformations of aromatic amino acids in the protein or cofactor chromophores like heme, chlorophyll, or the copper complex in hemocyanine. Further applications like differentiation of similar protein secondary structure ( $3_{10}$  and  $\alpha$ -helix) or DNA structures (A-, B- and Z-DNA) are conceivable and will boost the interplay between theoretical and experimental CD studies of various biomolecules. Furthermore, one can envisage the treatment of spectra of intercalation complexes of DNA and dyes or fluorescent tags (e.g., ethidium bromide). For medium-sized systems with molecules comprising of a few hundred atoms, the usage of an ordinary consumer computer is sufficient and bigger computational resources are only necessary for large molecules ( $>1000$  atoms). Thus, the sTDA-xTB method is a highly efficient yet still accurate method which opens up new opportunities to a wide community and even to laymen in quantum chemistry.

### 3.4. Technical Details

For an unbiased comparison of ECD spectra, the experimental absorption spectrum is used to shift the computed spectrum to match a distinctive absorption band (the standard deviation of sTDA-xTB for vertical valence excitations is roughly 0.4 eV, with no systematic under/overestimation tendency<sup>76</sup>). As observed in the original publication, computation of spectra along an MD trajectory seems to reduce the magnitude of the red-shift compared to the one obtained from the ground state minimum geometry. If an experimental absorption spectrum is absent, the shift is evaluated according to the (unsigned or squared) ECD spectrum, ideally keeping the shift smaller in magnitude than 0.8 eV. One 1L2Y structure (NMR) was taken from the PDB database<sup>121,134</sup> and optimized at the HF-3c level of theory<sup>112</sup> applying the COSMO implicit solvation model ( $\epsilon=78$ )<sup>127</sup>. (for more details see Figure A1.5). The optimized structure was taken as input for the xTB ground state calculation with the GBSA model for water.<sup>128</sup> All excitations up to 8 eV were calculated with the sTDA method<sup>76</sup>. The velocity formalism is used for the rotatory strengths. Starting from the myoglobin PDB X-Ray structure<sup>135</sup>, the missing hydrogen atoms were added and with GFN-xTB<sup>74</sup> the geometry was optimized with implicit solvent (GBSA). A molecular dynamics (MD) simulation was carried out for 6 ps with a timestep of 1 fs. From the resulting trajectory 100 snapshots were taken equidistantly and considered as input structures for the sTDA-xTB calculations. All excitations up to 9 eV were calculated in the sTDA part. The resulting rotatory strengths are scaled by a factor of three and the length formalism is used for the rotatory strengths. As initial structure of the photoactive yellow protein (PYP) the X-ray diffraction structure 2PHY<sup>136</sup> from the PDB was used and Hydrogen atoms were added. The full protein was then optimized with

## II. Development and Application of the sTDA-xTB method

a specific QM/MM approach, which is discussed in detail in the Supporting Information. The optimized structure was taken as input for the xTB ground state calculation with the GBSA model for water.<sup>128</sup> All excitations up to 9 eV were calculated with the sTDA method<sup>76</sup>. The length formalism is used for the rotatory strengths. Concerning the origin independence of the used formalism, we refer to the discussion in the SI. The intensities are displayed in arbitrary units to match the intensity of the most intense band. The structure of Poly[d(A)].poly[d(T)] is build up from scratch with the help of symmetry transformations available in the CRYSTAL<sup>137</sup> program and optimized at the 1D-periodic HF-3c level of theory. The optimization is carried out using the rotational-translational symmetry with 11 base pairs per unit cell and containing 72 water molecules in the grooves of the helix. After the optimization a fragment containing 12 base pair is prepared and saturated at the cut surface and the phosphate backbone, resulting in a neutral system. This fragment geometry is used in the sTDA-xTB procedure. All excited states up to 10 eV are computed in the sTDA step. The length formalism for the rotatory strengths is used and all rotatory strengths are scaled by a factor of 0.2. If not stated otherwise the TURBOMOLE<sup>138,139</sup> suite of programs (version 7.0) is used for all geometry optimizations with HF-3c while the CRYSTAL14<sup>137</sup> program is used for the periodic optimization of the DNA. For geometry optimization, MD simulation, ground state calculation with the xTB method, the stand-alone program xtb<sup>140</sup> is used. All excited state calculations are conducted with the stda stand alone program<sup>72</sup>.

**Acknowledgments** The authors thank Jan Gerit Brandenburg for the generation and periodic optimization of the DNA structure. This work was supported by the DFG in the framework of the 'Gottfried-Wilhelm-Leibniz' prize.

**Supporting Information** Additional information about the comparison of sTDA-xTB to hybrid sTDA-DFT calculations and convergence test for DNA fragments and further spectra and a full description of the methodological details.



## 4. Extension of the element parameter set for ultra-fast excitation spectra calculation (sTDA-xTB)

Jakob Seibert\*, Jana Pisarek\*, Sarah Schmitz\*, Christoph Bannwarth<sup>†</sup> and Stefan Grimme\*

*Received 14th of June 2018, Published online 16th of August 2018*

Reprinted (adapted) with permission from<sup>‡</sup>

Seibert, J; Pisarek, J; Schmitz, S; Bannwarth, C.; Grimme, S. *Mol. Phys.* **2018**, *117*, 1104–1116.

— Copyright © 2018, Taylor and Francis.

DOI [10.1080/00268976.2018.1510141](https://doi.org/10.1080/00268976.2018.1510141)

### Own manuscript contribution

- Performing the calculations for the comparison to theory/experiment
- Interpretation of the results
- Writing the manuscript

### Contribution of co-authors:

- Reference data collection (S. Grimme, J. Pisarek)
- Parameterization (J. Pisarek)

---

\*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

<sup>†</sup>Department of Chemistry, Stanford University Stanford, California 94305, United States

<sup>‡</sup>Permission requests to reuse material from this chapter should be directed to Taylor and Francis.

## II. Development and Application of the sTDA-xTB method

**Abstract** The extension of the parameter set for an ultra-fast electronic excitation spectra calculation is presented. The semiempirical theory based on a tight-binding approach, called extended tight-binding (xTB) in combination with the simplified Tamm-Dancoff approximation (sTDA) shows remarkable accuracy at very low computational cost for the calculation of vertical excitation energies of molecules. It enables the possibility for computing even large systems up to thousands of atoms or sampling along molecular dynamic (MD) trajectories. The original publication of the sTDA-xTB method included parameters for the most important elements (H-Zn,Br,I). In this work, element parameters for 4d and 5d metals, and the missing ones in 4p, 5p and 6p element blocks are presented and analyzed for their quality. Comparisons to theory and experiment show that sTDA-xTB provides similar good results as for the elements in the original publication with an average deviation of excitation energies of 0.3–0.5 eV.

### 4.1. Introduction

The spectroscopic characterization of a synthesized compound is an essential part of laboratory work. In addition to nuclear magnetic resonance (NMR) and mass spectrometry, vibrational spectra and optical excitation spectra are important in the assignment/identification of a substance. Theoretical methods and experimental practices work hand in hand thanks to the availability of fast computer resources and efficient algorithms to investigate various properties of molecules. In order to be able to analyze even larger systems, the theoretical method must use efficient approximations. Density functional theory (DFT) is, by now, the workhorse of most computational chemists. This allows the routine calculation of systems on a scale of up to hundreds of atoms. Beyond this system size, it is often difficult to provide efficient support for experimental chemists in the elucidation of the properties of a substance. Whereas for very large systems with thousands of atoms only classical mechanics are possible, semiempirical approaches bridge the gap between the worlds of force field and DFT. For the calculation ground state properties of organic molecules there are a multitude of semiempirical quantum mechanical methods available which are routinely used<sup>141–143</sup>. However, if the system to be studied contains heavier elements, the aforementioned methods quickly approach their limits, as appropriate parameterization is often missing. For excited states and excitation spectra, the mainstream method is time dependent DFT (TD-DFT). The computational limits in terms of system size for the aforementioned ground state methods hold for excited states as well while the number of semiempirical methods available for excited state treatments is limited. The accuracy under these methods<sup>144–147</sup> are comparable, but the number of parametrized elements is fairly limited. For an excellent review on semiempirical



## II. Development and Application of the sTDA-xTB method

Group → ↓ Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
Lanthanides				57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
Actinides				89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

Figure 4.1.: Overview of the parametrized elements for the sTDA-xTB method.

quantum mechanical methods see Ref. 148

Based on the tight binding idea, originally proposed by Seifert, Frauenheim and Elstner<sup>149</sup>, the extended tight binding (xTB) method was developed, which is parametrized for many elements of the periodic table<sup>76</sup>. This is made possible by an element-specific parameterization, in contrast to element pairwise potentials, as is the practice in other tight-binding methods.<sup>150,151</sup>

Within a semiempirical approach, the approximate description limits the flexibility and achievable accuracy of the method. Therefore, the xTB variants are purpose specific. Parameterization is focused on one or a few molecular properties, which increases the accuracy and robustness of the calculations. Within the xTB framework, a variant for calculating electronic excitations was initially presented, but not with a complete parameterization for the entire periodic table. After a version of the xTB approach (GFN-xTB<sup>74</sup>) has been published for geometries, frequencies and non-covalent interactions, the extension of the parameter set of the electronic spectra variant (sTDA-xTB) is presented in this work.

## 4.2. Theory

### 4.2.1. The simplified Tamm-Dancoff approximation to TD-DFT

Starting from the basic TDA-DFT equation 4.1

$$\mathbf{A}\mathbf{X}_{\text{TDA}} = \omega_{\text{TDA}}\mathbf{X}_{\text{TDA}} \quad (4.1)$$

three modifications lead to the simplified Tamm-Dancoff approximated density functional theory approach (sTDA-DFT):<sup>72</sup>

## II. Development and Application of the sTDA-xTB method

1. The response term of the exchange-correlation density functional ( $ia|f_{XC}|jb$ ) in matrix  $\mathbf{A}$  is neglected. Consequently, computationally demanding steps resulting from the numerical integration are avoided.
2. The remaining two-electron integrals in matrix  $\mathbf{A}$  are approximated as atom-centered transition charge density interactions utilizing the Mataga-Nishimoto-Ohno-Klopman<sup>152-154</sup> damped Coulomb law:

$$(pq|rs) \approx \sum_{\alpha,\beta}^{\text{atoms}} q_{pq}^{\alpha} q_{rs}^{\beta} \gamma_{\alpha\beta} \quad (4.2)$$

The elements of the resulting simplified matrix  $\mathbf{A}'$  are then given by the following equation:

$$A'_{ia,jb} = \delta_{ij} \delta_{ab} (\epsilon_a - \epsilon_i) + \sum_{\alpha,\beta}^{\text{atoms}} \left( s_k q_{ia}^{\alpha} \gamma_{\alpha\beta}^K q_{jb}^{\beta} - q_{ij}^{\alpha} \gamma_{\alpha\beta}^J q_{ab}^{\beta} \right) \quad (4.3)$$

The atom-centered monopole transition densities  $q_{pq}^{\alpha}$  and  $q_{pq}^{\beta}$  located on atom  $\alpha$  and  $\beta$ , respectively, are obtained from a Löwdin population analysis<sup>72,155</sup>. In the spin-restricted case,  $s_k$  is equal to a value of two for singlet-singlet and  $s_k$  equals zero for singlet-triplet excitations, as it results from the integration of the spin component in the electronic wave function. The variables  $\gamma_{\alpha\beta}^K$  and  $\gamma_{\alpha\beta}^J$  correspond to the Mataga-Nishimoto-Ohno-Klopman<sup>152-154</sup> damped Coulomb operators for exchange type ( $\gamma^K$ ) and Coulomb type ( $\gamma^J$ ) integrals, respectively:

$$\gamma_{\alpha\beta}^K = \left( \frac{1}{(R_{\alpha\beta})^{y_K} + \eta^{-y_K}} \right)^{\frac{1}{y_K}} \quad (4.4)$$

$$\gamma_{\alpha\beta}^J = \left( \frac{1}{(R_{\alpha\beta})^{y_J} + (a_x \eta)^{-y_J}} \right)^{\frac{1}{y_J}} \quad (4.5)$$

Here, the arithmetic average of the chemical hardness for the two atoms  $\alpha$  and  $\beta$  is given by  $\eta$  and tabulated  $\eta(\alpha)$  values are taken from work of Ghosh et al.<sup>156</sup>.  $y_K$  and  $y_J$  are parameters and the distance between two atoms is given by  $R_{\alpha\beta}$ . In a DFT context,  $a_x$  refers to the amount of Fock exchange in the functional, whereas it is a parameter ( $a_x = 0.5$ )<sup>76</sup> in sTDA-xTB.

3. The single excitation space is reduced by truncating all configurations at a user-specified energy threshold at only minor loss of accuracy. In order to describe important configurations beyond this threshold properly, such configurations are accounted for perturbatively (see Ref.<sup>72</sup> for a detailed

description).

Overall, the drastic two-electron repulsion integral approximations and the truncated single excitation space, in addition to the simple eigenvalue problem (Eq.4.1), lead to computational savings of more than three orders of magnitude compared to regular TD-DFT.<sup>72</sup> This allows extremely fast computation of electronic absorption and electronic circular dichroism (CD) spectra in the excitation energy range of up to 10 eV. As stated in earlier works,<sup>72,157</sup> CT states can principally be described correctly using the sTDA method since the scaling of  $(ij|ab)$  with  $a_x$  is absent at long-range (i.e., for  $R_{\alpha\beta} \rightarrow \infty$ ) in the approximated sTDA (see Eq. 4.5). This is a conceptual improvement over TD-DFT with standard density functionals and allows, in principle, a correct treatment even for large systems with up to thousands of atoms. The extension to the full linear response case provides the simplified TD-DFT (sTD-DFT) approach. Here, matrix  $\mathbf{B}$  is set up in a consistent, simplified manner.<sup>73</sup>

#### 4.2.2. The extended tight-binding composite scheme sTDA-xTB

The computational bottleneck of sTDA-DFT, as well as of sTD-DFT, calculations resides in solving the self-consistent field (SCF) equations for the electronic ground state to determine the necessary KS orbitals and eigenvalues. To overcome this problem, a specifically designed TB procedure has been introduced<sup>76</sup>. Orbitals and orbital energies which are used in the sTDA treatment are taken from this fast semi-empirical calculation. These orbitals are supposed to emulate the ones from a hybrid DFT treatment and are obtained in an extended atomic orbital (AO) basis set including minimal valence and diffuse AO functions. One of the key features of this TB method is the usage of two different Hamiltonian/basis set combinations with individual parameterizations. Hereafter, the two Hamiltonians used are termed VTB and XTB, respectively, and their corresponding basis sets are denoted by VBS and XBS, while the overall procedure is termed xTB. Since xTB replaces the ground state DFT treatment in sTDA-DFT, the overall scheme is termed sTDA-xTB (following the notation of Ref.<sup>76</sup>).

1. In the VTB treatment Gasteiger-type<sup>158</sup> atomic partial charges serve as input to construct an initial VTB Hamiltonian matrix. These only depend on the geometry of the corresponding molecule by

$$\Delta q_{\alpha} = (N_{\alpha}^{\text{EN}} - Z_{\alpha}^{\text{val}}) \frac{N_{\text{tot}}}{\sum_{\beta} N_{\beta}^{\text{EN}}} \quad (4.6)$$

with

$$N_{\alpha}^{\text{EN}} = Z_{\alpha}^{\text{val}} + \sum_{\beta}^{\text{atoms}} \frac{EN_{\alpha} - EN_{\beta}}{R_{\alpha\beta}^6}. \quad (4.7)$$

## II. Development and Application of the sTDA-xTB method

Here,  $N_{\alpha}^{\text{EN}}$  is the normalized number of electrons whereas  $N_{\text{tot}}$  represents the total number of valence electrons in the system.  $Z^{\text{val}}$  denotes the valence atomic charge and EN is the element-specific electronegativity. For the newly parametrized elements, EN values are determined in this work and explicit values are provided in the Supporting Information.

2. The calculated atomic Gasteiger-type charges are inserted for  $\Delta q_{\kappa}$  in the VTB Hamiltonian matrix according to

$$\langle \psi_{\mu} | F | \psi_{\nu} \rangle = \langle \psi_{\mu} | H_0 | \psi_{\nu} \rangle + k_q \frac{1}{2} S_{\mu\nu} \sum_{\kappa} (\gamma_{\alpha\kappa} + \gamma_{\beta\kappa}) \Delta q_{\kappa} \quad (\mu \in \alpha, \nu \in \beta) \quad (4.8)$$

The indices  $\alpha$ ,  $\beta$ , and  $\kappa$  denote atoms.  $S_{\mu\nu}$  is the AO overlap matrix element between the AO basis functions  $\mu$  and  $\nu$ . As in the second-order self-consistent charge (SCC) density functional tight-binding (DFTB) method<sup>149</sup>,  $H_0$  is the zeroth-order part.  $k_q$  is an empirical scaling parameter for the electron-electron repulsion. Similar to sTDA,  $\gamma$  denotes the damped inter-electronic repulsion function (Eq. 4.4 with  $y_{\kappa}=2$ ):

$$\gamma_{\alpha\beta} = \sqrt{\frac{1}{R_{\alpha\beta}^2 + \eta^{-2}}} \quad (4.9)$$

After diagonalization, the resulting density matrix is used for the computation of Mulliken atomic charges.

3. In the next step CM5 charges<sup>159</sup> are calculated based on the Mulliken partial charges.
4. A second VTB Hamiltonian matrix is constructed with these CM5 charges employed for  $\Delta q_{\kappa}$  and diagonalized. The SCC procedure is stopped after two iterations and the resulting new Mulliken-based CM5 charges and orbitals are used in the subsequent XTb treatment.
5. An XTb Hamiltonian matrix is constructed using Eq. 4.8 employing the final VTb CM5 charges. This matrix is then diagonalized once and the resulting molecular orbitals (MOs) and eigenvalues serve as input for the established and only slightly modified (local excitation correction and shifted virtuals, see Ref.<sup>76</sup>) sTDA method using the existing `stda` code.<sup>160</sup>

### 4.2.3. Extended AO basis set

In the xTB composite scheme correct occupied/virtual orbital energy splittings as well as proper Rydberg states can be obtained, because the general eigenvalue

problem is solved in a non-orthogonal, extended AO basis<sup>76</sup> which is an advantage over common NDDO type methods. In the VTB part a minimal valence basis set is used for the calculation of the atomic charges. Due to the fact that some electronegative main group elements of third and higher rows can form hypervalent electronic structures, a single d-polarization function is added to third through sixth row elements. To treat the important Rydberg states as accurately as possible, the XTB part is constructed using a minimal valence basis set with additional diffuse functions (XBS) on hydrogen and other non-metal main group elements. The type and composition of all AO basis functions is given in Table 4.1.

Table 4.1.: Description of the used AO basis sets.  $n$  denotes the principal quantum number of the valence shell of the respective element. Compare Figure 4.1 for differentiation between original work and this work.

element	VTB	XTB
H-He	ns	ns, (n+1)sp
group I/II	nsp	nsp
B-Ne	nsp	nsp, (n+1)sp
group IIb	nsp	nsp
remaining d-block elements	nd, (n+1)sp	nd, (n+1)sp
non-metal main group elements <sup>a,c)</sup>	nsp, (n+1)d	nsp, (n+1)sp
post transition metals <sup>b,c)</sup>	nsp, (n+1)d	nsp

a) non-metal main group elements: Si-Cl, Ge-Br, Sb-I, Po-At

b) post transition metals: Al, Ga, In-Sn, Tl-Bi

c) In the code no 6d AOs are implemented. Hence, for the elements Tl-At 6s6p5d AO basis sets are used in VTB. Furthermore, no 7sp AOs are available, leading to a 6s6p5sp AO basis sets for the elements Po and At in XTB.

## 4.3. Parameterization and technical details

### 4.3.1. Reference data collection

The fit set of molecules contains more than 420 closed shell systems for main group elements (Ga-Se, Rb-Te, Cs-At) as well as about 320 molecules for the d-block elements (Y-Cd and La, Hf-Hg). All geometries are optimized at the PBEh-3c<sup>161</sup> level with the TURBOMOLE 7.0.2<sup>138,139</sup> program suite to obtain accurate

## II. Development and Application of the sTDA-xTB method

equilibrium structures. Structures are then verified as minima by frequency calculations with the `aoforce`<sup>162,163</sup> program. Molecules are chosen to express a high variability of properties like oxidation states, coordination numbers, and chemical environments. Reference Hirshfeld<sup>164</sup> charges for all molecules are calculated with ORCA<sup>165</sup> version 3.0 using the hybrid functional PBE0<sup>166</sup> in combination with an Ahlrichs type triple-zeta basis set without f-functions (def2-TZVP(-f)<sup>167</sup>). Additionally, an exchange-correlation integration grid of size 4, an ORCA specific grid size, and the resolution of the identity (RI) approximation<sup>168–170</sup> are applied to the coulomb integrals with matching auxiliary basis sets (def2-TZVP/J). Effective core potentials (ECPs) from the Stuttgart-Dresden (SD)<sup>171–178</sup> collection are used for all fifth and sixth period elements. For the elements gold and mercury the default ECPs SD(60, MDF) differ from comparable TURBOMOLE 7.0.2 calculations and are therefore replaced by SD(60, MWB)<sup>171</sup>. The PBE0/def2-TZVP(-f) Hirshfeld charges serve as input for the computation of reference CM5 charges. The necessary empirical parameters for the XTB part are determined by fitting mainly to SCS-CC2<sup>179</sup>/aug-cc-pVTZ<sup>180</sup> vertical excitation energies. For transition metal containing systems TD-DFT(PBE0)/def2-TZVP data are used as reference. All excitation energies refer to vertical singlet states for given ground state equilibrium geometries. DFT calculations are performed with the `escf` module<sup>181</sup> of TURBOMOLE 7.0.2. For the SCS-CC2 calculations, the `ricc2` module<sup>182</sup> from TURBOMOLE 7.0.2 is utilized. In both cases no RI (density fitting) is used in the SCF equations, but in the `ricc2` part with matching auxiliary basis sets in the RI treatment<sup>183</sup> are applied (`cbas`). The fit set of molecules in the XTB part basically equals the one of the VTB part and contains about 380 reference values for the main group elements and 300 excitation states for the d-block elements (Y-Cd and La, Hf-Hg). Only low-lying and clearly assignable excitation energies are used as reference data. In some cases computed oscillator strengths are used for better assignment. About 10-20% of the molecules whose excited state assignments are unclear have been removed from the fit set.

### 4.3.2. Parameterization

The parameterization involves the determination of the atomic parameters  $H_s$ ,  $H_p$ ,  $H_d$ ,  $\zeta_s^0$ ,  $\zeta_p^0$ ,  $\zeta_d^0$ ,  $\zeta_D$ ,  $H_D$ , and EN (see below for a short description or Ref.<sup>76</sup> for more details). The actual number depends on the element and on the type of basis set (VBS or XBS). In the VTB part the number of atomic parameters varies between five and seven. In the XTB part the number of element-specific parameters is reduced to four to six because the determination of EN is not required. The necessary global empirical parameters were determined in former work<sup>76</sup> and remain unchanged. All atomic parameters of the transition

## II. Development and Application of the sTDA-xTB method

metal elements are obtained by linear interpolation regarding the nuclear charge  $Z$ , where only the start ( $Z=39$ , Y /  $Z=57$ , La) and end points ( $Z=47$ , Ag /  $Z=79$ , Au) are freely fitted. All remaining elements are fitted separately. For the group IIb elements Cd and Hg only the  $sp$ -shell is treated as valence whereas the  $d$ -electrons are omitted. Generally, a minimization of the root-mean-square deviation (RMSD) between calculated and reference data using the Levenberg-Marquardt (LM) algorithm<sup>184,185</sup> is conducted.

As noted before, a minimal basis set is used for the calculation of the atomic charges in the VTB part. Polarization functions are added to the main group elements with  $Z=31-34$ ,  $49-52$  and  $81-85$ . To avoid artificial states and to minimize the number of variable parameters, the exponents of these  $d$ -polarization functions are fixed to the value of the exponents of the  $p$ -functions. Then the element-specific parameters are derived by LM minimization between calculated VTB Mulliken-based CM5 charges and reference PBE0 Hirshfeld-based CM5 charges. The elements  $Z=1-17$ ,  $21-30$ ,  $35$  and  $53$  were parameterized in an earlier work<sup>76</sup> and their parameters served as initial guesses for homologous elements. For example the parameterization of the element gallium was started with fitted parameters of its lighter group homologue aluminum but with adapted principle quantum number  $n$ . Due to the linear interpolation of the atomic parameters of the transition metal elements with nuclear charge  $Z$ , the number of unknowns is reduced from 63 to 21 in the VTB part.

In the XTB part, a minimal valence basis set plus diffuse functions on hydrogen and main group non-metals is applied. As already mentioned, the parameters are fitted to reference values through minimization of the RMSD. Rydberg exponents ( $\zeta_D$ ) for fourth row main group non-metals are taken from Ref. 186 For fifth and sixth row main group non-metals the Rydberg exponents are determined during the parameterization process. As in the VTB parameterization, the element-specific parameters of the transition metals are interpolated with respect to their nuclear charge  $Z$ . Thus, the number of variables is reduced from 54 to 14 for each transition metal row. All excited state calculations are performed using the sTDA stand-alone program<sup>160</sup> employing the slightly adapted sTDA procedure. Here, the configuration selection threshold  $t_p$  is set to a value of  $10^{-4} E_h$  whereas the spectral range was limited to 10 eV (see Ref. 76 for details). The accuracy of the reference excitation energies with respect to experimental values is about 0.2 to 0.3 eV for closed shell systems<sup>179,187,188</sup> and hence, sTDA-xTB can achieve this accuracy at best.

In general, the dipole length formalism is used for the calculation of absorption spectra while the corrected<sup>76</sup> dipole velocity formalism is used for CD spectra. sTDA-xTB is used in combination with the generalized Born implicit solvent model with an solvent accessible surface term (GBSA)<sup>128,129</sup>.

## 4.4. Results and discussion

### 4.4.1. Charges and bond orders from the VTB Hamiltonian

The fit set of molecules for the parameterization of the partial atomic charges contains more than 740 closed shell systems, broken up into more than 420 molecules for the main group elements as well as approximately 320 structures for the d-block elements. In Figure 4.2 the results for these fit sets are depicted to enable a direct comparison of VTB(Mulliken) and reference PBE0/TZVP(Hirshfeld) computed CM5 charges. It can be seen that the reference charges are well reproduced by the VTB approach with a mean absolute deviation (MAD) of 0.04 and a standard deviation (SD) of 0.06 for the third to the sixth main group elements (statistical performance measures are given in units of electron charge). This trend also holds for the first and second main group elements (MAD: 0.04, SD: 0.06) as well as for the d-block elements with slightly larger MAD and SD values of 0.14 and 0.21, respectively. These enlarged deviations for the more polar d-block elements are probably caused by the interpolation scheme or the incomplete SCF procedure in the VTB step, which leads to an inaccurate treatment of very polar (ionic) environments. Additionally, the charges of these transition metals are typically larger on average.

Furthermore, chemically reasonable Wiberg bond order<sup>189</sup> values are obtained during the VTB procedure. A one to one comparison of VTB and reference PBE0/TZVP computed Wiberg bond orders for some main group element systems is depicted in Figure 4.3. Based on this plot, it is recognizable that mostly similar results are determined in both calculations. The MAD and SD values are slightly enlarged compared to the deviations obtained for CM5 partial charges, which have been used for fitting and seem to be more stable than the Wiberg bond orders. The latter ones serve as an additional crosscheck, and indicate the accuracy of the new method.

Overall, the simple VTB Hamiltonian can provide reasonably accurate charges as well as bond orders for a wide range of systems which contain main group or even transition metal elements. This is in line with results from the original paper<sup>76</sup>.



## II. Development and Application of the sTDA-xTB method

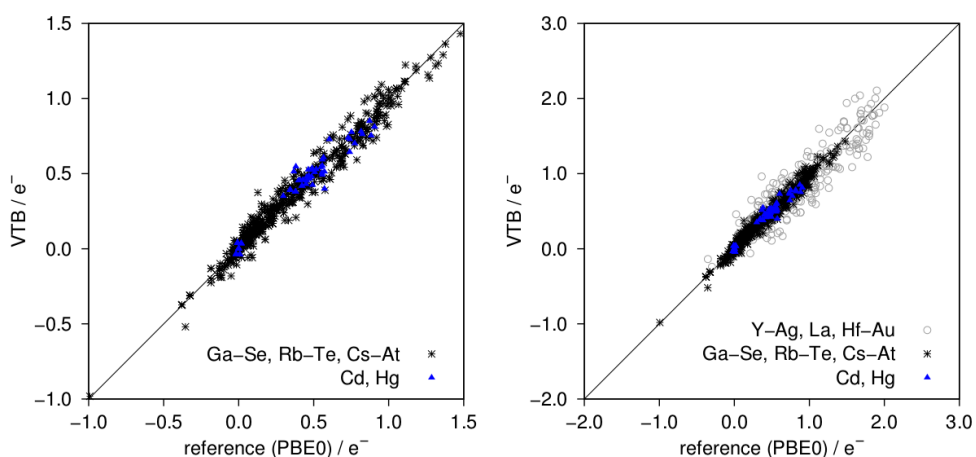


Figure 4.2.: Comparison of VTB and PBE0/TZVP data for CM5 atomic charges. The solid line denotes a one to one correspondence of the two data sets. Left: Charges of main group elements plus group IIb elements. Right: Charges for all parameterized elements including 4d- and 5d-block elements.

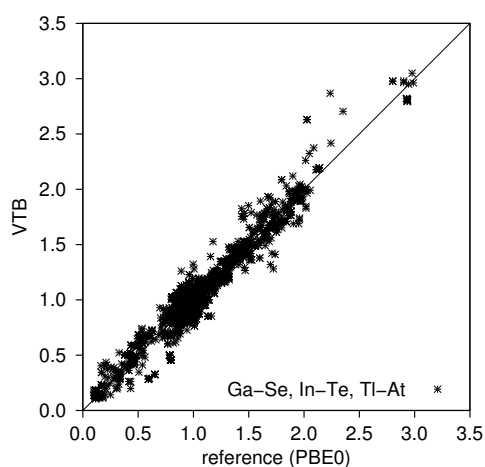


Figure 4.3.: Comparison of VTB and PBE0/TZVP data for Wiberg bond orders. The solid line denotes a one to one correspondence of the two data sets.

### 4.4.2. Excitation energies for the fit sets

The molecules used in the parameterization of the XTb part are identical to the VTb part except that in the former case roughly 10-20% had to be removed because of unclear excited state assignment. The final data are shown graphically

## II. Development and Application of the sTDA-xTB method

in Figure 4.4. As mentioned above, the accuracy of the reference excitation energies is about 0.2-0.3 eV (mostly for organic molecules) which limits the ideally achievable accuracy of sTDA-xTB MAD/SD values to this range.

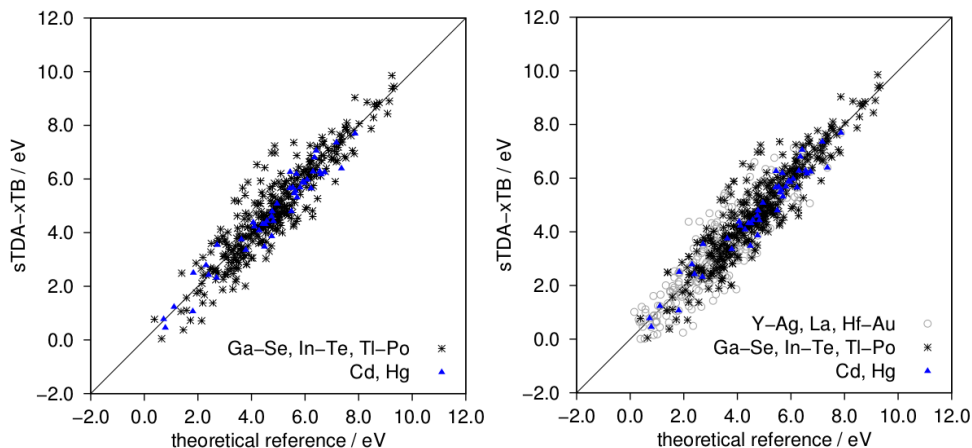


Figure 4.4.: Comparison of sTDA-xTB and SCS-CC2/aug-cc-pVTZ data for single excitation energies. The solid line denotes a one to one correspondence of the two data sets. Left: excitation energies of third and higher main group elements plus IIb elements. Right: excitation energies for all parameterized elements including 4d- and 5d-block elements.

Overall, the sTDA-xTB method performs reasonably well for third to sixth main group elements with MAD and SD values of only 0.53 eV and 0.70 eV, respectively. Outliers correspond either to very polar molecules or to electronically more complicated structures which are not very well described due to the incomplete SCF procedure in the sTDA-xTB scheme or fundamental problems of the pairwise TB Hamiltonian. The deviations for the d-block elements are only slightly larger with a mean absolute deviation of 0.58 eV and a standard deviation of 0.75 eV supporting the employed interpolation scheme.

### 4.4.3. UV and CD spectra in comparison to theory

In order to validate the obtained parameters, sTDA-xTB absorption spectra are calculated for various molecules not used in the training sets. These sTDA-xTB spectra are compared to corresponding sTD-DFT spectra, utilizing the hybrid density functionals BHLYP with the def2-SV(P) basis set of Ahlrichs and co-workers<sup>167</sup>. sTD-BHLYP and sTDA-xTB spectra are calculated up to 10 eV if not noted otherwise. For a more detailed information, see SI for additional technical

details.

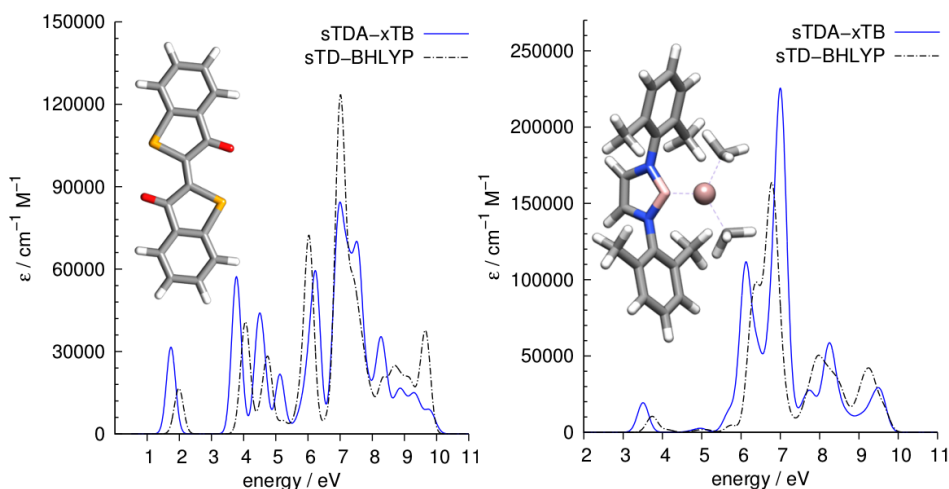


Figure 4.5.: Calculated absorption spectra of selenoindigo (left) and gallium containing complex (right) (sTDA-xTB: blue solid line, sTD-BHLYP: black dashed line, TD-BHLYP: gray solid line and red dotted line). The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV (sTDA-xTB) and 0.3 (sTD-BHLYP).

As a first example, two typical absorption spectra are depicted in Figure 4.5. It shows the spectra of selenoindigo and a gallium containing complex. For both molecules the sTDA-xTB calculated spectrum is in very good agreement with the sTD-BHLYP absorption spectrum. Each visible band in the sTD-DFT spectrum is also observable in the sTDA-xTB spectrum. Additionally, the absolute band intensities are comparable, which indicates that the STO exponents in the XTb part are also reasonable.

As a second comparison, Figure 4.6 shows the computed absorption spectra of arsenobetaine, a common substance in marine biological systems. Comparable to the first two examples, sTDA-xTB performs remarkably well in reproducing the reference spectra which are in this case simplified TD-DFT and the regular variant. The band at 179 nm is in case of sTDA-xTB quite intense. This could be attributed to the more diffuse basis set used in the sTDA-xTB method. Test computations with TD-DFT using a minimal augmented basis show, that this band is sensitive to such basis functions with diffuse character. Note that the entire sTDA-xTB calculation for the arsenobetaine systems for all excited states up to 15 eV takes only about one second on a common desktop computer whereas on the same machine the sTD-DFT treatment takes 90 times longer and the full TD-

## II. Development and Application of the sTDA-xTB method

DFT calculation 3720 times longer (7200 times longer for the minimal augmented basis).

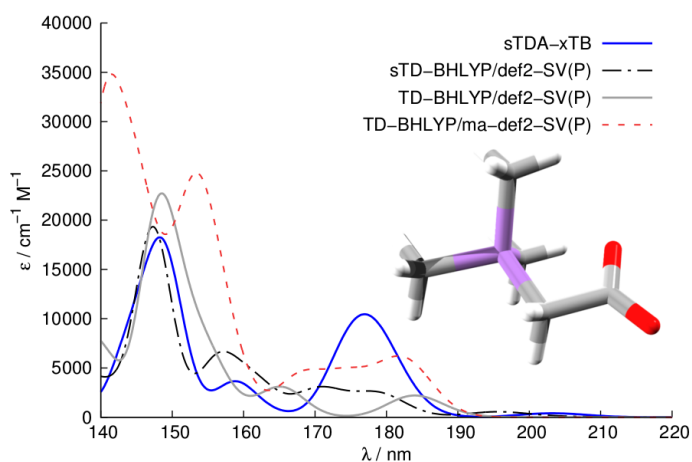


Figure 4.6.: Calculated absorption spectra of arsenobetaine (sTDA-xTB: blue solid line, sTD-BHLYP: black dashed line, TD-BHLYP: gray solid line and red dotted line). The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV (sTDA-xTB) and 0.7 (sTD-BHLYP).

As a fourth and final theory to theory comparison, Figure 4.7 shows the computed CD spectra of a bismuth and indium containing complex. The crystal structure of this complex was presented in Ref. 190. In comparison to the sTD-DFT computed spectrum, sTDA-xTB performs remarkably well. Almost all bands are reproduced with just a slight shift or minimal change in intensity. Such a rather unique system, containing silicon, indium and bismuth is not a trivial case for a semiempirical theory. In addition, it is noted that this calculation is 380 times faster than sTD-DFT for this molecule.

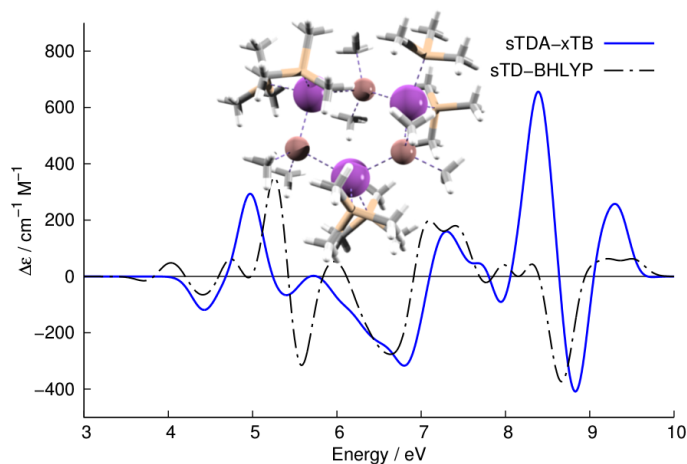


Figure 4.7.: Calculated absorption spectra of bismuth and indium containing complex<sup>190</sup> (sTDA-xTB: blue solid line, sTD-BHLYP: black dashed line, TD-BHLYP: gray solid line and red dotted line). The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at 1/e maximum of 0.5 eV and all spectra are red-shifted by 0.5 eV.

#### 4.4.4. UV and CD spectra in comparison to experiment

In the synthesis of chiral compounds, CD spectra are often measured to obtain information about an absolute configuration. Theoretically calculated spectra of the two enantiomers can be compared to experiment in order to be able to make an assignment. However, such calculations are often very expensive for larger systems and routine use has not yet taken place. The need for a method that can deliver such spectra very quickly and does not require special computer hardware is therefore high. After evaluating the performance against other theoretical methods, sTDA-xTB here will be compared to experimentally measured spectra.

## II. Development and Application of the sTDA-xTB method

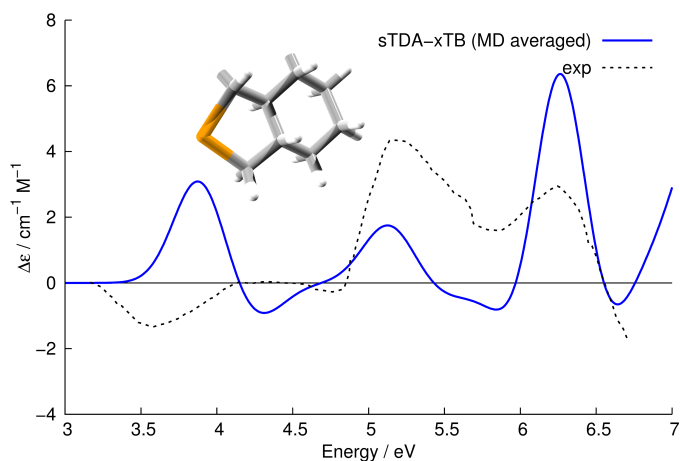


Figure 4.8.: Calculated CD spectrum of (2)-(3aS,7aS)-2-tellura-trans-hydrindan (blue solid line) averaged along a GFN-xTB computed MD trajectory. The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at 1/e maximum of 0.5 eV. The experimental curve (black dotted line) are taken from Ref. 191 and refer to a solution of (2)-(3aS,7aS)-2-tellura-trans-hydrindan in acetonitrile.

As a first example, Figure 4.8 shows the CD spectra of (2)-(3aS,7aS)-2-tellura-trans-hydrindan<sup>191</sup>. For a similar compound, substituting tellurium by its lighter homologue selenium, the CD spectra are shown in Figure A2.1 in the appendix. For both molecules, the common (vertical) single structure approach is not sufficiently accurate. In the frozen conformation of a minimum structure, transitions can have artificial rotatory strengths. In order to eliminate this problem MD simulations at room temperature were carried out. After sampling along the resulting trajectory, a spectrum is averaged from 100 equidistantly taken snapshots. The overall agreement to the experimental spectrum improved compared to the static case and is now quite good. Only the first band at 3.9 eV has a wrong sign. Despite the fact that the intensities for this molecule are very small ( $|\Delta\epsilon| < 6$ ), a clear assignment to this enantiomer can be given, which also holds for the selenium-hydrindan.

## II. Development and Application of the sTDA-xTB method

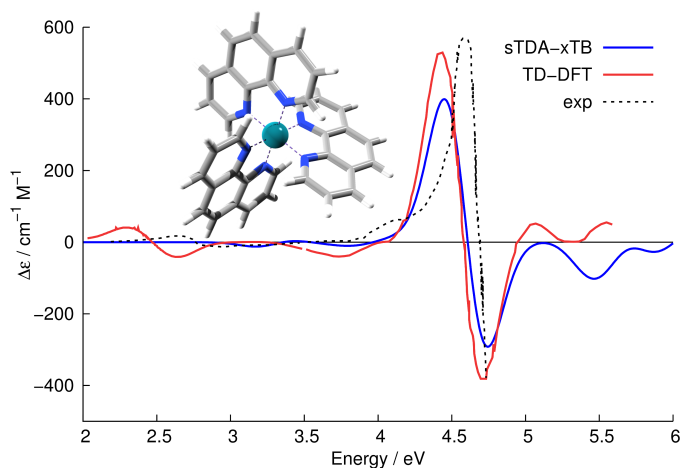


Figure 4.9.: Calculated CD spectrum of  $\Lambda$ -Ru(phen) $_3^{2+}$  (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and are red-shifted by 0.5 eV. The experimental (black dotted line) and TD-DFT (red solid line) spectra are taken from Ref. 192.

The second theory to experiment comparison is shown in Figure 4.9. The CD spectra of  $\Lambda$ -Ru(phen) $_3^{2+}$  are presented. Most of the characteristic features of the spectrum are reproduced by sTDA-xTB. This kind of transition metal complexes ( $[M(L)_3]^{n+}$  ( $M = \text{Fe, Ru, Os, Co, Rh, Ir}$ ;  $n=2,3$ ;  $L = 1,10$ -phenanthroline, 2,20-bipyridine)) were studied by Autschbach et al.<sup>192,193</sup>. As an additional comparison, the TD-DFT computed spectrum from Ref.<sup>192</sup> is provided. Except for the low-energetic part of the spectrum, where metal-ligand charge transfer excitations are assigned<sup>193</sup>, the two theoretical spectra match very well. For an assignment of an absolute configuration, the two spectra yield the same result. In Figure A2.2 in the appendix, the CD spectrum of  $\Lambda$ -Os(phen) $_3^{2+}$  is shown. The same conclusions made above hold for this complex as well.

To test the new sTDA-xTB parameter set even further, a third system is investigated. Figure 4.10 shows the CD spectra of a palladium-bis(helicene) complex. All characteristic features of the experimental spectrum are very well reproduced. Even the low lying band at 450 nm which is characterized to have a high charge transfer contribution<sup>194</sup>, is slightly shifted but clearly assignable in the computed spectrum.

## II. Development and Application of the sTDA-xTB method

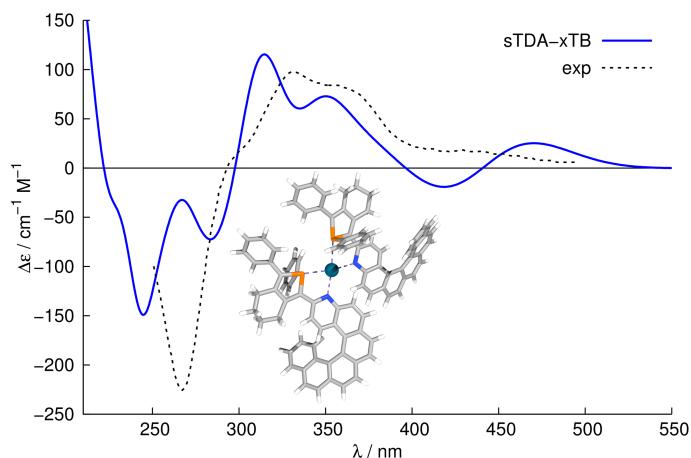


Figure 4.10.: Calculated CD spectrum of a metal bis-helicene assembly (blue solid line). The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at 1/e maximum of 0.5 eV and are red-shifted by 0.5 eV. The experimental (black dotted line) spectrum is taken from Ref. 194.

Up to now, the computation of UV and CD spectra for large molecules containing hundreds to thousands of atoms was not feasible with standard TD-DFT approaches. The single excitation space becomes extremely large for such systems and standard treatments are typically restricted to the computation of a small number of excited states.<sup>195</sup> Additionally, minimum structure approaches are often not sufficient to treat flexible complexes accurately. A possible way to go beyond the minimum structure approach is to compute spectra on structures obtained from MD simulations, as conducted for the tellura-hydrindan system. As further examples where this technique is required we examine two flexible and large palladium complexes<sup>196</sup>. The chosen approach is based on a similar MD scheme as it has already been used for large biomolecular systems<sup>197</sup>.



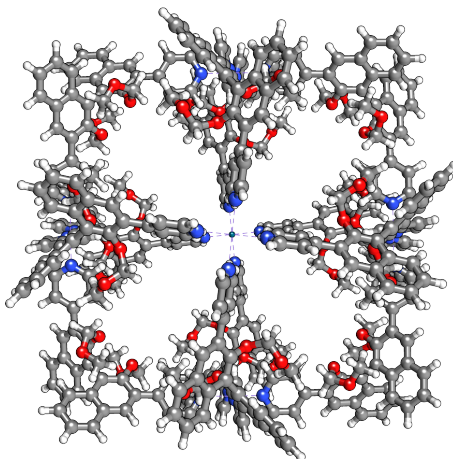


Figure 4.11.: Adopted structure of  $\text{Pd}_6\text{L}_{12}^{+12}$  (HF-3c geometry)<sup>196</sup>.

The  $\text{Pd}_6\text{L}_{12}^{+12}$  complex is depicted in Figure 4.11 and consists of 822 atoms with a molecular charge of +12. Calculated absorption and CD spectra are given in Figure 4.12. The CD spectrum of the minimum structure, computed with sTDA-xTB, is not in good agreement with the experimental one. The spectrum is blue-shifted by 0.5-1.0 eV. Nevertheless, all peaks are present and can be assigned to experimentally detected ones. The MD averaged sTDA-xTB spectrum coincides better with the experiment regarding the band positions while band intensities are somewhat too low and are thus scaled by a factor of two. Similar results are found for the absorption spectra on the right of Figure 4.12 where the experimental results are also scaled by a factor of two.

## II. Development and Application of the sTDA-xTB method

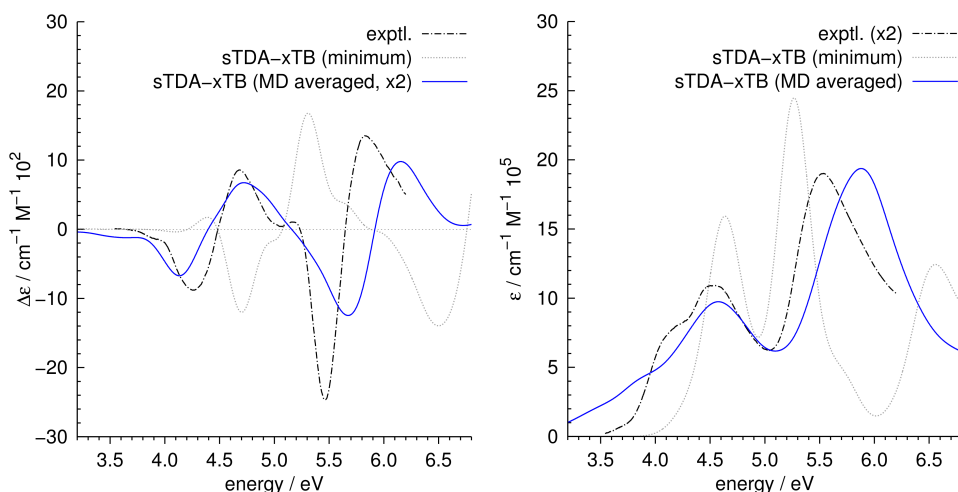


Figure 4.12.: Comparison of sTDA-xTB (minimum structure), MD averaged sTDA-xTB and experimental spectra for  $\text{Pd}_6\text{L}_{12}^{+12}$ . Left: CD spectra. The calculated intensity of the MD averaged sTDA-xTB spectrum is scaled by a factor of two. Right: Absorption spectra. The intensity of the experimental spectrum is scaled by a factor of two. The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at  $1/e$  maximum of 0.5 eV and are red-shifted by 0.5 eV.

Figure 4.13 shows a complex, which is twice as big as the former one with a charge of +24 and consisting of 1644 atoms. To the best of our knowledge, sTDA-xTB is the only quantum chemical method available which routinely can compute electronic spectra for such large systems. The calculated spectra are given in Figure 4.14. Looking at the absorption spectra on the right, it can be seen that the minimum structure does not yield an absorption spectrum matching the experimental one well. Like for the smaller complex it is blue-shifted by  $\sim 1$  eV. However, a total of three bands is observable, similar to the experiment. The MD averaged sTDA-xTB spectrum is in good agreement with the experimental one. Band positions and intensities are in good mutual agreement. Additionally, a reasonably good result is found for the MD averaged CD spectra shown on the left. The first peak is a bit too negative compared to the experiment, but the bands at 4.2 eV and 5.6 eV are in good agreement. Additionally, the slightly positive peak at around 5.2 eV exists as shoulder. The minimum structure spectrum still appears blue-shifted even though both theoretical spectra are red-shifted by 0.5 eV. This is similar to the discussion in the original work<sup>196</sup>. In principle, the treatment of these system is due to the high molecular  $O$  symmetry even possible with sTDA-DFT. But when using the symmetry of the system the Pd ligand sphere

## II. Development and Application of the sTDA-xTB method

has an artificial chirality. To overcome this issue, sampling along a MD trajectory is performed. However, the number of spectra to be calculated does not allow the use of higher level methods.

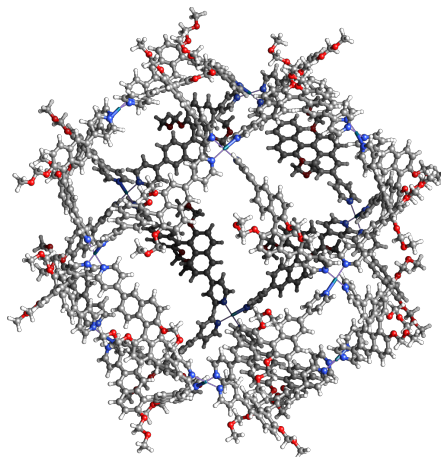


Figure 4.13.: Adopted structure of  $\text{Pd}_{12}\text{L}_{24}^{+24}$  (HF-3c geometry).

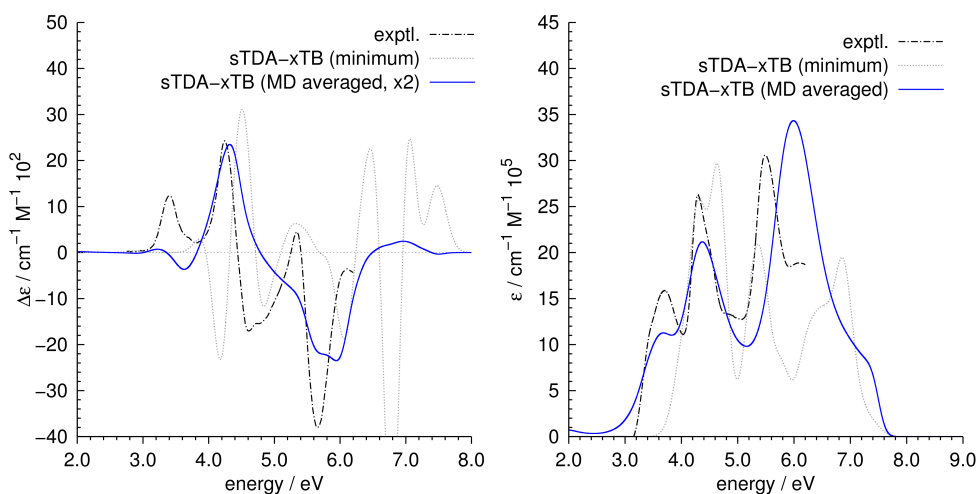


Figure 4.14.: Comparison of sTDA-xTB (minimum structure), MD averaged sTDA-xTB and experimental spectra for  $\text{Pd}_{12}\text{L}_{24}^{+24}$ . Left: CD spectra. The calculated intensity of the MD averaged sTDA-xTB CD spectrum is scaled by a factor of two. Right: Absorption spectra. The individual transition oscillator strengths are broadened by Gaussian shape functions with a full width at  $1/e$  maximum of 0.5 eV and are red-shifted by 0.5 eV.

For both complexes, slight blue-shifts are observed for bands above 5 eV (below 250 nm). These may be attributed to different aspects like the incomplete basis

## II. Development and Application of the sTDA-xTB method

set or neglected double excitations as well as the solvation treatment.<sup>196</sup> Nevertheless, the deviations are in the expected range based on the accuracy of larger systems ( $MAD \approx 0.4\text{-}0.5$  eV). However, the blue-shifts have to be considered in the peak assignment. In addition, it is noticeable that MD averaged spectra are shifted to lower energies compared to minimum spectra. This well known red-shift effect occurring in MD averaged spectra results from stretched bonds by thermal excitation in the MD simulation.

### 4.5. Conclusions

The extension of the parameter set for sTDA-xTB to almost all elements is presented. The good performance of sTDA-xTB compared to simplified TD-DFT is clearly visible from the presented examples. Furthermore, theory to theory comparisons show the potential of this method for computing spectra, e.g., to assign the absolute configuration of a molecule. The main area of application of this method is where other, higher level methods are computationally not feasible. This could be in cases, when large sets of molecules have to be screened or spectra of large molecules ( $> 1000$  atoms) have to be calculated. Those larger molecules are now for the first time within reach of theoretical models without using fragmentation schemes. Even sampling along a trajectory of a system with 2500 atoms is feasible<sup>197</sup> The computational speed of the sTDA-xTB method also allows quick checks during a computational chemistry project. As already shown for a variety of systems, there is good agreement of the computed spectra to experiments. The qualitative characterization of excited states or the determination of the absolute configuration is a routine task for sTDA-xTB. The GBSA solvation model not discussed here in detail is an essential part of the treatment for molecules in solution. Its speed is tailored to semiempirical models, allowing an efficient, reasonably accurate solvation treatment.

The drawbacks of sTDA-xTB are mostly assignable to the monopole approximation used to efficiently compute two-electron integrals. Very local (atomic) states as well as sometimes charge-transfer excitations are not perfectly described by this approximation. Therefore, the revision of the electrostatic description represents the first step towards an improved version in the future, maybe by including higher multipole moments.

sTDA-xTB and the whole xTB framework is not only targeted to specialists in theoretical and computational chemistry, but rather designed to be used by differently skilled users. The speed of this method allows to run on regular desktop or laptop hardware and molecules containing up to several hundreds of atoms can be routinely investigated on such machines. The use of this method by instrumentalists as well as theoreticians will hopefully boost cooperations in the future.

## **Disclosure statement**

The authors declare no competing financial interest.

## **Funding**

This work was supported by the DFG in the framework of the Gottfried-Wilhelm-Leibniz prize.



# 5. Benchmarking of structural approaches for an efficient quantum-mechanical protein spectra calculation

## 5.1. Introduction

CD spectroscopy is a very important and widely applied technique in biochemistry and medical research.<sup>31,32</sup> Together with nuclear magnetic resonance (NMR) and X-ray diffraction, CD spectroscopy plays an important role in structure elucidation of proteins or DNA<sup>21</sup>. CD is a much less ambitious technique, compared to X-ray crystallography, both in terms of sample concentration and preparation time.<sup>30</sup> While this is also true for NMR, the latter does not resolve the absolute configuration of a molecule. Thus, CD has become an indispensable tool in biochemists' daily routine. The far UV spectral region of protein absorption is dominated by electronic transitions in the peptide bonds. The respective CD spectra are highly sensitive to the protein secondary structure<sup>49</sup>. As already described in chapter 1, different secondary structure motifs have characteristic CD spectra. This allows for *in-vitro* studies of protein folding, which are crucial for diseases caused by prions like bovine spongiform encephalopathy (BSE)<sup>34–36</sup>. CD is furthermore capable of differentiating tertiary structures, e.g., between mutant and wild type proteins from the near UV region.<sup>27</sup>

As outlined in chapter 1, the theoretical simulation of CD spectra for large biomolecular systems is one of the main aims of this thesis. Chapter 3 already investigated the applicability of the sTDA-xTB method for the calculation of the excited state spectra of large biomolecular systems. However, the structural methods in this chapter ranged from specialized FFs, over SQMs, up to *ab-initio* WF theory. This chapter aims at finding an efficient structure method that can be applied to entire proteins without any fragmentation schemes. A further premise is the applicability of the structure method to a wide range of chemical elements and the absence of any topology based parameterizations, excluding common protein FFs. Due to the size of the target systems – typically larger than 1000 atoms – *ab-initio* WFT and *first-principles* DFT are no viable levels of theory for this

## II. Development and Application of the sTDA-xTB method

purpose. SQMs and universal FFs are reasonable choices for the structure generation for excited state calculations. GFN1-xTB<sup>74</sup> is used for some of the systems in chapter 3, showing its potential for protein structure optimization. GFN2-xTB, as successor to GFN1-xTB, uses a multipole electrostatic treatment for the electronic energy and a density based dispersion correction, superseding GFN1-xTB in almost any cases.<sup>75</sup> In this chapter, GFN2-xTB is thoroughly tested as a candidate for an efficient structure method for biomolecular structure optimizations. Chapter 3 and 4 show very clearly that for some systems the sampling along an MD trajectory is essential for a good reproduction of experimental spectra. However, the conduct of protein MD simulations on an SQM level of theory can take weeks to months, depending on the system size and the simulation time. FF are the methods of choice for this purpose, due to their speed since they only apply an atomistic treatment instead of an electronic treatment as in SQM, WFT, or DFT. The very recently published GFN-FF is an excellent candidate for the structure optimization for biomolecular systems.<sup>114</sup> It is an universal FF, but with remarkable performance, comparable to SQM and specialized FF.<sup>114</sup> The combination of GFN-FF as the structure method and sTDA-xTB for the calculation of CD is comprehensively tested in this chapter.

The chapter is organized as follows: First, the computational details of this chapter are provided. Then, the GFN2-xTB and GFN-FF structures are tested for the computation of CD spectra of model peptides with common and rare secondary structure motifs. Consecutively, the CD spectra of several proteins are computed, based on GFN-FF and GFN2-xTB optimized structures and the results are discussed. In the end, the results are summarized and concluded.

### 5.2. Computational Details

As initial structure of the proteins the X-ray diffraction structures from the protein database (PDB)<sup>134</sup> were used. Hydrogen atoms were added using the maestro program<sup>198,199</sup>. The proteins, computed with GFN-FF<sup>114</sup> were explicitly solvated with a 6 Å water shell, using the TIP3P water model<sup>200</sup>, as implemented in the UCSF Chimera package<sup>201</sup>. The resulting atom numbers and charges of all model peptides and proteins are displayed in table 5.1.



Table 5.1.: Name, PDB ID, charge, atom count of the peptide (#At), atom count of the peptide plus solvent shell (#At +solv) and number of residues of the treated systems.

name	PDB ID	charge	#At	#At+solv	# residues
[L-Lys] <sub>10</sub>	N/A	0	213	825	10
collagen	4Z1R	0	957	2628	90
Ac-[L-( $\alpha$ -Me)Val] <sub>8</sub> -OtBu	N/A	0	183	642	8
avidin	1RAV	6	1944	4334	124
monellin	1MOL	2	1562	3317	94
scorpion toxin	4KYP	-1	1082	2549	73
FATC domain	2KIT	-2	545	1595	33

For all geometry optimizations and MD simulations, the `xtb` program (version 6.2.2) is used.<sup>140</sup> For the structure optimization with GFN2-xTB, the implicit GBSA<sup>128,129</sup> water model is applied. In the case of GFN-FF, an adapted GBSA parameterization is not yet available, and therefore, the structures are optimized with an explicit water shell. For the geometry optimization, the convergence threshold for energy of  $5 \cdot 10^{-6}$  Eh and gradient norm of  $1 \cdot 10^{-3}$  Eh $\cdot\alpha^{-1}$  is applied (level normal in `xtb`). The MD simulations were performed with preceding energetic equilibration and an applied time step of 1 fs. The SHAKE<sup>202,203</sup> algorithm was used to constrain all hydrogen-containing bonds.

If not stated otherwise, the experimental CD spectra are taken from the protein circular dichroism database (PCDDDB)<sup>204</sup>. The theoretical spectra are computed at the sTDA-xTB level of theory<sup>76</sup>. For this purpose, the electronic ground state is computed with the `xtb4stda` program with applied implicit GBSA solvation for water. The excited state is then computed with the `stda` program<sup>205</sup>. For the energy range of the configuration states, a threshold of 10 eV is applied. To ensure origin independent properties, the velocity formalism is used for the rotatory strengths. For an unbiased comparison of CD spectra, the computed spectrum is shifted to match a distinctive experimental band. If not stated otherwise, the computed spectra is red-shifted by 0.5 eV. The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV.

### 5.3. Results and Discussion

This section comprises the results of computing CD spectra with sTDA-xTB on GFN2-xTB and GFN-FF structures. First, some secondary structure motifs like the  $\alpha$ -helix,  $3_{10}$ -helix and collagen PII-helix are investigated. Secondly, the CD spectra of entire proteins are analyzed.

#### 5.3.1. Model peptides

The first model peptide is poly-lysine with 10 residues in an  $\alpha$ -helix conformation ( $[L\text{-Lys}]_{10}$ ). The  $\alpha$ -helix and the  $\beta$ -sheet are the most commonly occurring secondary structure motifs and have been discovered first by Pauling and Corey.<sup>23</sup> The  $\alpha$ -helix is stabilized by a hydrogen bond network between the backbone atoms, generating a 13-membered ring. Hence, the  $\alpha$ -helix is also systematically denoted as  $3.6_{13}$ -helix, since the average number of residues per helical turn is 3.6.

Already in the original publications of sTDA-DFT and sTDA-xTB, the spectrum of an  $\alpha$ -helical model peptide was presented with great agreement with the experiment. In chapter 3, many proteins comprise large amounts of  $\alpha$ -helical arrangements. Also there, the sTDA-xTB shows great performance for this type of secondary structure. To test the performance of GFN-FF, the model peptide is optimized and an MD simulation of 500 ps is performed. On the single optimized structures and on 100 equidistantly taken snapshots from two different MD simulations (20 ps and 500 ps), the CD spectrum is computed by sTDA-xTB. The calculated and experimental spectra are depicted in Figure 5.1.

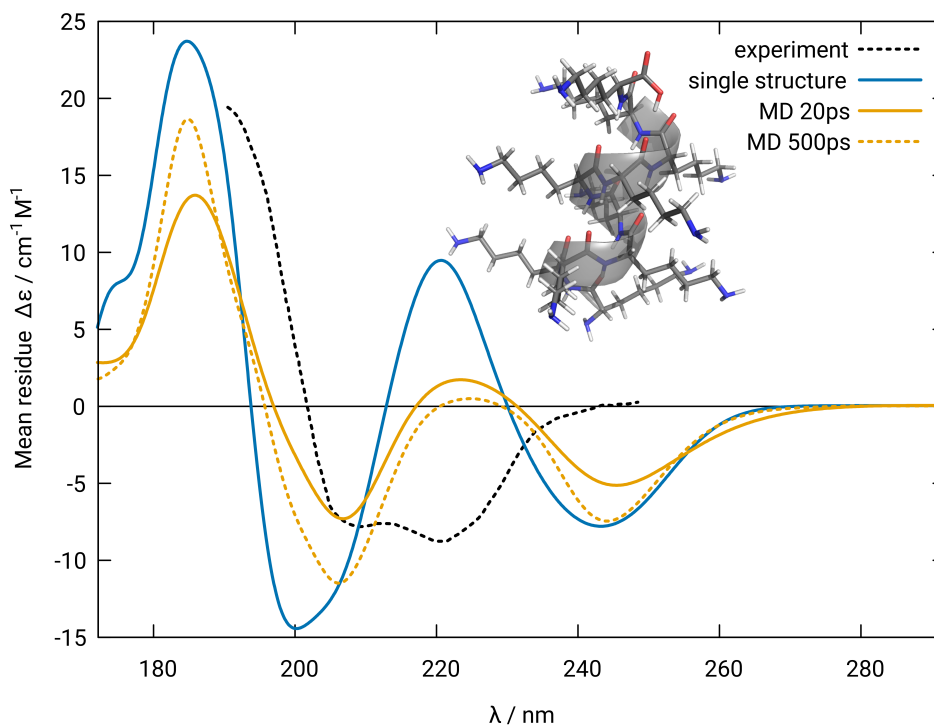


Figure 5.1.: ECD spectrum of  $[L\text{-Lys}]_{10}$  computed with sTDA-xTB on a single optimized (blue) and MD (orange) structures. The experimental spectrum (black) refers to a solution of polylysine in water.<sup>28</sup>

The CD spectra, computed on GFN-FF structures are in good agreement with the experiment. The three characteristic bands of the  $\alpha$ -helical<sup>206</sup> secondary structure are reproduced, i.e., are a positive band at 190 nm (perpendicular  $\pi \rightarrow \pi^*$  transition) and two negative bands at 208 nm and 222 nm, respectively (parallel  $\pi \rightarrow \pi^*$  and  $\pi \rightarrow \pi^*$  transitions)<sup>207–210</sup>. However, the single optimized structure has an additional positive band at 220 nm that is not present in the experimental spectrum that decreases in the averaged MD spectrum. The averaged MD spectrum from the full 500 ps simulation shows the best agreement with the experimental spectrum. This is in agreement with the findings of chapter 3 and 4: in some cases, the MD sampling is crucial for an accurate reproduction of experimental spectra.

The second investigated secondary structure is the  $3_{10}$ -helix. The systematic notation of the helix indicates that the with three residues a full helical turn is reached and a 10-membered ring is build from the hydrogen bonds. Thus, the  $3_{10}$ -helix is more tightly wound in comparison to the  $\alpha$ -helix. A structural comparison of the two helices is given in Figure 5.2.

## II. Development and Application of the sTDA-xTB method

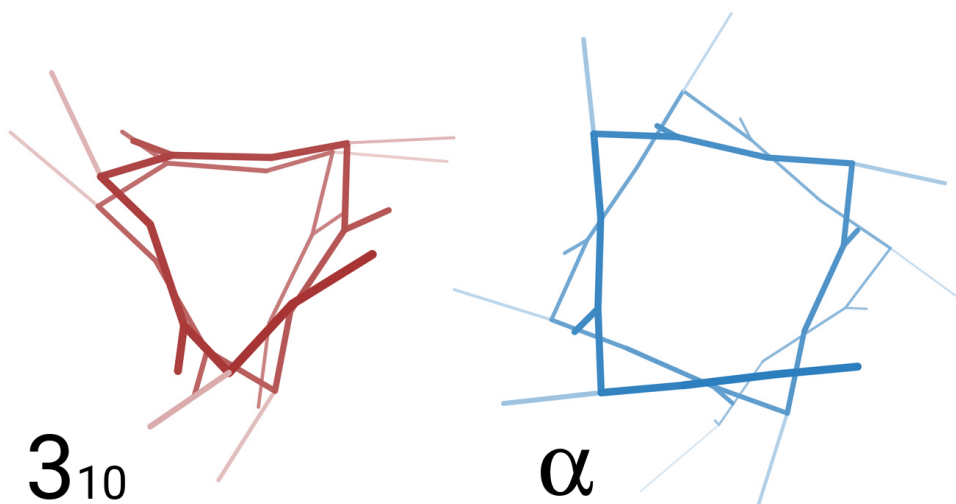


Figure 5.2.: Comparison of  $3_{10}$  (triangular, red) versus  $\alpha$  (offset square, blue) helix. Top view of backbone atoms and sidechains out to  $C_{\beta}$ .

The  $3_{10}$ -helix constitutes 10-15% of all helical arrangements in proteins and extends often  $\alpha$ -helical secondary structures.<sup>211</sup> As a model for the  $3_{10}$ -helix the molecule Ac-[L-(\alpha-Me)Val]\_8-OtBu is used.<sup>212</sup> The model peptide is optimized with GFN-FF and an MD simulation of 20 ps is performed. The computed CD spectra for the single minimum structure and the MD average, and the experimental spectrum is shown in Figure 5.3.

The experimental spectra of the  $\alpha$ - and  $3_{10}$ -helix differ strongly. The CD intensities are much weaker for the more compressed  $3_{10}$ -helix. Because of the deviations between both helices, CD spectroscopy can be used to characterize these secondary structures in solution. The computed spectrum on the GFN-FF optimized minimum structure gives the correct position of the CD bands, but with too strong intensity. Also for this system, the MD sampling improves the result and the averaged spectrum over 100 snapshots is in excellent agreement with the experiment. This shows very clearly the potential of GFN-FF, since the MD simulation of the model peptide only takes 1.5 h on a quad core computer (Intel® Xeon® CPU E3-1270 v5 @ 3.6GHz). GFN2-xTB is on the other hand more than three orders of magnitude slower,<sup>114</sup> making MD sampling of larger biomolecular systems at SQM level a time-intensive task.

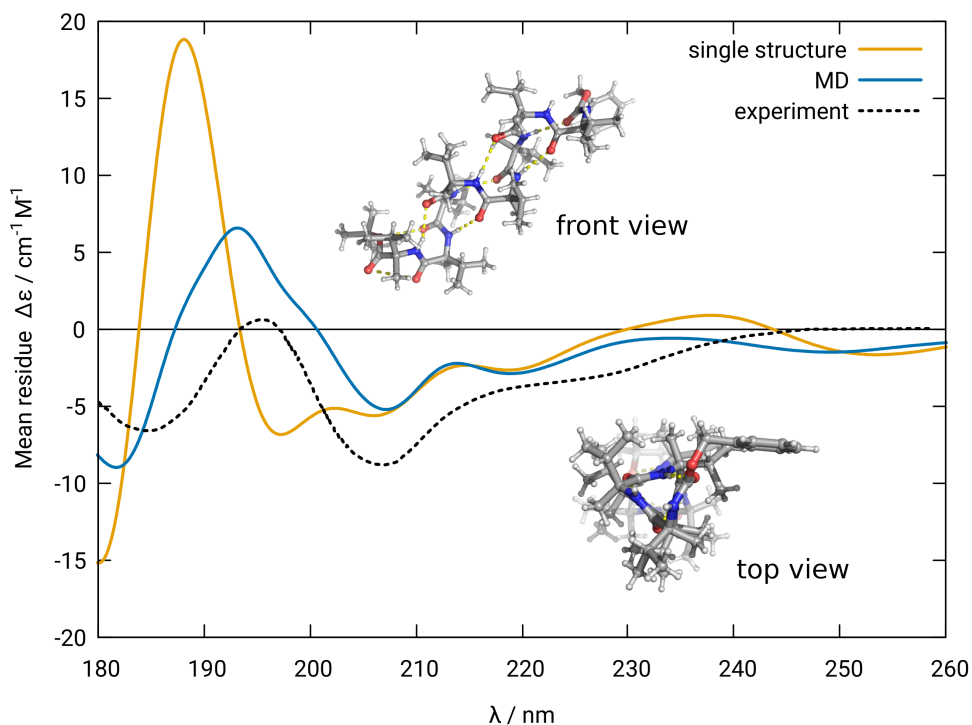


Figure 5.3.: ECD spectrum of Ac-[L-( $\alpha$ -Me)Val]<sub>8</sub>-OtBu computed with sTDA-xTB on GFN-FF structures (orange and blue). The experimental spectrum (black) refers to a solution of Ac-[L-( $\alpha$ -Me)Val]<sub>8</sub>-OtBu in trifluoroethyl alcohol (TFE).<sup>212</sup>

The next secondary structure motif is the collagen triple helix. It consists of three P<sub>II</sub>-type helices with the repetitious amino acid sequence of glycine - X - Y (with X and Y being frequently occupied by proline or hydroxyproline). The three strands wrap parallel around a common axis forming a right-handed superstructure that is stabilized by interchain hydrogen bonds. The experimental CD spectrum shows characteristic, but weak positive band at 226 nm ( $n \rightarrow \pi^*$  transition) and a strong, negative band at 206 nm ( $\pi \rightarrow \pi^*$  transition).<sup>213</sup> Due to the rich content of tertiary amides, the transitions are red shifted in comparison to secondary amides, e.g., in  $\alpha$ - or  $3_{10}$ -helices.

The crystal structure of the collagen-like peptide from Ref. 214 serves as structural model for the collagen secondary structure motif. The initial structure is optimized with GFN-FF and GFN2-xTB. The sTDA-xTB computed and experimental spectra are depicted in Figure 5.4.

## II. Development and Application of the sTDA-xTB method

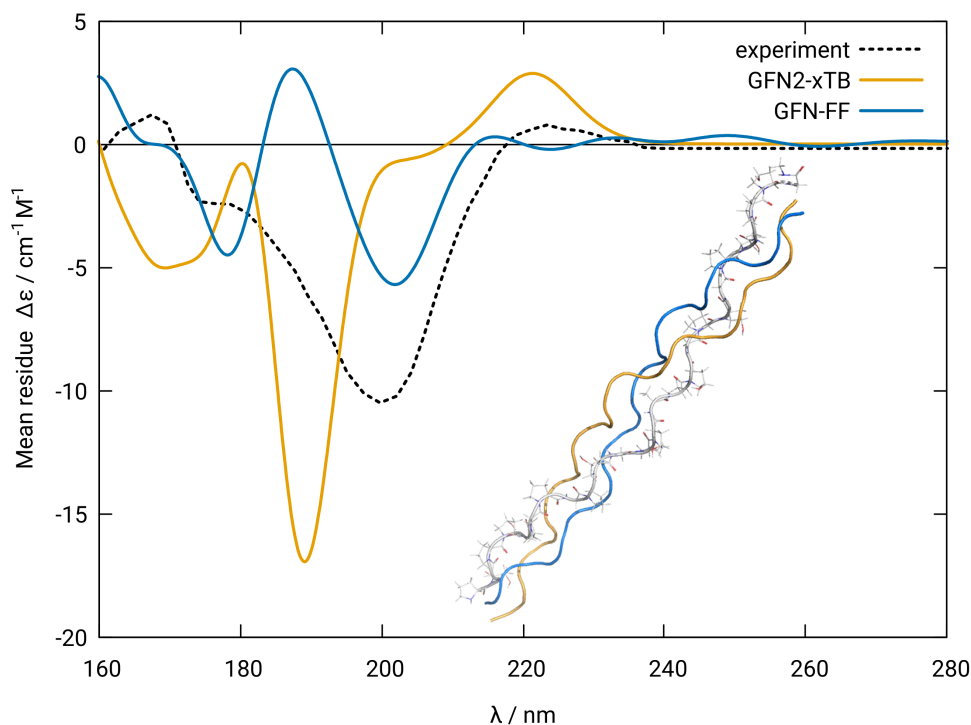


Figure 5.4.: ECD spectrum of a collagen model peptide computed with sTDA-xTB on GFN-FF (blue) and GFN2-xTB (orange) optimized geometries. The experimental spectrum (black) refers to a solution of collagen in water.<sup>24</sup>

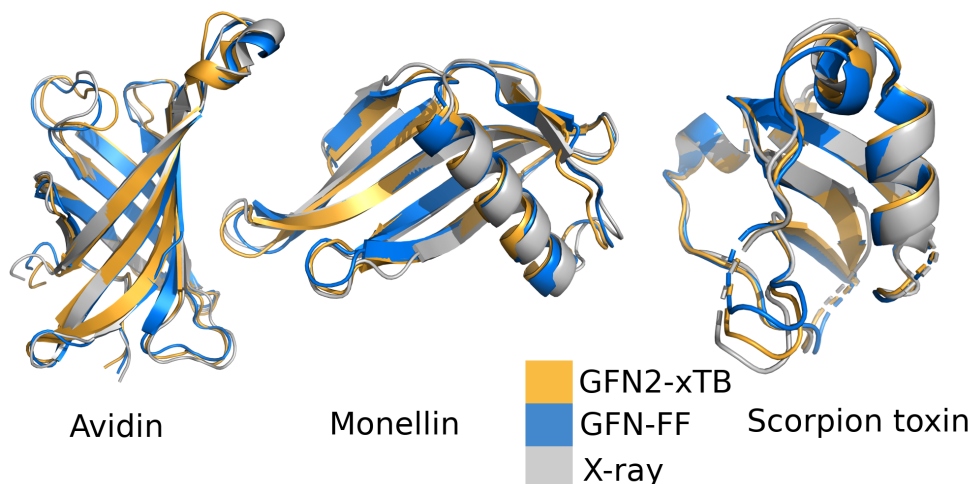
The computed spectrum on the GFN-FF optimized structure shows in principle all characteristic CD bands, but with wrong relative intensities. The negative band at 206 nm shows a too weak intensity and the shoulder in the experiment at 180 nm corresponds to a positive band at 190 nm in the GFN-FF spectrum. The spectrum based on the GFN2-xTB structure is in very good agreement with the experiment. The bands at 226 and 206 nm are slightly too intense. For a better agreement with experiment, the computed rotatory strengths could be scaled. As already discussed in chapter 3, this is not unusual for systems of this size (957 atoms). All examples of this sections show that GFN-FF has huge potential for providing optimized structures for the calculation of peptide CD spectra. Due to its low computational cost and good accuracy, it is perfectly suited in combination with sTDA-xTB.

### 5.3.2. Proteins

The next step is the calculation of CD spectra of entire proteins. This is a challenging task due to the high structural sensitivity of the CD spectroscopy and the

## II. Development and Application of the sTDA-xTB method

flexibility of polypeptide systems. Before comparing computed with experimental spectra, the optimized structures are evaluated with respect to the crystal structure. The structural overlays of three proteins from this section are depicted in Figure 5.3.2.



The comparison of the secondary structure, determined by X-ray crystallography and theoretical calculations, shows that rigid, intrinsically stabilized motifs (like helices and sheets) are conserved very well. The areas of highest deviations from the crystal structure are the unstructured loops. This is in accordance with previous, more comprehensive analyses of GFN-FF and GFN2-xTB for the optimization of protein structures.<sup>114,215</sup> The effect of these minor deviations on the computed CD spectra is tested in the following.

First, the sweet protein monellin – discovered in 1972<sup>216</sup> – is investigated. Monellin has two noncovalently connected polypeptide chains. The protein secondary structure of this protein comprises five  $\beta$ -strands that form an antiparallel  $\beta$ -sheet and a 17-residue  $\alpha$ -helix.<sup>217</sup> The experimental spectrum and the computed spectra on GFN-FF and GFN2-xTB optimized structures are depicted in Figure 5.5.

## II. Development and Application of the sTDA-xTB method

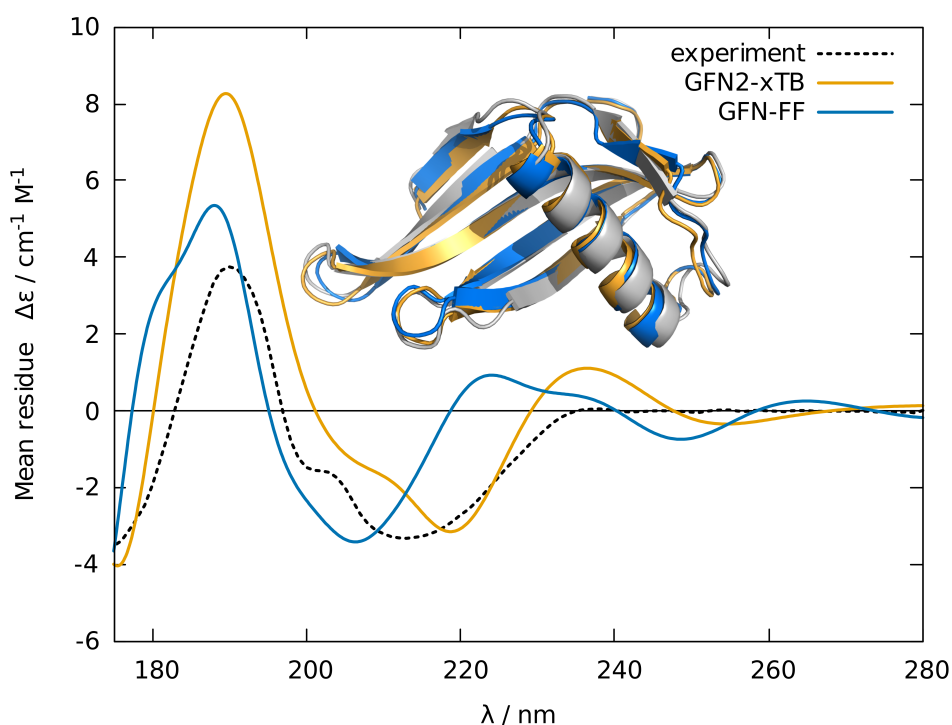


Figure 5.5.: ECD spectrum of monellin (PDB: 1MOL), computed with sTDA-xTB on optimized GFN-FF (blue) and GFN2-xTB (orange) structures. The experimental spectrum is taken from Ref. 204 and refers to a solution of monellin in water.

The CD spectrum of monellin is dominated by the characteristic features of the antiparallel  $\beta$ -sheet. Due to the use of synchrotron radiation for the CD measurement<sup>204</sup>, the spectrum is accessible down to 175 nm. Both computed spectra show excellent agreement with the experiment. The GFN2-xTB based spectrum shows overall a better agreement, since also fine details like the shoulder at 205 nm are reproduced. The additional signals in the computed spectra above 240 nm can be attributed to the use of a single structure for the computation of the CD spectrum. The inset of Figure 5.5 shows the overlay of monellin structures from X-ray, GFN-FF, and GFN2-xTB. The main secondary structure motifs of the protein, i.e., the  $\alpha$ -helix and the  $\beta$ -sheet show only minor deviations. But even such small variations in the structure result in different CD spectra.

The next example is the  $\beta$ -barrel protein avidin, a tetrameric biotin-binding protein in egg whites.<sup>218,219</sup> One monomer unit serves as structure for the computation of the CD spectrum, which are presented in Figure 5.6 together with the experimental spectrum.



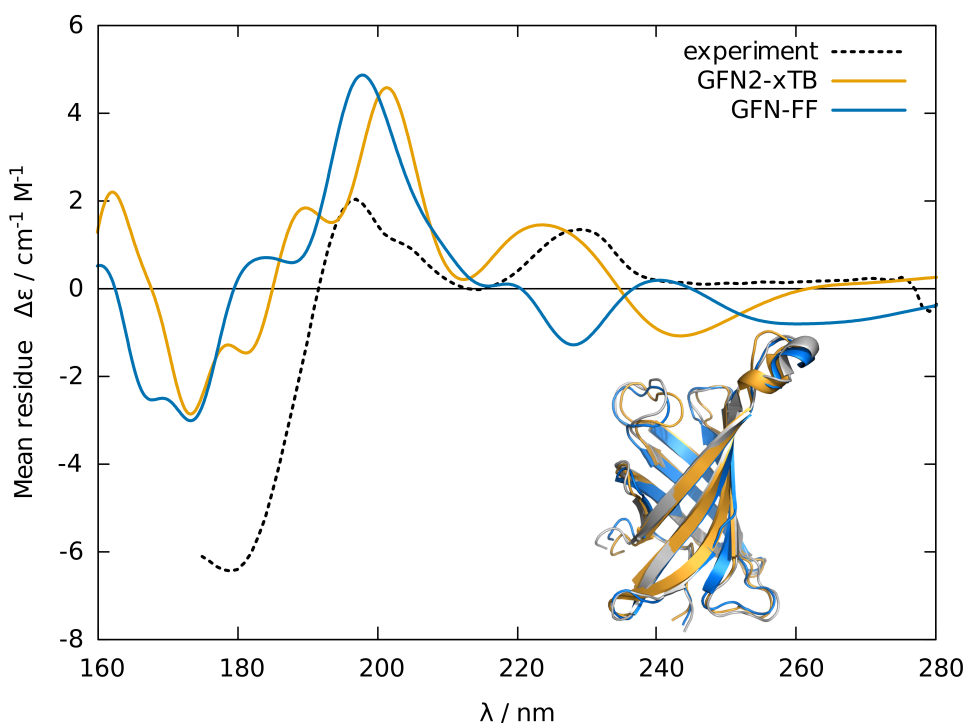


Figure 5.6.: ECD spectrum of avidin (PDB: 1RAV), computed with sTDA-xTB on optimized GFN-FF (blue) and GFN2-xTB (orange) structures. The experimental spectrum is taken from Ref. 204 and refers to a solution of avidin in water.

The experimental CD spectrum of avidin shows no characteristic bands of common secondary structure motifs. This can be attributed to the unique secondary structure of the protein. The  $\beta$ -barrel protein subunit also contains small  $\alpha$ -helical regions and a substantial amount of unstructured loops. This mixture of secondary structure motifs is responsible for the unique CD spectrum. Both theoretical spectra reproduce all features of the experimental spectrum very well. The spectrum on the GFN2-xTB optimized structure reproduces the intensity and the band position (energy) of the negative band at 180 nm and the two positive band at 195 nm and 230 nm, respectively, very well. The spectrum on the structure, optimized with GFN-FF, is in good agreement for the negative band and the first positive band (195 nm). The second positive band at 230 nm from the experiment seems to have the wrong sign. However, it is not possible to clearly assign this band from the computed spectrum to the one from the experiment. But the overall agreement for GFN-FF is good and due to the low computational cost, a reasonable alternative to GFN2-xTB in this case.

The next example is a member of the  $\beta$  scorpion toxin class. It is found in the

## II. Development and Application of the sTDA-xTB method

venom of scorpions and inactivates members of the voltage-gated ion channels.<sup>220</sup>. The protein comprises multiple types of secondary structure motifs, including  $\alpha$ -helices,  $\beta$ -sheets and random coil formations. The experimental and computed spectra are shown in Figure 5.7.

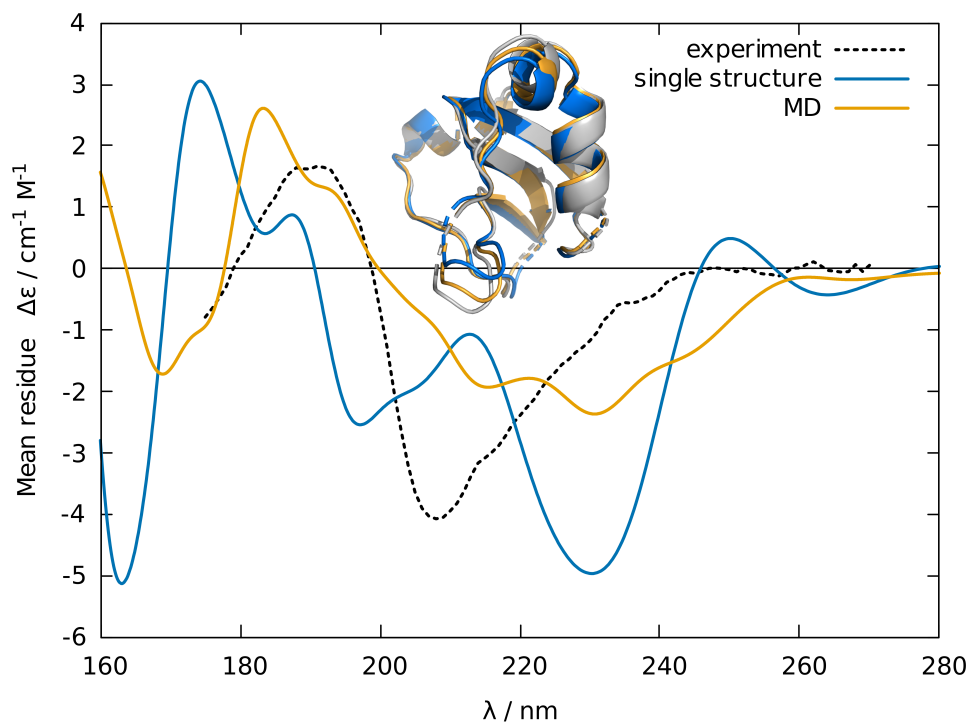


Figure 5.7.: ECD spectrum of a scorpion toxin (PDB: 4KYP), computed with sTDA-xTB on optimized (blue) and MD (orange) GFN-FF structures. The experimental spectrum is taken from Ref. 204 and refers to a solution of the toxin in water.

The mixture of secondary structure motifs is represented in the experimental CD spectrum, which is dominated by an intense negative feature at 209 nm and a positive band at 190 nm. Because of the flexible secondary structure – caused by the floppy random coil formations – the computed CD spectrum on the GFN-FF optimized structure is not in very good agreement. However, this deviations originates from some additional bands in the computed spectrum, the main features from the experiment are reproduced. To analyze the effect of MD sampling, the spectrum is averaged over 100 snapshots from a GFN-FF MD simulation. The comparison of the computed spectra shows that by sampling along an MD trajectory, the additional features from the spectrum on a single structure vanish. The computed CD spectrum on MD structures is in excellent agreement with the experiment.

## II. Development and Application of the sTDA-xTB method

The last example is the solution structure of a FATC domain, which is involved in redox-dependent structural and cellular stability.<sup>221</sup> The secondary structure consists of a single  $\alpha$ -helix. The experimental and computed spectra are depicted in Figure 5.8.

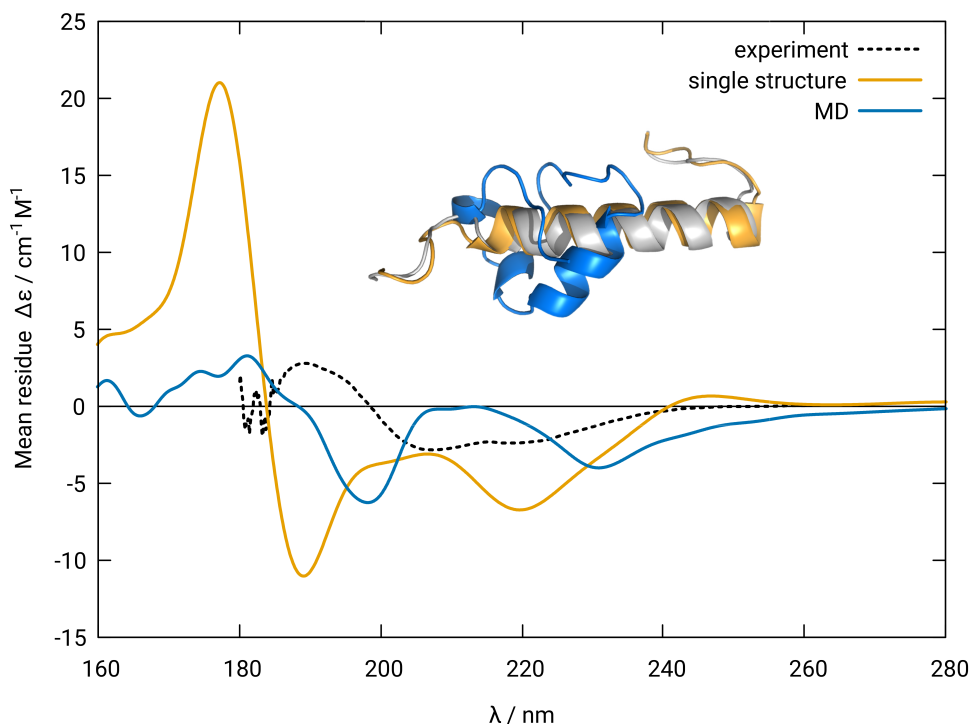


Figure 5.8.: ECD spectrum of a FATC domain (PDB: 4KIT), computed with sTDA-xTB on optimized (orange) and MD (blue) GFN2-xTB structures. The experimental spectrum is taken from Ref. 204 and refers to a solution of the FATC domain in water.

The experimental CD spectrum of the FATC-domain shows the characteristic features of the  $\alpha$ -helix. As demonstrated earlier, this type of secondary structure does not challenge sTDA-xTB or the different structure methods. Thus, the computed CD spectrum on a single optimized GFN2-xTB structure also shows the characteristic features. However, the intensity of the computed spectrum on the single structure is too large by a factor of five. The MD sampled CD spectrum agrees better in terms of intensity with the experimental spectrum. But the average structure from the MD simulations is more compressed than the reference structure from the NMR solution experiment. This agrees with the intensity of the experimental spectrum. The spectra in Figure 5.1 show that prolonged  $\alpha$ -helices yield large intensities. This is not the case for the spectrum of the FATC-domain, indicating short  $\alpha$ -helical structural parts. The MD simulation

## II. Development and Application of the sTDA-xTB method

identifies this folded version as more stable. This is an excellent example for the tandem application of efficient structural and excited states methods that are able to elucidate structural arrangements in solution.

### 5.4. Conclusion

In the present chapter, the structure methods GFN2-xTB and GFN-FF were tested in the context of providing optimized geometries or MD simulations for the computation of protein CD spectra with sTDA-xTB. Multiple model peptides with characteristic secondary structures and proteins were investigated for this purpose. This work extends the studies from chapter 3, which focused on the performance of sTDA-xTB for the computation of protein CD spectra. Consecutively, this chapter aimed at finding also a reliable structure method that can be applied without restrictions to provide an almost black-box workflow for the theoretical simulation of electronic excitation spectra for proteins and other large biomolecular aggregates.

CD spectra of model peptides for the  $\alpha$ -,  $\beta_{10}$ - and collagen P<sub>II</sub>-helix secondary structure motif were computed and the comparison with the experiment shows that structures from GFN-FF and GFN2-xTB are of equal quality and in very good agreement with the experiment. Note that this is especially remarkable for the universal FF GFN-FF: without any special parameterization for proteins, the results are equally reliable compared to the more expensive SQM GFN2-xTB. To test the two structure methods even further, the CD spectra of four proteins were studied. For all proteins, the agreement with the experimental spectra are remarkable. The sampling along an MD simulation, by taking snapshots equidistantly from a trajectory and averaging over the snapshot spectra, improves the results in some cases significantly. This is in line with prior studies in this context. In this regard it is important to have efficient structure methods. The SQM GFN2-xTB is computationally inapplicable for the structure optimization of very large proteins (> 10000 atoms) or long simulation times. GFN-FF, on the other hand, provides structures of remarkable quality at a fraction of the cost of GFN2-xTB calculations. Note that GFN-FF is conceptually unable to compute some reactive chemical processes due to its internal design with fixed bonding partners. GFN2-xTB is in this context more robust because of the QM foundation. A proper strategy for future studies on electronic excitation spectra is the combined use of GFN2-xTB and GFN-FF. Initial structures and long simulations are performed at the GFN-FF level. If the universal FF reaches its limits, GFN2-xTB can be used for further investigations.

In future studies, CD spectra of proteins with prosthetic groups – including metal centers – can be investigated. Because of the progresses from chapter 4 there

## II. Development and Application of the sTDA-xTB method

is no limitation to certain chemical elements. Furthermore, more rare secondary structure motifs (e.g.,  $\pi$ -,  $\beta$ -helices, or different strand types) can be studied with respect to their CD spectra.



**Part III.**

**Application of the  
sTD-DFT-xTB method**





### III. Application of the sTD-DFT-xTB method

In part II of this work, the sTDA-xTB method was comprehensively tested for the application on protein UV-Vis and CD spectra. As described in chapter 1, protein CD spectra can be utilized in medical research to characterize the protein secondary structure in solution.<sup>31,32</sup> SHIM is also a popular medical imaging technique, used to visualize non-centrosymmetric cell and tissue structures.<sup>222</sup> The underlying physical effect for this imaging technique is the SHG phenomenon, a scattering process in which the optical frequency of incident photons is doubled. Theoretical models can access this NLO phenomenon via the first hyperpolarizability ( $\beta$ ).

In part III, first hyperpolarizabilities are computed by the sTD-DFT-xTB<sup>223</sup> method, which was designed to enable the calculation of entire biomolecular structures that are measured in SHIM. Until the advent of sTD-DFT-xTB, systems with more than 1000 atoms were computationally unfeasible, and large systems must be treated in fragmentation approaches.<sup>60,61,224</sup> Chapter 6 deals with the computation of exemplary endogenous biotags for SHIM. For the first time, a systematic investigation of the structural ensemble of tryptophan-rich oligopeptides<sup>225</sup> and the computation of their spectroscopic properties is conducted. The very efficient potential energy surface (PES) sampling algorithm (CREST)<sup>48</sup> is applied for the generation of the conformational ensemble. Furthermore, the effect of MD sampling on the first hyperpolarizability is investigated. Additionally, the first hyperpolarizabilities are compared to extrapolated experimental and *ab-initio* TD-HF values.



## 6. Dynamic structural effects on the second-harmonic generation of short tryptophan-rich peptides and gramicidin A

Jakob Seibert\*, Benoît Champagne<sup>†</sup>, Stefan Grimme\* and Marc de Wergifosse\*

*Received 23th of January 2020, Published online 23th of March 2020*

Reprinted (adapted) with permission from<sup>‡</sup>

Seibert, J; Champagne, B.; Grimme, S.; de Wergifosse, M. *J. Phys. Chem. B* **2020**, *124*, 2568.

— Copyright © 2020, American Chemical Society.

DOI [10.1021/acs.jpcb.0c00643](https://doi.org/10.1021/acs.jpcb.0c00643)

### Own manuscript contribution

- Performing the calculations
- Interpretation of the results
- Writing the manuscript

---

\*Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

<sup>†</sup>Laboratoire de Chimie Théorique, Université de Namur, rue de Bruxelles, 61, B-5000 Namur, Belgium

<sup>‡</sup>Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

### III. Application of the sTD-DFT-xTB method

**Abstract** Peptide chains can model endogenous biotags for applications in second-harmonic imaging microscopy. Such structures are flexible which may strongly affect their structure–property relationship. Here, we explore quantum-mechanically the conformational space of a set of tryptophan-rich model peptides. This has become feasible because of the recently proposed meta-dynamics method based on efficient tight-binding (TB) calculations. The TB version of the simplified time-dependent density functional theory (sTD-DFT-xTB) method is used to evaluate the first hyperpolarizability ( $\beta$ ). These new tools enable us to calculate nonlinear optical properties for systems with several thousand atoms and/or to screen large structure ensembles. First, we show that the indole chromophore in tryptophan residues dominates the  $\beta$  response of these systems. Their relative orientation mostly determines the global  $\beta$  tensor and affects the static  $\beta$  response. The results underline the importance of finding low-energy conformers for modeling  $\beta$  of flexible molecules. Additionally, we compare calculated and extrapolated experimental static  $\beta$ . The sTD-DFT-xTB method is capable of providing reliable second-harmonic generation values for tryptophan-rich systems at a fraction of the computational cost of the commonly used TD-DFT/TD-HF levels of theory.

## 6.1. Introduction

Second harmonic imaging microscopy (SHIM)<sup>55–58</sup> has been developed for contrast enhancement of non-centrosymmetric molecular arrangements where the so-called second-harmonic generation (SHG) occurs. Information about the molecular organization of the chromophores can be extracted from SHG imaging data because the signal is polarization-dependent. SHIM is used as a high-resolution biological imaging technique where the SHG polarization anisotropy yields information about molecular orientation. Furthermore, it enables to analyze the degree of organization of proteins in tissues, related to their healthy or unhealthy state<sup>57</sup>. SHIM biotags could be endogenous like ordered structures of collagen<sup>59–61</sup>, microtubule, or myosin<sup>55,57</sup>. Exogenous biotags on the other hand should be carefully applied to avoid phototoxicity problems.<sup>55</sup>

Generally, the SHG process is not directly photodamaging the living environment because it is a scattering effect. However, at the laser wavelength, two-photon absorption and subsequent emission may occur so that excited state photochemistry could damage the sample. Thus, to avoid this, one must tune the molecular properties of such biotags.<sup>56</sup> A large first hyperpolarizability ( $\beta$ ) is required for efficient SHG. Then, one can record the SHG signal of bright dyes with a low laser power, limiting photodamage. The exogenous biotags should also have a large  $\beta$  within the tissue transparency window (700–900 nm) and should involve minor one- and two-photon activities. Biotags such as fluorescent proteins can also be

introduced by genetic engineering.

Some of us theoretically characterized the nonlinear optical (NLO) properties of fluorescent proteins at different theoretical levels of theory, including an ONIOM MP2:HF scheme where the first shell of residues around the chromophore was included in.<sup>224,226–229</sup> These investigations showed the importance to account for the H-bond network close to the chromophore and how the  $\beta$  of FPs strongly depends on the  $\pi$ -conjugation pathway, the degree of bond length alternation, and the presence of  $\pi$ -stacking interactions. Theoretical studies on compounds of interest for SHIM are still scarce.

In this contribution, we calculate and analyze the SHG response of small peptide chains and the peptide gramicidin A as models to understand the NLO properties of tryptophan-rich endogenous dyes. This is a follow-up work of the study of Duboisset *et al.*<sup>225</sup> who investigated experimentally the SHG response of an ensemble of tryptophan-rich peptides and gramicidin A. This set of systems is composed of KWK, KWWK, KWWWK, and KWWKWWK compounds where W and K denote tryptophan and lysine units, respectively. Gramicidin A is a natural peptide with the sequence VGALAVVWLWLWLW, comprising four W units. Here, V, G, A, and L are the one-letter codes for the amino acids valine, glycine, alanine, and leucine, respectively. SHG signals were measured by hyper-Rayleigh scattering experiments at a wavelength of 784 nm, in Tris buffered aqueous solutions. While providing experimental reference first hyperpolarizabilities for this set of compounds, Duboisset *et al.*<sup>225</sup> showed that the  $\beta$  response follows an additive scheme with increasing number of W units. It also appears that the first hyperpolarizability of a sole tryptophan strongly differs from a KWK one. They argued that the local environment created by both lysine units decreases its SHG response.

From a theoretical point of view, the first hyperpolarizability of these systems is difficult to evaluate not only because of their size but also because of conformational flexibility. This implies a large number of relevant structures at room temperature. Recently, two of us proposed the simplified time-dependent density functional theory (sTD-DFT) method<sup>223</sup> in its tight-binding version to evaluate the frequency-dependent first hyperpolarizability of large compounds with up to about 3000 atoms. With respect to a full TD-DFT treatment, the simplified method applies three approximations: i. the exact Coulomb and exchange integrals are approximated by short-range damped Coulomb interactions of transition density monopoles, ii. the CI excitation space is truncated as controlled by a single energy selection threshold parameter, and iii. the response of the exchange-correlation functional is neglected.<sup>72</sup>

For the evaluation of the first hyperpolarizability, two minor additional approximations are introduced, i.e., both the Hartree exchange-correlation kernel and the

### III. Application of the sTD-DFT-xTB method

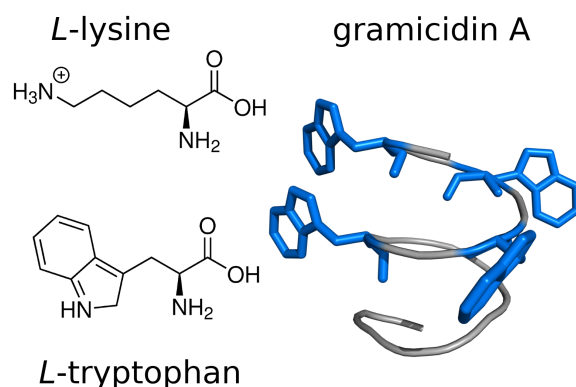


Figure 6.1.: Lewis structures of amino acids lysine and tryptophan, building residues for the model peptides. Secondary structure of gramicidin A with blue highlighted tryptophan units.

third functional derivative of the exchange-correlation functional are neglected.<sup>223</sup> A semi-empirical tight-binding version called sTD-DFT-xTB also has been developed where instead of using KS-DFT input data, orbitals and eigenvalues from an extended basis set tight-binding calculation are employed.<sup>76</sup> It was shown that the sTD-DFT-xTB method can reproduce reasonably well the  $\beta$  frequency dispersion of a collagen triple helix  $[(\text{Pro-Pro-Gly})_{10}]_3$  and of fluorescent proteins with respect to ONIOM reference calculations.<sup>223</sup>

Here, the sTD-DFT-xTB method is used to evaluate and understand dynamical structural effects on the SHG response for the set of tryptophan-rich short peptides and gramicidin A characterized experimentally by Duboisset *et al.*<sup>225</sup>.

This article is organized as followed. First, we detail the calculations for this study and then present in the first results part, the analysis of the conformer ensembles for all systems with respect to their first hyperpolarizability properties. Structure-property relationships and the effect of sampling structures along a molecular dynamic (MD) trajectory is discussed. The last part of the results section compares experimental first hyperpolarizabilities to computed values at different levels of theory. The summary section concludes the main findings and possible implications of this work for future applications.

## 6.2. Computational Details

Figure 6.1 presents the structures of L-lysine (K) and L-tryptophan (W), which are the building blocks of all tryptophan-rich peptides compounds considered in this study. The secondary (rigid helical) structure of gramicidin A with highlighted tryptophan units is also shown. In order to evaluate the impact of conformational

### III. Application of the sTD-DFT-xTB method

flexibility on the first hyperpolarizability as well as for determining lowest energy conformers, we used the RMSD-based meta-dynamics approach recently proposed by Grimme<sup>230,231</sup>, except for gramicidin A. First, the standard, ground-state tight-binding GFN2-xTB<sup>75</sup> method is used for the generation of the conformer ensemble. Solvation effects for water are implicitly accounted for using the GBSA<sup>74,129</sup> continuum model. Second, all conformers within a 6 kcal/mol GFN2-xTB energy window were optimized at the PBEh-3c(COSMO)<sup>127,161</sup> level of theory. Third, within a 4 kcal/mol PBEh-3c energy window all remaining structures were used as input for PW6B95/def2-QZVP<sup>167,232</sup> single point energy calculations. Free energies were computed for the lowest energy conformers by adding solvation free energies with COSMO-RS<sup>233,234</sup> and thermostatical contributions within the modified<sup>235</sup> rigid-rotor harmonic oscillator approximation based on the GFN2-xTB computed Hessian ( $\Delta G = \Delta E_{\text{PW6B95}} + \Delta G_{\text{solv}}^{\text{COSMO-RS}} + \Delta G_{\text{RRHO}}^{\text{GFN2-xTB}}$ ). The Boltzmann weights used correspond to a temperature of 298.15 K. The first hyperpolarizability of relevant conformers (population larger than 1.5%) were determined at the sTD-DFT-xTB/GBSA level of theory with modified Coulomb  $y_J$  and exchange  $y_K$  parameters set to 0.55 and 1.0, respectively. To further explore dynamic structural effects, molecular dynamic (MD) simulations were carried out for 1 ns at the GFN2-xTB level of theory with preceding equilibration. In the simulations, a time step of 4 fs was used and the SHAKE<sup>202,203</sup> algorithm was applied, constraining all covalent bonds. The lowest conformer was used as starting point for the MD simulations. From the resulting trajectory 200 snapshots were taken equidistantly and used as structural input for the sTD-DFT-xTB/GBSA<sup>223</sup> calculations.

All sTD-DFT-xTB/GBSA calculations were performed with a configuration selection threshold of 10 eV. Note that the GBSA solvation model is only applied in the (ground state) orbitals generation step, meaning that non-equilibrium solvent effects on the hyperpolarizability are not accounted for. These effects usually enhance the SHG response<sup>236–239</sup>. As already mentioned in the original publication of the method<sup>223</sup>, the underlying sTD-DFT-xTB parameterization – originally developed for excitation energies and absorption spectra – is not perfectly suited for nonlinear optical properties which more strongly depend on the high-energy part of the excitation manifold. We employ a small model system for benchmarking and adjustment purposes, i.e., performing high-level calculations for reference  $\beta$  values and then to tune sTD-DFT-xTB Coulomb  $y_J$  and exchange  $y_K$  parameters accordingly. We computed the static  $\beta_{\text{HRS}}$  of tryptophan at the CCSD(T)/aug-cc-pVDZ<sup>180</sup> level of theory using the Romberg's automatic finite-field (FF) differentiation procedure.<sup>240</sup> Its missing frequency dispersion was accounted for using the multiplicative approximation ( $\beta_{\text{CC}}^{\omega} = \beta_{\text{CC}}^0 \frac{\beta_{\text{TDForM06-2X}}^{\omega}}{\beta_{\text{TDForM06-2X}}^0}$ ). Figure 6.2 shows the adjusted sTD-DFT-xTB  $\beta_{\text{HRS}}$  curve matching almost perfectly the CCSD(T)

### III. Application of the sTD-DFT-xTB method

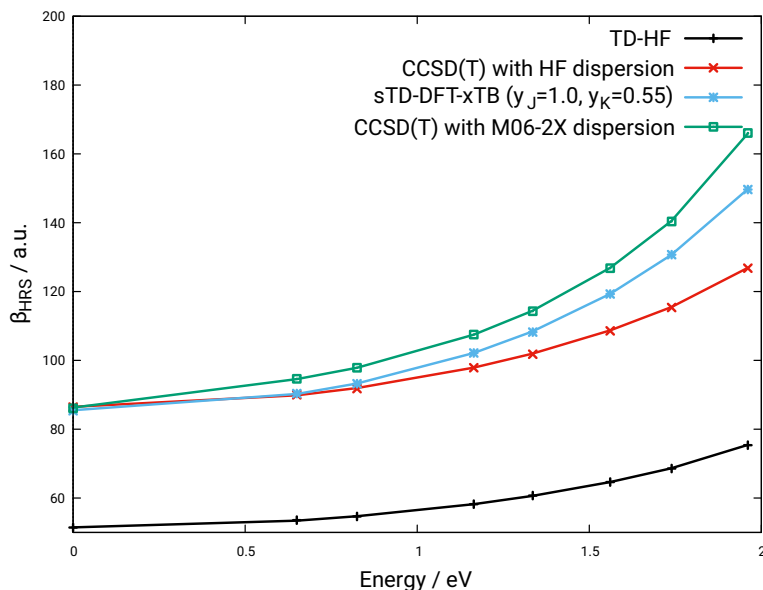


Figure 6.2.: Frequency dispersion for tryptophan computed with CCSD(T)/aug-cc-pVDZ (red), TDHF/aug-cc-pVDZ (black) and sTD-DFT-xTB (blue) methods.

one with the M06-2X frequency dispersion when using a  $y_J$  parameter of 0.55 instead of the original value of 4.0 and a  $y_K$  of 1.00 instead of 2.0. For all following sTD-DFT-xTB calculations, the  $y_J$  and  $y_K$  parameters are set to 0.55 and 1.0, respectively.

Time-dependent Hartree-Fock (TDHF) calculations were also conducted for lowest energy conformers with the 6-31+G(d) basis set with and without solvent effects accounted for using the IEF-PCM scheme.<sup>241,242</sup> Note that when comparing TDHF results to experiment, the static and dynamic dielectric constants of water differ largely. Thus, we select a value at a large wavelength of 1900 nm instead of the static one. This wavelength is chosen to enable non-biased comparisons and should be large enough to prevent any (near)resonance effects. For the smaller systems W and KWK we also computed the response with the slightly larger aug-cc-pVDZ basis set showing very small differences with respect to 6-31+G(d) (see table A3.1 and A3.2). In the experimental work of Duboisset *et al.*<sup>225</sup>, the hyper-Rayleigh scattering value  $\beta_{\text{HRS}}$ <sup>243</sup> was determined. Theoretically, the following definition of  $\beta_{\text{HRS}}$  is used as the mean of  $\beta$ -tensor orientations

$$\beta_{\text{HRS}}(-2\omega; \omega, \omega) = \sqrt{\{\langle \beta_{ZZZ}^2 \rangle + \langle \beta_{ZXX}^2 \rangle\}}, \quad (6.1)$$

where molecular averages without assuming Kleinman's conditions<sup>244</sup> are defined



in the laboratory frame as

$$\begin{aligned}
 \langle \beta_{ZZZ}^2 \rangle &= \frac{1}{7} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta}^2 + \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} \\
 &+ \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\zeta\zeta\eta} + \frac{4}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\eta\eta\zeta} + \frac{1}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 \\
 &+ \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\xi\xi} + \frac{1}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\xi\xi} + \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\xi\xi\eta} \\
 &+ \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2 + \frac{4}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi} \beta_{\eta\zeta\xi},
 \end{aligned} \tag{6.2}$$

and

$$\begin{aligned}
 \langle \beta_{ZXX}^2 \rangle &= \frac{1}{35} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{4}{105} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} - \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\eta\eta\zeta} \\
 &+ \frac{8}{105} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta}^2 + \frac{3}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\eta\eta}^2 - \frac{2}{35} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\zeta\zeta} \\
 &+ \frac{1}{35} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\eta} \beta_{\zeta\xi\xi} - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\xi} \beta_{\eta\eta\xi} - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\zeta\eta} \beta_{\eta\xi\xi} \\
 &+ \frac{2}{35} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2 - \frac{2}{105} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi} \beta_{\eta\zeta\xi}.
 \end{aligned} \tag{6.3}$$

In addition from these quantities, the depolarization ratio can be obtained

$$\text{DR} = \frac{I_{VV}^2 \omega}{I_{HV}^2 \omega} = \frac{\langle \beta_{ZZZ}^2 \rangle}{\langle \beta_{ZXX}^2 \rangle}. \tag{6.4}$$

where a value of 1.5 corresponds to a fully octupolar response, of 5 to a one-dimensional push-pull  $\pi$ -conjugated system, and of 9 to a fully dipolar system. Furthermore, it is useful to analyze the  $\beta$  tensor in terms of its dipolar ( $J=1$ ) and octupolar ( $J=3$ ) tensorial  $\beta_J$ -components<sup>245</sup>:

$$\beta_{\text{HRS}} = \sqrt{\langle \beta_{\text{HRS}}^2 \rangle} = \sqrt{\frac{10}{45} |\beta_{J=1}|^2 + \frac{10}{105} |\beta_{J=3}|^2}, \tag{6.5}$$

$$\begin{aligned}
 \text{where } |\beta_{J=1}|^2 &= \frac{3}{5} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 + \frac{6}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} \\
 &+ \frac{3}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 + \frac{3}{5} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\xi\xi},
 \end{aligned} \tag{6.6}$$

### III. Application of the sTD-DFT-xTB method

$$\begin{aligned}
 |\beta_{J=3}|^2 = & \frac{2}{5} \sum_{\zeta}^{x,y,z} \beta_{\zeta\zeta\zeta}^2 - \frac{6}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\zeta\zeta\zeta} \beta_{\zeta\eta\eta} + \frac{12}{5} \sum_{\zeta \neq \eta}^{x,y,z} \beta_{\eta\zeta\zeta}^2 \\
 & - \frac{3}{5} \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\eta\zeta\zeta} \beta_{\eta\xi\xi} + \sum_{\zeta \neq \eta \neq \xi}^{x,y,z} \beta_{\zeta\eta\xi}^2.
 \end{aligned} \tag{6.7}$$

Using this decomposition, the electronic character of a NLO chromophore can be analyzed. We used a loss-less 3D visualization of the first hyperpolarizability tensor, to have a more intuitive analysis tool for the  $\beta$  tensor.<sup>246</sup> This so-called unit-sphere representation (USR) uses effective SHG dipoles, which are defined as:

$$\vec{\beta}_{\text{eff}} = \vec{\beta} : \hat{E}(\theta, \phi) \hat{E}(\theta, \phi) \tag{6.8}$$

Taking all possible incident polarization directions defined by  $(\theta, \phi)$ , a unit-sphere is mapped out (at a field value of  $\hat{E}(\theta, \phi)$  1 a.u.). At these sampled points on the unit-sphere surface, the corresponding  $\vec{\beta}_{\text{eff}}$  is plotted. This scheme allows a three-dimensional visualization of the  $\beta$  tensor without losing any information. Another approach for visualizing the  $\beta$  tensor, but at cost of losing anisotropic information, consists in defining a vector, having the following components:<sup>247</sup>

$$\beta_x = \beta_{xxx} + \beta_{xyy} + \beta_{xzz}; \beta_y = \beta_{yxx} + \beta_{yyy} + \beta_{yzz}; \beta_z = \beta_{zxx} + \beta_{zyy} + \beta_{zzz} \tag{6.9}$$

All reported  $\beta$  values are given in atomic units [1 a.u. of  $\beta=3.6213 \cdot 10^{-42} \text{ m}^4 \text{ V}^{-1}=3.2064 \cdot 10^{-53} \text{ C}^3 \text{ m}^3 \text{ J}^{-2} = 8.639 \cdot 10^{-33} \text{ esu}$ ] within the Taylor series convention<sup>248</sup>. The CCSD(T) finite-field first hyperpolarizabilities were computed by the T-REX program natively interfaced with QChem Version 5.1<sup>249</sup>. All TDHF calculations were carried out using the Gaussian 09 package.<sup>250</sup> For all DFT calculations TURBOMOLE Version 7.2<sup>138,139</sup> was used with COSMOtherm Version C3.0 release 1601<sup>233,234</sup> for COSMO-RS. The remaining calculations were conducted with the xtb<sup>140</sup> and stda<sup>205</sup> codes.

The first hyperpolarizability values of the tryptophan-rich peptides ( $\beta_{\text{HRS}}^{800}$ ) were experimentally determined by Duboisset *et al.*<sup>225</sup>. To eliminate the resonance effects and extrapolate to the static value<sup>251,252</sup>, we process the experimental data with the two-state approximation (TSA) proposed by Oudar and Chemla<sup>253</sup>.

$$F(\omega, \omega_{ge}, \gamma) = \frac{\beta_{zzz}(-2\omega, \omega, \omega)}{\beta_{zzz}(0, 0, 0)} = \frac{\omega_{ge}^2 (\omega_{ge} - i\gamma)^2}{([\omega_{ge} - i\gamma]^2 - 4\omega^2) ([\omega_{ge} - i\gamma]^2 - \omega^2)} \tag{6.10}$$

A homogeneous broadening of  $\gamma = 0.35 \text{ eV}$  is applied while the experimental excitation energy of tryptophan  $\omega_{ge} = 4.44 \text{ eV}$  is used. Aside from the conventional TSA, inhomogeneous broadening or even the vibronic structure of the excited states could be accounted for<sup>60,251,252</sup>. However, using and comparing these refined extrapolation schemes for the studied systems goes beyond the scope of

Table 6.1.: Measured and extrapolated (TSA) first hyperpolarizabilities in atomic units.

<b>system</b>	$\beta_{\text{HRS}}^{800}$	$\beta_{\text{HRS}}^{\infty}$
<b>W</b>	544	240
<b>KWK</b>	100	44
<b>KWWK</b>	396	175
<b>KWWWK</b>	828	365
<b>KWWKWWK</b>	1215	536
<b>gramicidin A</b>	872	384

this study and would be inaccessible since some required experimental data are missing. Table 6.1 shows the dynamic and extrapolated experimental values.

## 6.3. Results and Discussion

### 6.3.1. Conformers

Peptides in solution at ambient temperature are represented as a set of various conformers. Depending on the temperature, these are populated differently and the accessed conformational space can become quite large. The objective of this section is to analyze the conformational dependence of the SHG response, shedding light on this structure-property relationship. The above described theoretical multi-level approach provides a very reasonable conformational ensemble including solvation effects. For these sets of conformers, molecular first hyperpolarizabilities (static  $\beta_{\text{HRS}}$ ) were computed. With the help of the described visualization techniques for the  $\beta$  tensor, the conformers of each system are compared in terms of their electronic and geometrical structures to show how they impact the NLO properties.

First, the tryptophan molecule is analyzed. For this system, the indole unit is essentially responsible for the first hyperpolarizability. Therefore, only small variations in  $\beta_{\text{HRS}}$  values are expected due to its rigid  $\pi$ -conjugated system. The amine and carboxylic acid moieties are flexible and the obtained conformer ensemble is mainly determined by their different orientations. Figure 6.3 shows the most contributing conformers at room temperature for tryptophan and their respective first hyperpolarizability values which do not differ much. The largest differences are caused by the re-orientation of the amine group. A small SHG enhancement is observed when local dipoles of the indole and amine units are perpendicular (con-

### III. Application of the sTD-DFT-xTB method

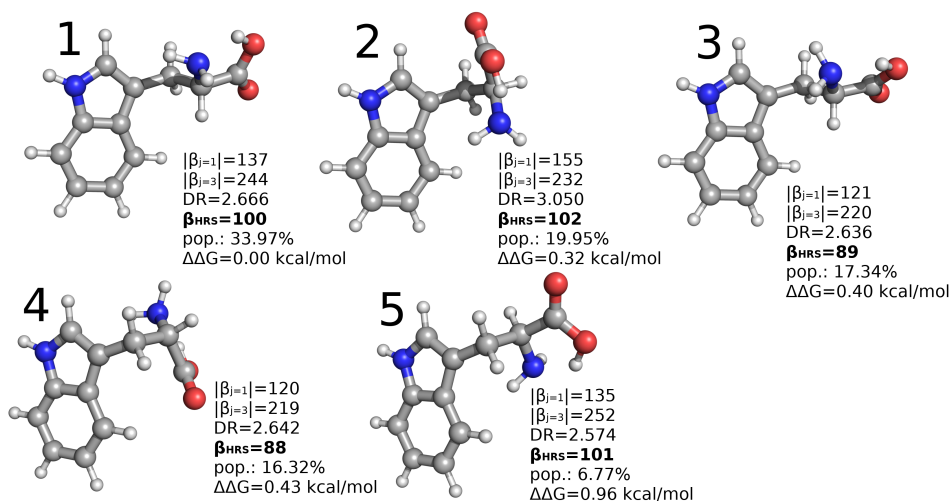


Figure 6.3.: Conformer ensemble for tryptophan. First hyperpolarizability data ( $|\beta_{J=1}|$ ,  $|\beta_{J=3}|$ , DR,  $\beta_{\text{HRS}}$ ), population and relative free energies are depicted with the plotted structures.

former 2). This can be further visualized by the USR (see Figure A3.1). There, two conformers with the largest and smallest  $\beta_{\text{HRS}}$  values are shown (conformers 2 and 4). The pattern of effective SHG dipoles indicates a mix of dipolar and octupolar contributions. According to Equation 5, e.g., the  $\beta_{\text{HRS}}^2$  value of conformer 2 contains 51% of dipolar and 49% of octupolar contribution.

The second system is the model peptide KWK, a lysine capped tryptophan. The lysine groups were introduced for solubility reasons.<sup>225</sup> Experimentally, the observed difference in responses between tryptophan and KWK was attributed to the lysine side chains that shield the indole unit from the solvent. A large variety of conformers is found due to the high flexibility of these side chains. However, the rigid indole chromophore is unchanged among these conformers. Figure 6.4 presents the conformers and their first hyperpolarizability properties. Note that sTD-DFT-xTB/GBSA calculations cannot account for non-equilibrium solvent effects on the response and hence part of the solvent-induced difference between tryptophan and KWK is missing. Thus, we conducted TDHF/IEF-PCM calculations for comparison that are discussed in the SI. Mostly, the picture emerging from the calculations is similar for both systems: several conformers of KWK are significantly populated, as for W. The USR for KWK (Figure A3.2) is more complex to analyze because of small dipolar contributions from both peptide bonds, but overall the response is similar to the one for W (i.e., a mix of dipolar and octupolar contributions). Note, however, that the relative  $|\beta_{J=1}|$  and  $|\beta_{J=3}|$  contributions to  $\beta_{\text{HRS}}^2$  are quite different for, e.g., conformers 2 and 4 (with factors of 2.4 and 0.6 between the two contributions).

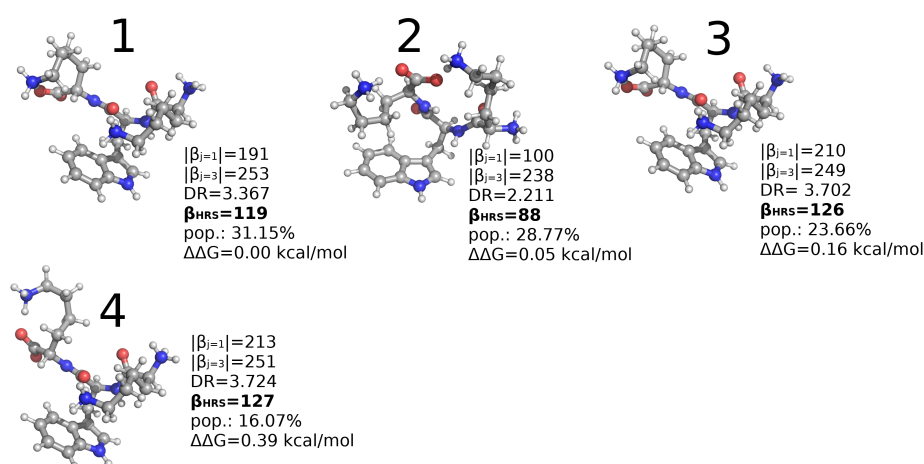


Figure 6.4.: Conformer ensemble for model peptide KWK. First hyperpolarizability data ( $|\beta_{j=1}|$ ,  $|\beta_{j=3}|$ , DR,  $\beta_{\text{HRS}}$ ), population and relative free energies are depicted with the plotted structures.

The third model peptide KWWK includes two chromophore units. The respective orientation of the two indole groups is mainly responsible for the change of the first hyperpolarizabilities among the ensemble. Figure 6.5 displays conformers of the KWWK model peptide, where the lysine groups and the peptide backbone are hidden to improve the visibility of chromophores. Based on the results of tryptophan and KWK systems, we conclude that the effects of the saturated side chains are negligible for the SHG response. Therefore, the structure-property analysis is focused on the orientation of the indole moieties. For KWWK, the extreme  $\beta_{\text{HRS}}$  values differ by more than a factor of two. The values among the ensemble show a larger spread than for the mono-chromophoric examples. The USR of both conformers – with the lowest (5) and highest (2)  $\beta_{\text{HRS}}$  values – shows that the orientation of the indole units plays an important role on the SHG response (see Figure 6.6). The indole units are aligned parallel in the second conformer with the largest first hyperpolarizability (232 a.u.). On the other hand, one can observe a drastic decrease of the  $\beta_{\text{HRS}}$  value when the indole units are antiparallel (71 a.u. for conformer 5). The USR shows a very strong dipolar character of the  $\beta$  tensor, when the chromophore units are aligned parallel. In that case, the DR value of 4.95 confirms this. For opposite orientations, the octupolar component dominates the  $\beta$  tensor, as indicated by a DR value of 1.99. For the larger KWWK and KWWKWWK systems, similar observations as for KWWK are made. The orientation of the indole groups directly correlates with the magnitude of the first hyperpolarizability. The conformers of KWWK and KWWKWWK are depicted in Figure 6.7 and 6.8, respectively. The two highest

### III. Application of the sTD-DFT-xTB method

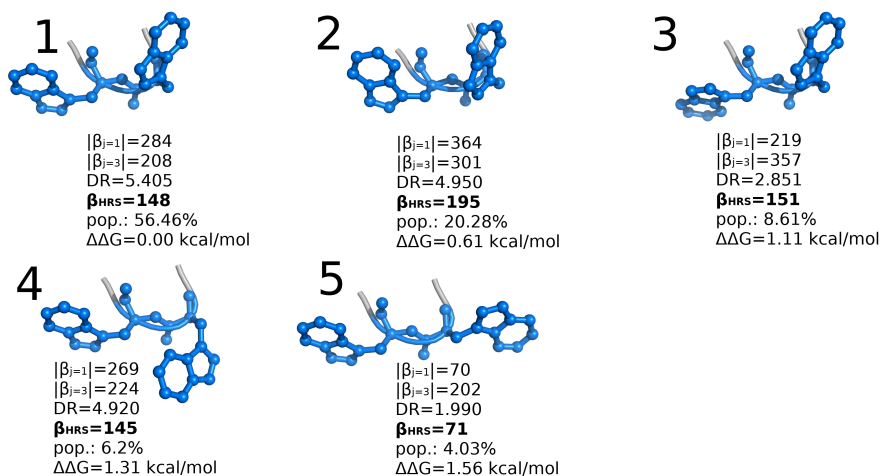


Figure 6.5.: Conformer ensemble for model peptide KWWK. First hyperpolarizability data ( $|\beta_{J=1}|$ ,  $|\beta_{J=3}|$ , DR,  $\beta_{\text{HRS}}$ ), population and relative free energies are depicted with the plotted structures.

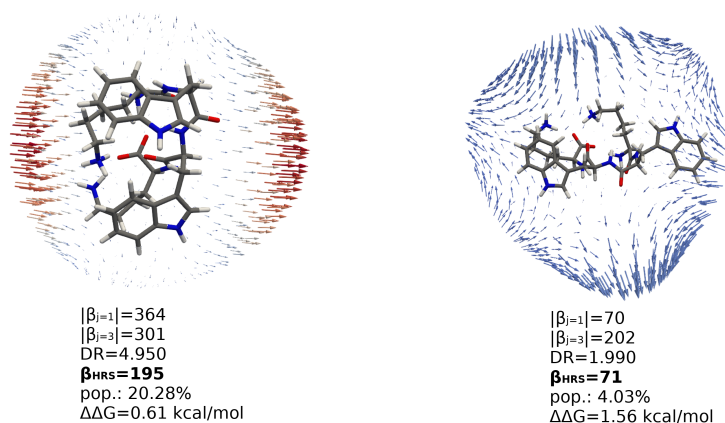


Figure 6.6.: Unit-sphere representation for KWWK conformers 2 (left) and 5 (right). To increase visibility, the vector fields are differently scaled but arrow colors are consistent.

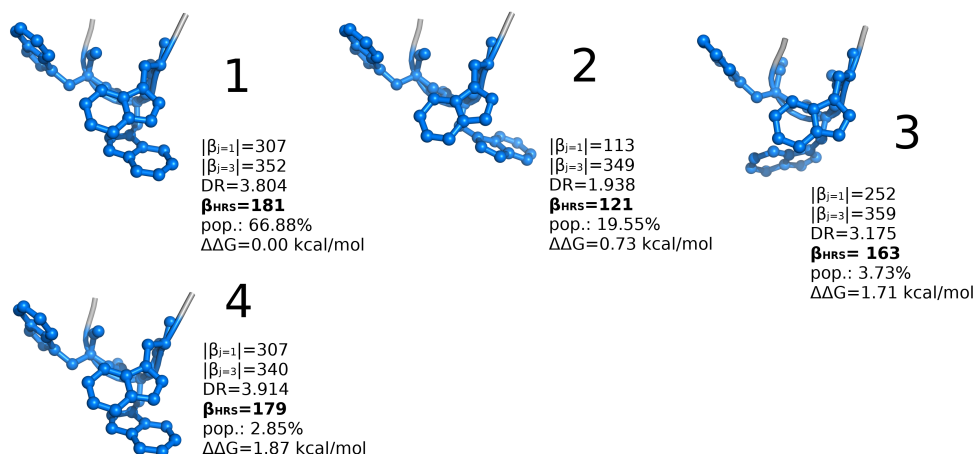


Figure 6.7.: Conformer ensemble for model peptide KWWWK. First hyperpolarizability data ( $|\beta_{J=1}|$ ,  $|\beta_{J=3}|$ , DR,  $\beta_{\text{HRS}}$ ), population and relative free energies are depicted with the plotted structures.

populated conformers of KWWWK differ by almost a factor of 1.5 in  $\beta_{\text{HRS}}$  values (1 and 2 with  $\beta_{\text{HRS}}=181$  a.u. and 121 a.u., respectively). Conformer 1 has almost equal dipolar  $|\beta_{J=1}|$  and octupolar  $|\beta_{J=3}|$  values. Considering their weighting factors –  $\frac{10}{45}$  and  $\frac{10}{105}$ , respectively – a strong dipolar contribution to the  $\beta_{\text{HRS}}$  is observed, confirmed by a DR of 3.80. For this conformer, two indole units are more or less pointing in the same direction while one points in a different one (but not opposite). For conformer 2, this third indole unit changes its orientation to point in an opposite direction with respect to the first one, canceling each others. The dipolar contribution  $|\beta_{J=1}|$  then equals only to 113 a.u. where the octupolar component is more than three times larger ( $|\beta_{J=3}|=349$  a.u.). The DR value of 1.94 indicates an octupolar case for this conformer, which is corroborated by the USR depicted in Figure A3.3. This analysis shows that the inclusion of several conformers changes the  $\beta_{\text{HRS}}$  value compared to the one of the lowest energy conformer. In a later section, we will discuss whether this approach improves the first hyperpolarizability values in comparison to experiment.

With four chromophore units, KWWKWWK is the largest example of all model peptides studied here. Figure 6.8 presents the six significantly populated conformer structures. Among this set, the conformers differ mostly by the orientation of their indole units. The first hyperpolarizability values spread by a factor of two between the conformers 3 and 4 ( $\beta_{\text{HRS}}=234$  and 112 a.u.). For conformers 1, 3, and 6, the indole units are oriented roughly in the same manner which is mainly due to the shared secondary structure. In the second group of trans-like conformers (2, 4 and 5), the indole units are partially oriented anti-parallel. This results in a decrease

### III. Application of the sTD-DFT-xTB method

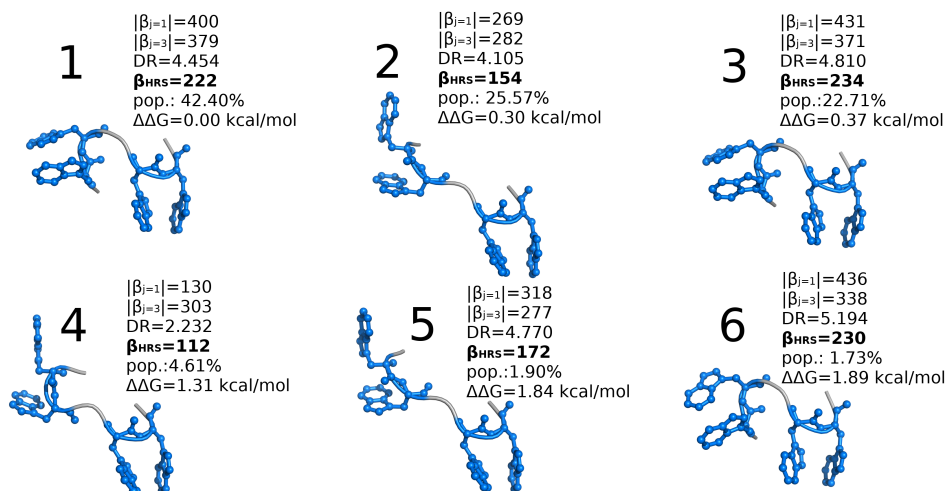


Figure 6.8.: Conformer ensemble for model peptide KWWKWWK. First hyperpolarizability data ( $|\beta_{j=1}|$ ,  $|\beta_{j=3}|$ , DR,  $\beta_{\text{HRS}}$ ), population and relative free energies are depicted with the plotted structures.

of the dipolar contribution to the  $\beta_{\text{HRS}}$ . The values for  $|\beta_{j=1}|$  and  $|\beta_{j=3}|$  indicate that for this cis-like group the dipolar contribution is dominant. The secondary structure can bend only because of the flexibility introduced by the third lysine unit. While the KWWK ensemble is only dominated by a few conformers due to a rigid peptide backbone, the KWWKWWK ensemble is clearly enlarged by this added flexibility. For the interested reader we provide in section A3.3 an analysis in terms of indole unit  $\beta$  vectors. They explain further the significant differences observed in calculated SHG response among conformers.

Table 6.2 shows the first hyperpolarizability values for the lowest energy conformer and the Boltzmann-weighted ensemble. For tryptophan and KWK, the two values are very close to each other. As already described above, this is due to the rigidity of the chromophore. For the larger systems, we observed slightly larger differences between a single structure approach and the ensemble average. We explain this with the orientation of multiple chromophore units in ways that enhance or cancel the SHG response. The difference in  $\beta_{\text{HRS}}$  values (in a.u.) between KWK and KWWK amounts to 30, to 32 for KWWK and KWWWK, and to 41 a.u. for KWWWK and KWWKWWK. Since not all lowest energy conformers share the same indole orientation, these enhancements are not equal. When considering the Boltzmann weighted  $\beta_{\text{HRS}}$ , the differences can be quantified to 41, 13, and 35 a.u. Also here, the enhancement is not perfectly linear. However, such a perfect linear enhancement is not realistic since the different conformers for the individual model peptides have a strong influence on the SHG response. These results clearly indicate that the addition of a tryptophan unit to these model peptides does not



Table 6.2.: Static first hyperpolarizability values (in a.u.) for the lowest free energy conformer and the Boltzmann weighted ensemble computed with sTD-DFT-xTB/GBSA.

system	minimum	ensemble
W	100	97
KWK	119	113
KWWK	149	154
KWWWK	181	167
KWWKWWK	222	202

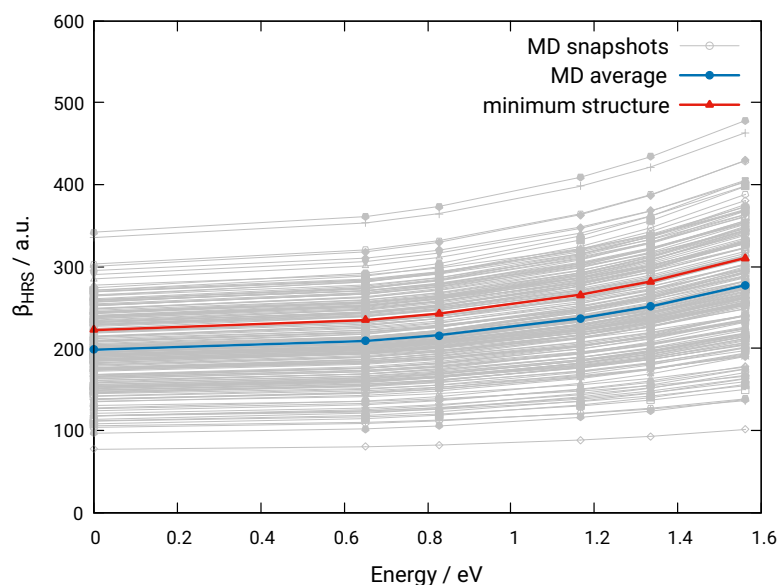


Figure 6.9.: Frequency dispersion computed with sTD-DFT-xTB for 200 snapshots (gray), the MD average (blue) and for the optimized minimum structure (red) for the KWWKWWK peptide.

equally enhance the first hyperpolarizability.

### 6.3.2. Molecular dynamics sampling

In this section, we analyze the effect of using several structures from a MD simulation as input for the evaluation of  $\beta$ . For the simulation of electronic circular dichroism spectra, some of us have already applied this approach successfully<sup>197,254,255</sup>. From a MD trajectory, snapshots are taken equidistantly and serve as input structures for the property calculations which are simply averaged over all included snapshots. In this process, no structural relaxations are included and the snapshots are equally weighted. By using a reasonably long simulation time, the considered structures should represent a Boltzmann ensemble.

Figure 6.9 shows the frequency dispersion of  $\beta$  for each snapshot as well as their

### III. Application of the sTD-DFT-xTB method

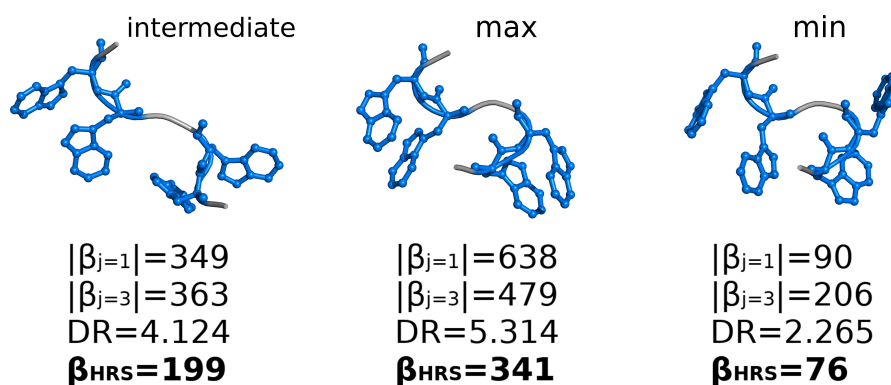


Figure 6.10.: Selected MD structures for model peptide KWWKWWK. First hyperpolarizability data ( $|\beta_{j=1}|$ ,  $|\beta_{j=3}|$ , DR and  $\beta_{\text{HRS}}$ ) are depicted with the plotted structures.

average for the largest peptide KWWKWWK. The value of  $\beta_{\text{HRS}}$  changes drastically within a factor of six between the most extreme structures. This illustrates the sensitivity of this NLO property with respect to structural subtleties of a flexible system. However, when all snapshots are averaged the resulting frequency dispersion of  $\beta_{\text{HRS}}$  does not differ much from the one that considers only the lowest energy conformer with a reduction of around 2-22%. The frequency dispersion of the Boltzmann weighted conformer ensemble does not differ much from the MD averaged ensemble (cf. Figure A3.8). Figure 6.10 shows selected snapshots from the MD trajectory giving the lowest, the highest, and intermediate values for the first hyperpolarizability. Analyzing the structures in terms of indole orientations corroborates the findings from the previous section. When all indole units are oriented along the same direction, the largest  $\beta_{\text{HRS}}$  values are obtained. Considering the structure with the minimum response, we observe that two indole units are pointing in opposite directions, canceling out their dipolar contributions to  $\beta$ , leaving a dominant octupolar character.

Table 6.3 shows MD averaged static first hyperpolarizabilities for all model peptides. The difference in  $\beta_{\text{HRS}}$  values between the minimum structure (minimum) and MD averaged structures (MD average) increases with system size. This observation holds also for the difference between the minimum energy conformer and the conformer weighted ensemble (ensemble). The first hyperpolarizability values of the ensemble differ from the ones of the MD average, except for the biggest system. Here, emphasize in the discussion on two systems. First, for tryptophan we obtain similar values for all three approaches. As already stated, this small difference is due to the negligible difference in SHG response among the conform-

Table 6.3.: Static first hyperpolarizability values for lowest energy conformer (minimum), Boltzmann weighted ensemble (ensemble) and averaged MD snapshots (MD average) computed with sTD-DFT-xTB/GBSA.

system	minimum	ensemble	MD average
W	100	97	100
KWK	119	113	128
KWWK	149	154	142
KWWWK	181	167	152
KWWKWWK	222	202	198

ers. Second, for KWWKWWK, the small difference in SHG response of ensemble and MD average can be assigned to a shallow PES. The energy difference between the first and the third most contributing conformers is only 0.35 kcal/mol and those conformers amount up to 91% of the entire population (see Figure 6.8). Thus, the MD simulation covers an ensemble that is comparable to the one from the equilibrium structure sampling procedure. However, differences between conformer weighted ensemble and MD average are expected and present for the remaining systems, since the simulation time of 1 ns is rather short to fully explore the conformational PES.

### 6.3.3. Comparison with extrapolated experimental values

In this section, we compare the experimental values to ones computed at the TDHF level with and without accounting for implicit solvation effects and at both sTD-DFT-xTB and sTD-DFT-xTB/GBSA levels. The frequency dispersion of  $\beta$  for all systems, computed with TDHF and sTD-DFT-xTB, is discussed in section A3.5.

As geometries, we consider the lowest energy conformers for each system optimized with the PBEh-3c(COSMO) method. In the case of averaging all relevant conformers, the static  $\beta_{\text{HRS}}$  values were weighted by their respective Boltzmann weights. The static first hyperpolarizabilities (and at 1900 nm for TDHF/IEF-PCM results) are shown in Figure 6.11, which is divided in two panels: one for comparing tryptophan with KWK and one for every systems except tryptophan. First, we compare the  $\beta$  response of tryptophan and KWK. The experimental first hyperpolarizability values reveal a much weaker response (-80%) for KWK compared to tryptophan. This observation is only reproduced by the TDHF/IEF-PCM method, demonstrating the role of non-equilibrium solvent effects, though it yields only a difference of 20%. Due to the flexible side chains of KWK, the chromophore is partially shielded from the solvent. As speculated in the experimental paper, this could have a large effect on the electronic structure of the indole unit and thus

### III. Application of the sTD-DFT-xTB method

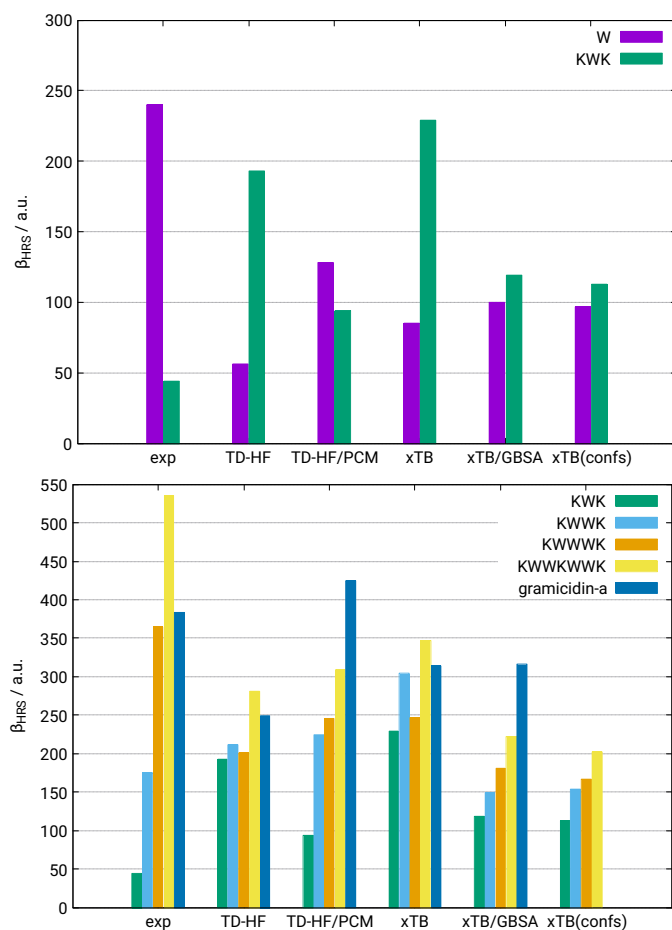


Figure 6.11.: Static first hyperpolarizabilities for tryptophan and KWK (top) and KWK, KWWK, KWWWK, KWWKWWK, gramicidin A (bottom) extrapolated from experiment and computed with TDHF/6-31+G(d), TDHF/6-31+G(d)/PCM, sTD-DFT-xTB, sTD-DFT-xTB/GBSA and Boltzmann weighted ensemble with sTD-DFT-xTB/GBSA.

also on the first hyperpolarizability. However, this hypothesis could not be corroborated by our calculations. The direct follow-up of this could be to investigate this effect by adding an explicit solvent shell. Secondly, the remaining systems are compared. The TDHF method without accounting for solvent effects is not able to reproduce the experimental trends. For a comparison of the frequency dispersion of TDHF and sTD-DFT-xTB see Figure A3.9. Applying an implicit solvation model improves the quality of the computed data. Except for gramicidin A, the experimental ordering is reproduced. The IEF-PCM scheme improves the  $\beta_{\text{HRS}}$  values with respect to the experiment. The sTD-DFT-xTB/GBSA method is also able to provide correct trends except for gramicidin A, but a systematic underestimation with respect to experiment is observed, at least for the largest systems considered here. Note that the value obtained for the rigid gramicidin A is not so far from the experimental one. Including all relevant conformers has a small impact on the  $\beta_{\text{HRS}}$  values. The effect of weighting the conformers becomes more pronounced when changes between indole subunit orientations become important. However, accurate computation of free energies in solution for a proper Boltzmann weighting is difficult. Since the individual conformers largely differ in their  $\beta_{\text{HRS}}$  values, a slightly miscalculated population could lead directly to a bad result

Table A3.5 shows the relative error for all methods. The relative mean absolute errors (MAEs, in a.u.) of all methods are substantial, ranging from 0.94 (TDHF) to 0.46 (TDHF/IEF-PCM). The sTD-DFT-xTB/GBSA method provides a MAE of 0.62 very close to TDHF/IEF-PCM. Advances in including non-equilibrium solvent effects at the sTD-DFT-xTB/GBSA level could close this gap. For (s)TD-DFT, the paramount importance is already proved for calculations of first hyperpolarizabilities<sup>236–239</sup> and other properties.<sup>256–258</sup> This is especially remarkable considering that the sTD-DFT-xTB/GBSA calculations are 3-5 orders of magnitude faster than at the TDHF/IEF-PCM level of theory (see Table A3.3).

## 6.4. Conclusion

We have presented quantum chemical calculations and a structure-property analysis for NLO properties of tryptophan-rich model peptides. The sTD-DFT-xTB scheme enables computations for systems with up to several thousands of atoms and/or to screen large sets of structures. We used this method to sample the first hyperpolarizabilities with respect to structural changes for the flexible tryptophan-rich peptide chains. For this purpose, molecular dynamics simulations as well as conformational sampling were carried out. This was done with the help of the efficient tight-binding based method GFN2-xTB and a recently proposed approach for exploring the potential energy surface with meta-dynamics. To fine-tune the sTD-

### III. Application of the sTD-DFT-xTB method

DFT-xTB scheme for the desired model peptides, two parameters were adjusted to reproduce CCSD(T) reference  $\beta$  values for tryptophan.

We first studied the conformer ensemble in terms of the relation between the relative orientations of indole moieties and the first hyperpolarizability. We found that a drastic change in  $\beta_{\text{HRS}}$  values relates to different alignments of indole dipoles among conformers. When they are aligned parallel, the value is enhanced while it diminishes when the dipole moments are pointing in opposite directions leading to first hyperpolarizability values of conformers that can differ by a factor of up to two. This highlights the importance of finding the lowest energy conformer when calculating SHG response. The unit-sphere representations clearly showed in most of the case the dipolar character of the  $\beta$  tensor. The assumption that the sum of indole dipole vectors is correlated to the intensity of the first hyperpolarizability was confirmed by model system calculations. The results of the MD simulations indicate additionally a very strong sensitivity of the first hyperpolarizability to details of the molecular structure. Overall it seems essential to properly explore the conformational space of flexible chromophores or when multiple chromophore orientations are possible.

In the second part, we compared sTD-DFT-xTB computed first hyperpolarizabilities to the TDHF values and to experiment. sTD-DFT-xTB and TDHF perform similarly in terms of reproducing the experimental trend of SHG response. Cases where the methods produce incorrect order of values with respect to the size of the system was found at both level of theories. This is mostly when solvent effects are not included. The sTD-DFT-xTB method is able to provide SHG response values for tryptophan-rich systems at a fraction of the cost of the usually-used TDHF level of theory. The comparison of theoretical with experimental values shows that getting a quantitative agreement is a challenging task. TDHF/IEF-PCM performs slightly better than sTD-DFT-xTB/GBSA probably due to the inclusion of additional non-equilibrium solvent effects. The sTD-DFT-xTB scheme can be routinely applied to systems that are inaccessible with TDHF, while still expecting reasonable accuracy. Considering this efficiency, a future study could investigate how the relation between the first hyperpolarizability and the number of tryptophan units evolve for larger peptides. This relation is not expected to be linear because of the sensitivity of the property on indole unit orientations. Another area of future research is the impact of an explicit solvation on the conformational ensembles and their SHG response.

### Disclosure statement

The authors declare no competing financial interest.

## **Acknowledgments**

This work was supported by the DFG in the framework of the Gottfried-Wilhelm-Leibniz prize.





## **Part IV.**

# **Final Summary and Conclusion**



Spectroscopic techniques like CD or SHIM are often used in medical research to obtain information about the three-dimensional structure of the systems in solution or *in vivo*. Theoretically simulated spectra can provide additional information about electronic and structural properties of the systems and support experimental studies. In this thesis, the accurate and efficient computation of electronic excitation spectra of entire large biomolecular systems ( $\approx 3000$  atoms) was addressed. Part II has been dedicated to the simulation of CD and UV-Vis spectra with the efficient sTDA-xTB method.<sup>76</sup> In particular, chapter 3 has been devoted to the application of sTDA-xTB to large biomolecular systems like proteins or DNA fragments. The focus of this chapter was to unveil the challenges arising from the simulation of systems of this size and develop possible solutions. The variety of examples in chapter 3 highlighted the all-round applicability of the sTDA-xTB method. No limitation to a certain spectral range or type of chromophore was observed, since excellent results were achieved for protein and DNA CD spectra. The sampling along an MD trajectory was necessary for some cases, because artificial CD transitions on floppy side chains interfere with the main transitions of the peptide backbone. These transitions are present in a single structure approach while they average out by the MD sampling. sTDA-xTB in combination with the structure method GFN1-xTB showed excellent results for this MD sampling approach.

Chapter 4 was dedicated to the extension of the parameter set for the sTDA-xTB method, which allowed the efficient computation of electronic excitation spectra for systems, including almost all elements of the entire periodic table. During the parameterization process, first, atomic charges were fitted to reference charges from DFT calculations and then excitation energies were fitted to match CC2<sup>259</sup>/TD-DFT reference excitation energies. Comparisons to theory and experiment showed that sTDA-xTB provides similar reliable results as for the elements in the original publication<sup>76</sup> with an average deviation of excitation energies of 0.3–0.5 eV. The comparison of UV-Vis absorption and CD spectra for a variety of compounds revealed an accuracy, which is comparable to the much more computationally expensive sTD-DFT method. For the first time, electronic excitation spectra for systems with up to 1700 atoms (comprising the newly parametrized elements) were computed with remarkable agreement with experimental spectra after MD sampling.

In chapter 5, the performance of GFN2-xTB<sup>75</sup> and GFN-FF<sup>114</sup> was tested with respect of providing reasonable structures for the protein CD spectra calculation with sTDA-xTB. The aim of this chapter was the development of an almost black-box approach for protein CD simulation. As shown in chapter 3, the excited states calculation of large biomolecular systems poses no difficulty for sTDA-xTB. However, both the optimization and the MD simulation of the protein structures required in

#### IV. Final Summary and Conclusion

chapter 3 a variety of theoretical levels. The comparison of CD spectra of multiple model peptides and proteins revealed that GFN-FF and GFN2-xTB yield equally reliable structures for this purpose and the computed spectra show an excellent agreement with the experiment. Even the characteristic CD spectra of challenging secondary structure motifs (e.g.,  $3_{10}$ - or collagen-helix) were reproduced well. The necessity of MD sampling was also shown for some systems in this chapter. Such tasks are typically conducted by highly specialized FFs (e.g., OPLS<sup>260</sup> or AMBER variants<sup>113,261</sup>). These methods sacrifice generality for accuracy, so that only a limited range of chemical systems can be treated. Chapter 5 underlined that the GFN-FF method has an outstanding accuracy for an universal FF and is able to provide very reliable structures for subsequent sTDA-xTB calculations at very low computational effort. So that in conclusion, the tandem use of GFN-FF for all structural aspects and sTDA-xTB for the computation of excited state spectra is an ideal combination for an efficient and reliable protein spectra calculation.

Part III addressed the computation of the first hyperpolarizability with the sTD-DFT-xTB<sup>223</sup> method. This property correlates with the SHG response of a molecule, which is important for the medical imaging technique SHIM. In chapter 6, tryptophan-rich peptides – as models for endogenous SHG biotags – were investigated in terms of the structure-property relationship of the first hyperpolarizability. Preceding advances in efficient potential energy surface sampling (CREST<sup>48</sup>) and the development of the sTD-DFT-xTB method enabled this comprehensive investigation for the first time. The latter is capable of computing NLO properties of molecules with up to several thousand of atoms and to screen such properties for large amounts of structures in short time frames (in days for systems up to hundred atoms). In this chapter, it was shown that the indole chromophore of the tryptophan units causes mainly the SHG response of the model peptides. Furthermore, the analysis of conformational sampling and MD simulation in combination with the calculation of first hyperpolarizabilities revealed that the relative orientations of the indole moieties affect the SHG response of the model peptides heavily. Hence, the structure-property relationship of the first hyperpolarizability could be enlightened to a certain extent. Additionally, the comparison to TDHF calculations examined that sTD-DFT-xTB is capable of providing similar accurate results than the higher level of theory, but 3–5 orders of magnitude faster.

At this point, the medical applications from chapter 1 can be recalled, in particular the CD spectrum analysis of the prion protein (PrP). Part II was devoted to the fundamental research of the electronic excitation spectra calculation of large biomolecular systems. The worked-out combination of structure and excited state methods is in the following applied exemplary to the CD spectrum of the prion protein. For this purpose, the structure methods GFN2-xTB and GFN-FF are used for geometry optimization and meta-dynamics (MTD) simula-

tion<sup>230</sup>, respectively. Based on these structures, CD spectra are calculated with sTDA-xTB. The computed spectrum of the natural form of the prion protein is depicted together with the experimental spectrum in Figure 7.1.

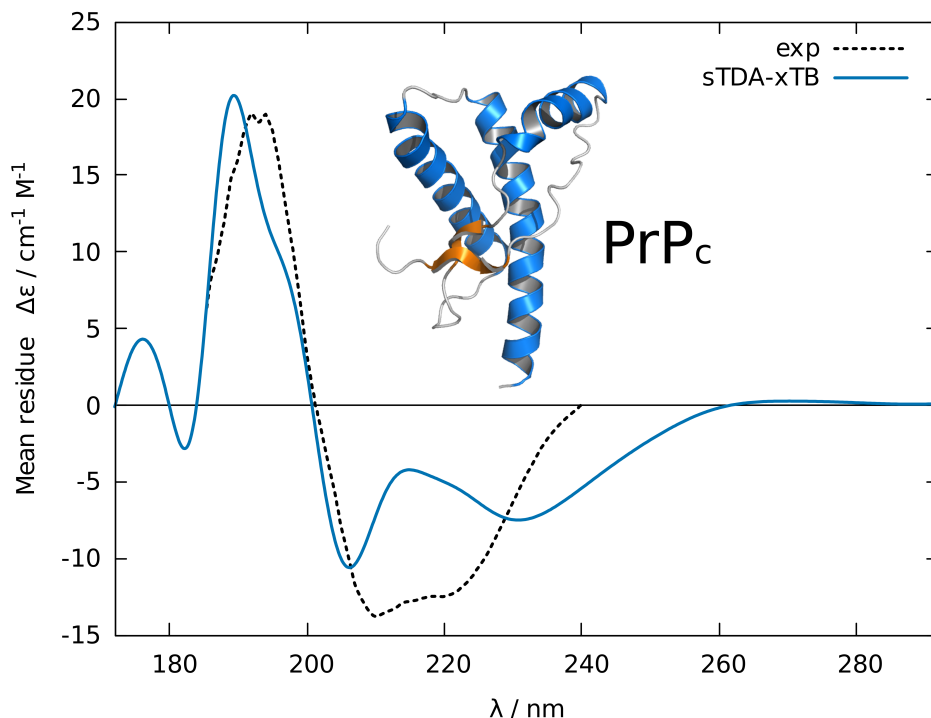


Figure 7.1.: CD spectrum of the prion protein in natural conformation ( $\text{PrP}_c$ ) computed with sTDA-xTB on a GFN2-xTB optimized structure (blue). The individual transition strengths are broadened by Gaussians with a full width at  $1/e$  maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV. The experimental spectrum (black) is taken from Ref.<sup>41</sup> and refers to a solution of  $\text{PrP}_c$  in water.

As depicted in the inset of Figure 7.1, the protein secondary structure of the prion protein comprises mainly  $\alpha$ -helical secondary structure motifs. Hence, the characteristic CD bands of this structure dominates the experimental spectrum. The computed spectrum is in excellent agreement with the experiment and it reproduces the characteristic features with correct intensity and excitation energy. The accuracy of this fully QM-generated spectrum is remarkable, even more so with regard to the size of the molecule ( $\approx 1700$  atoms) and the short amount of time for the computation (20 h for GFN2-xTB structure optimization and 23 h for the spectrum calculation\*). An atomically resolved structure has been discovered for

\*The calculations were performed on a quad core computer (Intel® Xeon® CPU E3-1270 v5 @ 3.6GHz).

#### IV. Final Summary and Conclusion

the natural form of the prion protein<sup>40</sup>, which remains, however, largely unknown for the misfolded (malignant) form ( $\text{PrP}_{sc}$ ).<sup>262</sup> Since a high-resolution structure determination is inaccessible for the malignant form due to insolubility and tendency to aggregate, it is of particular interest to study the folding process and the structure of the malignant form with spectroscopic methods like CD in solution. To some extent, the misfolding of the protein can be approximated with the help of MTD simulations. The additional bias potential<sup>†</sup> enables fast exploration of the PES, and forces the protein structure to unfold. The obtained structures can be used for further excited states calculation to monitor the effect of the misfolding on the CD spectrum. Note that this setup is not representative for the actual natural process of misfolding *in-vivo*. It is rather an example to show the potential of the worked-out theoretical approach of this thesis. Figure 7.2 depicts structures and CD spectra of the prion protein within 10 ps of MTD simulation.

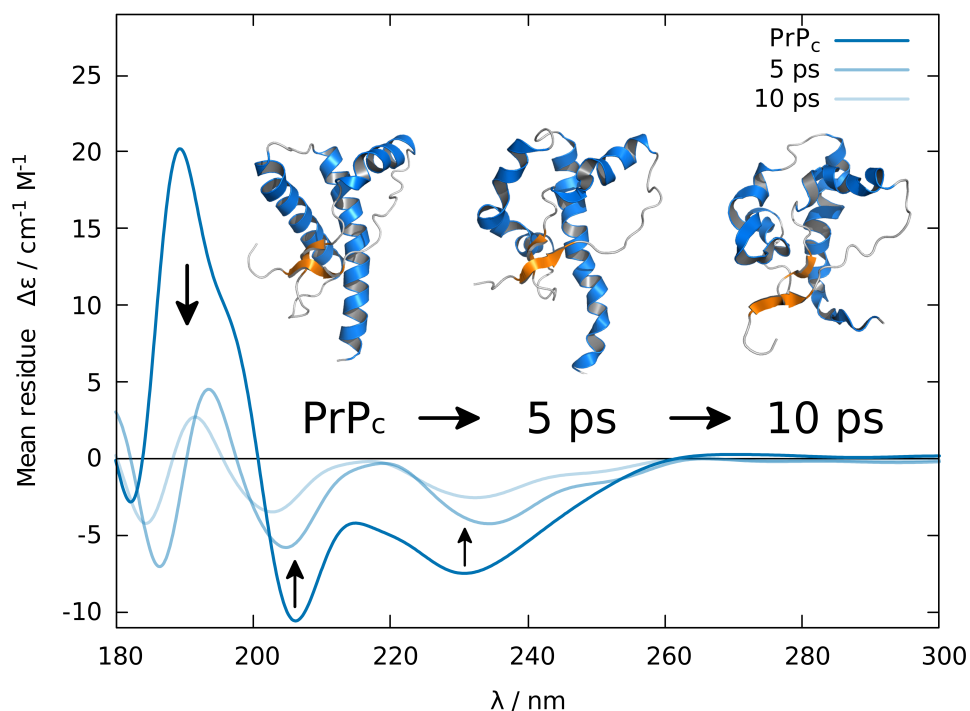


Figure 7.2.: CD spectrum of the prion protein in natural conformation ( $\text{PrP}_c$ ) and after 5 and 10 ps of MTD simulation, computed with sTDA-xTB on GFN-FF structures. The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV.

<sup>†</sup>In MTD simulations, a root-mean-square deviation (RMSD)-based bias potential is added to the electronic energy ( $E_{\text{tot}} = E_{\text{el}} + E_{\text{bias}}^{\text{RMSD}}$ ).

The secondary structure of the prion protein rapidly loses the high  $\alpha$ -helical content and the CD spectrum changes accordingly. This shows once more the high structural sensitivity of CD spectroscopy that is reproduced in the theoretical calculations. Furthermore, this example sketches a potential application of the in this thesis developed methodology for a fast and accurate protein CD spectrum calculation, which is designed to extend experimental medical studies.

All studies of this thesis confirm the remarkable performance of the xTB-based methods for a variety of optical properties. Even large biomolecular systems pose no difficulty and with the help of the GFN based methods, an efficient MD and/or conformational sampling is feasible. As seen in every chapter of this work, this structural sampling is very important for a good agreement with experimentally derived properties. An automatic black-box approach for the calculation of protein spectra is desirable and the findings of the preceding chapters underline that xTB-based methods are excellent candidates for this purpose. Based on the results of this work, a recommended general workflow for the efficient calculation of optical properties is depicted in Figure 7.3.

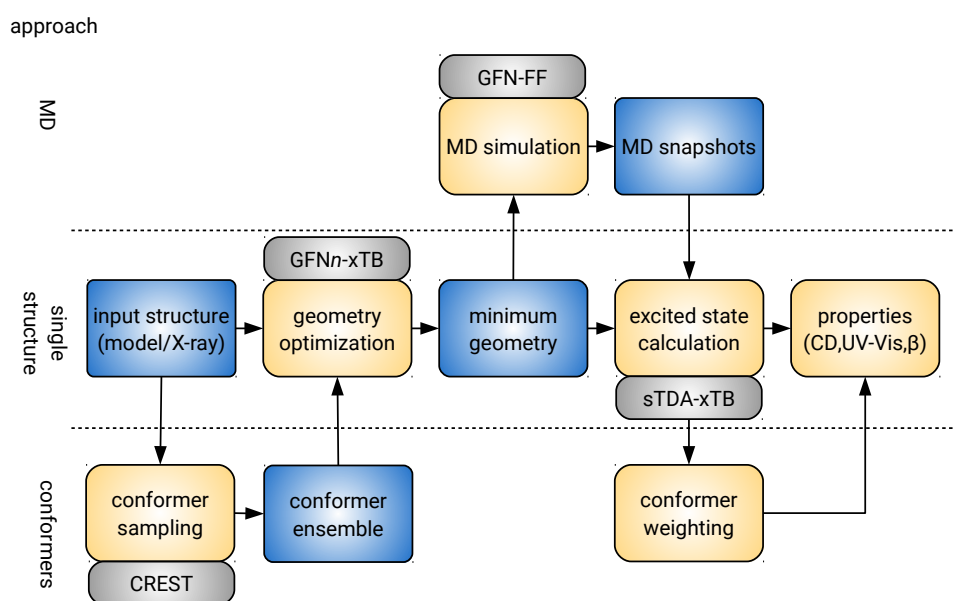


Figure 7.3.: Flowchart for the efficient computation of optical properties.

To improve some parts of this workflow even further, certain aspects could be addressed:

1. sTDA-xTB shows remarkable performance regarding its computational cost.

#### IV. Final Summary and Conclusion

However, the approximations in the ground and excited state reach for some systems their limit (as seen in chapter 4 and Ref. 76). As a possible improvement of the sTDA-xTB method, the basis in the xTB part could be enlarged. This would allow for a more robust treatment of electronically more complicated systems (e.g., transition metal complexes). Additionally, the monopole approximation for the transition moments could be extended by including higher-order multipoles, similar to the CAMMs in GFN2-xTB.

2. The example of the prion protein in this chapter underlined the relevance of an CD monitoring of folding processes. On the one hand, the sTDA-xTB method is fully capable of computing accurate spectra even for unusual or rare protein conformations (cf. chapter 5). However, the underlying structure techniques (MD or MTD) are by default inappropriate or require very long simulation times for the modeling of folding mechanisms.<sup>263,264</sup> Therefore, the PES sampling approach of CREST<sup>48</sup> – which uses the mentioned structural techniques in combination – could be modified for protein folding purposes (e.g., separating the sampling of the protein back-bone and side chains).
3. The research of excited states dynamics remains inaccessible for the sTDA-xTB method, since analytical gradient expressions are not available. Theoretical studies on fluorescent proteins<sup>265,266</sup> or photosystems<sup>267,268</sup> would become possible, if the accuracy and efficiency of the sTDA-xTB method could be transferred to a excited states dynamic version. This deficiency should be addressed for revised versions, mentioned in the first item.

In conclusion, for the first time a full QM treatment of the simulation of CD spectra and NLO properties of large biomolecular systems was reported. Furthermore, all parts of an efficient computational workflow for the calculation of optical properties (cf. Figure 7.3) were comprehensively tested and successfully elaborated. The diversification and benchmarking of the sTDA/sTD-DFT-xTB method in the context of an accurate and efficient computation of optical properties for large biomolecular systems can help computational chemists and biologist for their particular research. Perhaps even new, ground-breaking investigations have been made possible by the fundamental research of this thesis.



## Bibliography

- [1] Waldrop, M. M. *Nature* **2016**, *530*, 144–147.
- [2] Houk, K. N.; Liu, F. *Acc. Chem. Res.* **2017**, *50*, 539–543.
- [3] Thiel, W. *Angew. Chemie - Int. Ed.* **2011**, *50*, 9216–9217.
- [4] Cavalleri, M. *Int. J. Quantum Chem.* **2013**, *113*, 1.
- [5] Jain, A.; Shin, Y.; Persson, K. A. *Nat. Rev. Mater.* **2016**, *1*, 1–13.
- [6] Baik, M. H.; Friesner, R. A. *J. Phys. Chem. A* **2002**, *106*, 7407–7412.
- [7] Cernak, T. A.; Gleason, J. L. *J. Org. Chem.* **2008**, *73*, 102–110.
- [8] Rostamikia, G.; Janik, M. J. *Energy Environ. Sci.* **2010**, *3*, 1262–1274.
- [9] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- [10] Garbuzynskiy, S. O.; Melnik, B. S.; Lobanov, M. Y.; Finkelstein, A. V.; Galzit-skaya, O. V. *Proteins Struct. Funct. Genet.* **2005**, *60*, 139–147.
- [11] Warnke, I.; Furche, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 150–166.
- [12] Berova, N.; Bari, L. D.; Pescitelli, G. *Chem. Soc. Rev.* **2007**, *36*, 914–931.
- [13] Crawford, T. D.; Tam, M. C.; Abrams, M. L. *J. Phys. Chem. A* **2007**, *111*, 12057–12068.
- [14] Grimme, S. In *Rev. Comput. Chem.*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2004; Vol. 20; pp 153–218.
- [15] Autschbach, J. *Chirality* **2009**, *21*, 116–152.
- [16] Autschbach, J.; Nitsch-Velasquez, L.; Rudolph, M. *Top. Curr. Chem.* **2011**, *298*, 1–98.
- [17] Pescitelli, G.; Di Bari, L.; Berova, N. *Chem. Soc. Rev.* **2014**, *43*, 5211–5233.
- [18] Pescitelli, G.; Bruhn, T. *Chirality* **2016**, *28*, 466–474.
- [19] Woody, R. W.; Toniolo, C.; Formaggio, F.; Woody, R. W. In *Compr. Chiroptical Spectrosc. Vol. 2 Appl. Stereochem. Anal. Synth. Compd.*; N. Berova P. L. Polavarapu, K. N., Woody, R. W., Eds.; Wiley: New York, 2012; pp 499–574.
- [20] Miles, A. J.; Wallace, B. A. Circular dichroism spectroscopy of membrane proteins. 2016; [www.rsc.org/chemsocrev](http://www.rsc.org/chemsocrev).

## Bibliography

- [21] Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophys. Acta* **2005**, *1751*, 119–139.
- [22] Ranjbar, B.; Gill, P. *Chem. Biol. Drug Des.* **2009**, *74*, 101–120.
- [23] Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205–211.
- [24] Wallace, B. A.; Janes, R. W. *Curr. Opin. Chem. Biol.* **2001**, *5*, 567–571.
- [25] Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D. E.; Baumgärtel, P.; Lengefeld, J.; Reichardó, G.; Horsley, D. A.; Seckler, R.; Bakajin, O.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 105–110.
- [26] Riley, M. L.; Wallace, B. A.; Flitsch, S. L.; Booth, P. J. *Biochemistry* **1997**, *36*, 192–196.
- [27] Sreerama, N.; Manning, M. C.; Powers, M. E.; Zhang, J.-X.; Goldenberg, D. P.; Woody, R. W. *Biochemistry* **1999**, *38*, 10814–10822.
- [28] Greenfield, N. J. *Nat. Protoc.* **2007**, *1*, 2527–2535.
- [29] Ellett, L. J.; Johanssen, V. A. *Methods Mol. Biol.* **2017**, *1658*, 27–34.
- [30] Johnson, W. C. *Proteins Struct. Funct. Bioinforma.* **1990**, *7*, 205–214.
- [31] Martin, S. R.; Schilstra, M. J. *Methods Cell Biol.* **2008**, *84*, 263–293.
- [32] Yerabham, A. S.; Mas, P. J.; Decker, C.; Soares, D. C.; Weiergräber, O. H.; Nagel-Steger, L.; Willbold, D.; Hart, D. J.; Bradshaw, N. J.; Korth, C. *J. Biol. Chem.* **2017**, *292*, 6468–6477.
- [33] Petchanikow, C.; Saborio, G. P.; Anderes, L.; Frossard, M. J.; Olmedo, M. I.; Soto, C. *FEBS Lett.* **2001**, *509*, 451–456.
- [34] Prusiner, S. B. *Science (80- )*. **1997**, *278*, 245–251.
- [35] Horwich, A. L.; Weissman, J. S. *Cell* **1997**, *89*, 499–510.
- [36] Prusiner, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363–13383.
- [37] Mastrianni, J. A.; Nixon, R.; Layzer, R.; Telling, G. C.; Han, D.; DeArmond, S. J.; Prusiner, S. B. *N. Engl. J. Med.* **1999**, *340*, 1630–1638.
- [38] Babelhadj, B.; Di Bari, M. A.; Pirisinu, L.; Chiappini, B.; Gaouar, S. B. S.; Riccardi, G.; Marcon, S.; Agrimi, U.; Nonno, R.; Vaccari, G. *Emerg. Infect. Dis.* **2018**, *24*, 1029–1036.
- [39] Nitrini, R.; Rosemberg, S.; Passos-Bueno, M. R.; Teixeira Da Silva, L. S.; Iughetti, P.; Papadopoulos, M.; Carrilho, P. M.; Caramelli, P.; Albrecht, S.; Zatz, M.; LeBlanc, A. *Ann. Neurol.* **1997**, *42*, 138–146.
- [40] Zahn, R.; Liu, A.; Lührs, T.; Riek, R.; Von Schroetter, C.; Garcia, F. L.; Billeter, M.; Calzolari, L.; Wider, G.; Wüthrich, K. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 145–150.

- [41] Jackson, G. S.; Hill, A. F.; Joseph, C.; Hosszu, L.; Power, A.; Waltho, J. P.; Clarke, A. R.; Collinge, J. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1999**, *1431*, 1–13.
- [42] Schellman, J. A. *Chem. Rev.* **1975**, *75*, 323–331.
- [43] Rosenfeld, L. Z. *Phys.* **1929**, *52*, 161–174.
- [44] Laurent, A. D.; Jacquemin, D. *Int. J. Quantum Chem.* **2013**, *113*, 2019–2039.
- [45] Yoshikawa, T.; Kobayashi, M.; Fujii, A.; Nakai, H. *J. Phys. Chem. B* **2013**, *117*, 5565–5573.
- [46] Ahmadi, F.; Jafari, B.; Rahimi-Nasrabadi, M.; Ghasemi, S.; Ghanbari, K. *Toxicol. Vitr.* **2013**, *27*, 641–650.
- [47] Giovannini, T.; Olszowska, M.; Cappelli, C. *J. Chem. Theory Comput.* **2016**, *12*, 5483–5492.
- [48] Pracht, P.; Bohle, F.; Grimme, S. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.
- [49] Sreerama, N.; Woody, R. W. *Methods Enzym.* **2004**, *383*, 318–351.
- [50] Franken, P. A.; Hill, A. E.; Peters, C. W.; Weinreich, G. *Phys. Rev. Lett.* **1961**, *7*, 118–119.
- [51] Bloembergen, N.; Chang, R. K.; Jha, S. S.; Lee, C. H. *Phys. Rev.* **1968**, *174*, 813–822.
- [52] Freund, I.; Deutsch, M. *Opt. Lett.* **1986**, *11*, 94.
- [53] Cohen, B. E. *Nature* **2010**, *467*, 407–408.
- [54] Pantazis, P.; Maloney, J.; Wu, D.; Fraser, S. E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14535–14540.
- [55] Campagnola, P. J.; Loew, L. M. *Nat. Biotechnol.* **2003**, *21*, 1356–1360.
- [56] Reeve, J. E.; Anderson, H. L.; Clays, K. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13484–13498.
- [57] Campagnola, P. *Anal. Chem.* **2011**, *83*, 3224–3231.
- [58] Pavone, F. S.; Campagnola, P. J. *Second harmonic generation imaging*; CRC Press, 2013.
- [59] Deniset-Besseau, A.; Duboisset, J.; Benichou, E.; Hache, F.; Brevet, P.-F.; Schanne-Klein, M.-C. *J. Phys. Chem. B* **2009**, *113*, 13437–13445.
- [60] de Wergifosse, M.; De Ruyck, J.; Champagne, B. *J. Phys. Chem. C* **2014**, *118*, 8595–8602.
- [61] Harczuk, I.; Vahtras, O.; Ågren, H. *J. Phys. Chem. Lett.* **2016**, *7*, 2132–2138.

## Bibliography

- [62] Alowami, S.; Troup, S.; Al-Haddad, S.; Kirkpatrick, I.; Watson, P. H. *Breast Cancer Res.* **2003**, *5*, R129.
- [63] Provenzano, P. P.; Eliceiri, K. W.; Campbell, J. M.; Inman, D. R.; White, J. G.; Keely, P. J. *BMC Med.* **2006**, *4*, 1–15.
- [64] Tokarz, D.; Cisek, R.; Joseph, A.; Golaraei, A.; Mirsanaye, K.; Krouglov, S.; Asa, S. L.; Wilson, B. C.; Barzda, V. *Front. Oncol.* **2019**, *9*, 272.
- [65] Keikhosravi, A.; Bredfeldt, J. S.; Sagar, A. K.; Eliceiri, K. W. *Methods Cell Biol.*; Academic Press Inc., 2014; Vol. 123; pp 531–546.
- [66] Nadiarnykh, O.; LaComb, R. B.; Brewer, M. A.; Campagnola, P. J. *BMC Cancer* **2010**, *10*, 94.
- [67] Papadopoulos, M., Sadlej, A., Leszczynski, J., Eds. *Non-Linear Optical Properties of Matter*, vol. 1 ed.; Springer, Dodrecht, 2006.
- [68] Plaquet, A.; Guillaume, M.; Champagne, B.; Rougier, L.; Mançois, F.; Rodriguez, V.; Pozzo, J. L.; Ducasse, L.; Castet, F. *J. Phys. Chem. C* **2008**, *112*, 5638–5645.
- [69] Sergeyev, S.; Didier, D.; Boitsov, V.; Teshome, A.; Asselberghs, I.; Clavs, K.; Velde, C. M.; Plaquet, A.; Champagne, B. *Chem. - A Eur. J.* **2010**, *16*, 8181–8190.
- [70] Mançois, F.; Sanguinet, L.; Pozzo, J. L.; Guillaume, M.; Champagne, B.; Rodriguez, V.; Adamietz, F.; Ducasse, L.; Castet, F. *J. Phys. Chem. B* **2007**, *111*, 9795–9802.
- [71] Loison, C.; Simon, D. *J. Phys. Chem. A* **2010**, *114*, 7769–7779.
- [72] Grimme, S. *J. Chem. Phys.* **2013**, *138*, 244104.
- [73] Bannwarth, C.; Grimme, S. *Comput. Theor. Chem.* **2014**, *1040-1041*, 45–53.
- [74] Grimme, S.; Bannwarth, C.; Shushkov, P. *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- [75] Bannwarth, C.; Ehlert, S.; Grimme, S. *J. Chem. Theory Comput.* **2018**, *15*, 1652–1671.
- [76] Grimme, S.; Bannwarth, C. *J. Chem. Phys.* **2016**, *145*, 054103.
- [77] Schrödinger, E. *Ann. Phys.* **1926**, *384*, 361–376.
- [78] Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *389*, 457–484.
- [79] Slater, J. C. *Phys. Rev.* **1929**, *34*, 1293–1322.
- [80] Pauli, W. *Zeitschrift für Phys.* **1925**, *31*, 765–783.
- [81] Hartree, D. R.; Hartree, W. *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.* **1935**, *150*, 9–33.

- [82] Roothaan, C. C. *Rev. Mod. Phys.* **1951**, *23*, 69–89.
- [83] Koch, W.; Holthausen, M. C. *A Chem. Guid. to Density Funct. Theory*; Wiley-VCH, 2001; p 300.
- [84] Jensen, F. *Introduction to Computational Chemistry 2nd Ed.*; John Wiley & Sons Ltd.: Chichester, 2007.
- [85] Thomas, L. H. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542–548.
- [86] Fermi, E. *Zeitschrift für Phys.* **1928**, *48*, 73–79.
- [87] Dirac, P. A. *Math. Proc. Cambridge Philos. Soc.* **1930**, *26*, 376–385.
- [88] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*.
- [89] Perdew, J. P.; Schmidt, K. *AIP Conf. Proc.* **2001**, *577*, 1–20.
- [90] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.
- [91] Grimme, S.; Steinmetz, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16031–16042.
- [92] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- [93] Grimme, S. *J. Chem. Phys.* **2006**, *124*.
- [94] Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- [95] Marques, M. A., Ullrich, C. A., Nogueira, F., Rubio, A., Burke, K., Gross, E. K. U., Eds. *Time-Dependent Density Functional Theory*; Lecture Notes in Physics; Springer Berlin Heidelberg: Berlin, Heidelberg, 2006; Vol. 706.
- [96] Casida, M. E. In *Recent Adv. Density Funct. Methods*; Chong, D. P., Ed.; World Scientific: Singapore, 1995; Vol. 1; pp 155–192.
- [97] Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009–4037.
- [98] Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett* **1996**, *256*, 454–464.
- [99] Fetter, A. L.; Walecka, J. D. *Quantum Theory of Many-Particle Systems*; McGraw-Hill: New York, 1971.
- [100] Hirata, S.; Head-Gordon, M. *Chem. Phys. Lett.* **1999**, *314*, 291–299.
- [101] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- [102] Hubbard, J. *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **1963**, *276*, 238–257.
- [103] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- [104] Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- [105] Jones, J. E. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1924**, *106*, 463–477.
- [106] Mermin, N. D. *Phys. Rev.* **1965**, *137*.

## Bibliography

- [107] Caldeweyher, E.; Bannwarth, C.; Grimme, S. *J. Chem. Phys.* **2017**, *147*.
- [108] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. *J. Chem. Phys.* **2019**, *150*.
- [109] Koester, A. M.; Leboeuf, M.; Salahub, D. R. In *Molecular electrostatic potentials : concepts and applications*, volume 3 ed.; Murray, J. S., Sen, K., Eds.; Elsevier, 1996; pp 105–142.
- [110] Peacock, A. F. A.; Pecoraro, V. L. *Natural and Artificial Proteins Containing Cadmium*; Springer, Dordrecht, 2013; pp 303–337.
- [111] Freisinger, E.; Vašák, M. *Cadmium in Metallothioneins*; Springer, Dordrecht, 2013; pp 339–371.
- [112] Sure, R.; Grimme, S. *J. Comput. Chem.* **2013**, *34*, 1672–1685.
- [113] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- [114] Spicher, S.; Grimme, S. *Angew. Chemie Int. Ed.* **2020**, in print.
- [115] Bulheller, B. M.; Hirst, J. D. *Bioinformatics* **2009**, *25*, 539–540.
- [116] Woody, R. W.; Sreerama, N. *J. Chem. Phys.* **1999**, *111*, 2844.
- [117] Kurapkat, G.; Krüger, P.; Wollmer, A.; Fleischhauer, J.; Kramer, B.; Zobel, E.; Koslowski, A.; Botterweck, H.; Woody, R. W. *Biopolymers* **1997**, *41*, 267–287.
- [118] Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. *J. Phys. Chem. B* **2010**, *114*, 8270–8277.
- [119] Chin, D.-H.; Woody, R. W.; Rohl, C. A.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15416–15421.
- [120] Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95–99.
- [121] Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- [122] Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
- [123] Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15519–15523.
- [124] Imamoto, Y.; Kataoka, M. *Photochem. Photobiol.* **2007**, *83*, 40–49.
- [125] Borucki, B.; Otto, H.; Meyer, T. E.; Cusanovich, M. A.; Heyn, M. P. *J. Phys. Chem. B* **2005**, *109*, 629–633.

- [126] Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327–335.
- [127] Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *0*, 799–805.
- [128] Shushkov, P.; Grimme, S. manuscript in preparation.
- [129] Still, W.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- [130] Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.
- [131] Körzdörfer, T.; Sears, J. S.; Sutton, C.; Brédas, J. L. *J. Chem. Phys.* **2011**, *135*, 204107.
- [132] Brkljača, Z.; Čondić-Jurkić, K.; Smith, A. S.; Smith, D. M. *J. Chem. Theory Comput.* **2012**, *8*, 1694–1705.
- [133] Glättli, A.; Daura, X.; Seebach, D.; Van Gunsteren, W. F. *J. Am. Chem. Soc.* **2002**, *124*, 12972–12978.
- [134] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [135] Chu, K.; Vojtchovský, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921–923.
- [136] Borgstahl, G. E. O. G.; Williams, D. R. D.; Getzoff, E. E. D. *Biochemistry* **1995**, *34*, 6278–6287.
- [137] Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.
- [138] Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.
- [139] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- [140] The xtb code can be obtained from: <https://github.com/grimme-lab/xtb>.
- [141] Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.
- [142] Stewart, J. J. P. *J. Mol. Mod.* **2013**, *19*, 1–32.
- [143] Voityuk, A. A. *J. Chem. Theory Comput.* **2014**, *10*, 4950–4958.
- [144] Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. *Phys. Rev. B* **2001**, *63*, 85108.
- [145] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

## Bibliography

- [146] Stewart, J. J. P. *J. Mol. Mod.* **2007**, *13*, 1173.
- [147] Gadaczek, I.; Krause, K.; Hintze, K. J.; Bredow, T. *J. Chem. Theory Comput.* **2011**, *7*, 3675–3685.
- [148] Christensen, A. S.; Kubar, T.; Cui, Q.; Elstner, M. *Chem. Rev.* **2016**, *116*, 5301–5337.
- [149] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- [150] Wahiduzzaman, M.; Oliveira, A. F.; Philipsen, P.; Zhechkov, L.; van Lenthe, E.; Witek, H. A.; Heine, T. *J. Chem. Theory Comput.* **2013**, *9*, 4006.
- [151] Oliveira, A. F.; Philipsen, P.; Heine, T. *J. Chem. Theory Comput.* **2015**, *11*, 5209–5218.
- [152] Nishimoto, K.; Mataga, N. *Z. Phys. Chem.* **1957**, *12*, 335–338.
- [153] Ohno, K. *Theor. Chim. Acta* **1964**, *2*, 219–227.
- [154] Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550–4557.
- [155] Löwdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365–375.
- [156] Ghosh, D. C.; Islam, N. *Int. J. Quantum Chem.* **2010**, *110*, 1206–1213.
- [157] Risthaus, T.; Hansen, A.; Grimme, S. *Phys. Chem. Chem. Phys.* **2014**, *28*, 14408–14419.
- [158] Gasteiger, J.; Marsili, M. *J. Comput. Chem.* **1978**, *19*, 3181–3184.
- [159] Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.
- [160] See: <http://www.thch.uni-bonn.de/>.
- [161] Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. *J. Chem. Phys.* **2015**, *143*, 54107.
- [162] Deglmann, P.; Furche, F.; Ahlrichs, R. *Chem. Phys. Lett.* **2002**, *362*, 511–518.
- [163] Deglmann, P.; Furche, F. *J. Chem. Phys.* **2002**, *117*, 9535–9538.
- [164] Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129–138.
- [165] Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
- [166] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- [167] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [168] Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514–518.



- [169] Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124.
- [170] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- [171] Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. *Theor. Chim. Acta* **1990**, *77*, 123–141.
- [172] Dolg, M.; Stoll, H.; Savin, A.; Preuss, H. *Theor. Chim. Acta* **1989**, *75*, 173–194.
- [173] Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1993**, *85*, 441–450.
- [174] Leininger, T.; Nicklass, A.; Küchle, W.; Stoll, H.; Dolg, M.; Bergner, A. *Chem. Phys. Lett.* **1996**, *255*, 274–280.
- [175] Kaupp, M.; Schleyer, P. v. R.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1991**, *94*, 1360–1366.
- [176] Metz, B.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2000**, *113*, 2563–2569.
- [177] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113–11123.
- [178] Metz, B.; Schweizer, M.; Stoll, H.; Dolg, M.; Liu, W. *Theor. Chem. Acc.* **2000**, *104*, 22–28.
- [179] Hellweg, A.; Grün, S. A.; Hättig, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4119–4127.
- [180] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- [181] Furche, F.; Rappoport, D. In *Theor. Comput. Chem. Comput. Photochem.*; Olivucci, M., Ed.; Elsevier: Amsterdam, 2005; Vol. 16; Chapter III, pp 93–128.
- [182] Hättig, C.; Weigend, F. *J. Chem. Phys.* **2000**, *113*, 5154–5161.
- [183] Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- [184] Levenberg, K. *Quart. Appl. Math.* **1944**, *2*, 164–168.
- [185] Marquardt, D. *SIAM J. Appl. Math.* **1963**, *11*, 431–441.
- [186] Chong, D. P. *Mol. Phys.* **2005**, *103*, 749–761.
- [187] Goerigk, L.; Grimme, S. *J. Chem. Phys.* **2010**, *132*, 184103.
- [188] Brückner, C.; Engels, B. *J. Phys. Chem. A* **2015**, *119*, 12876–12891.
- [189] Wiberg, K. B. *Tetrahedron* **1968**, *24*, 1083–1096.
- [190] Thomas, F.; Schulz, S.; Mansikkamäki, H.; Nieger, M. *Angew. Chemie - Int. Ed.* **2003**, *42*, 5641–5644.

## Bibliography

- [191] Honda, Y.; Kurihara, A.; Hada, M.; Nakatsuji, H. *J. Comp. Chem.* **2008**, *29*, 612–621.
- [192] Rudolph, M.; Autschbach, J. *J. Phys. Chem. A* **2011**, *115*, 2635–2649.
- [193] Le Guennic, B.; Hieringer, W.; Görling, A.; Autschbach, J. *J. Phys. Chem. A* **2005**, *109*, 4836–4846.
- [194] Graule, S.; Rudolph, M.; Vanthuyne, N.; Autschbach, J.; Roussel, C.; Crassous, J.; Réau, R. *J. Am. Chem. Soc.* **2009**, *131*, 3183–3185.
- [195] Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87–94.
- [196] Gütz, C. et al. *Angew. Chem. Int. Ed.* **2014**, *53*, 1693–1698.
- [197] Seibert, J.; Bannwarth, C.; Grimme, S. *J. Am. Chem. Soc.* **2017**, *139*, 11682–11685.
- [198] Schrödinger Release 2016-2: Schrödinger Suite 2016-2 Protein Preparation Wizard; Epik version 3.6, Schrödinger, LLC, New York, NY, 2016; Impact version 7.1, Schrödinger, LLC, New York, NY, 2016; Prime version 4.4, Schrödinger, LLC, New York, NY, 2016.
- [199] Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *J. Comput. Mol. Des.* **2013**, *27*, 221–234.
- [200] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- [201] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [202] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- [203] van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311–1327.
- [204] Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A. *Bioinformatics* **2006**, *22*, 1955–1962.
- [205] The stda code can be obtained from: <https://github.com/grimme-lab/stda/>.
- [206] Holzwarth, G.; Doty, P. *J. Am. Chem. Soc.* **1965**, *87*, 218–228.
- [207] Schellman, J. A.; Oriol, P. *J. Chem. Phys.* **1962**, *37*, 2114–2124.
- [208] Woody, R. R. W.; Tinoco Jr., I.; Woody, R. R. W. *J. Chem. Phys.* **1967**, *46*, 4927–4945.
- [209] Moffitt, W. *J. Chem. Phys.* **1956**, *25*, 467–478.
- [210] Tinoco, I.; Woody, R. W.; Bradley, D. F. *J. Chem. Phys.* **1963**, *38*, 1317–1325.

- [211] Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L. *Stryer Biochem.*; Springer Berlin Heidelberg, 2018.
- [212] Toniolo, C.; Polese, A.; Formaggio, F.; Crisma, M.; Kamphuis, J. *J. Am. Chem. Soc.* **1996**, *118*, 2744–2745.
- [213] Toniolo, C.; Formaggio, F.; Woody, R. W. *Compr. Chiroptical Spectrosc.*; John Wiley and Sons: Hoboken, NJ, USA, 2012; Vol. 2; pp 499–544.
- [214] Plonska-Brzezinska, M. E.; Bobrowska, D. M.; Sharma, A.; Rodziewicz, P.; Tomczyk, M.; Czyrko, J.; Brzezinski, K. *RSC Adv.* **2015**, *5*, 95443–95453.
- [215] Schmitz, S.; Seibert, J.; Ostermeir, K.; Hansen, A.; Göller, A. H.; Grimme, S. *J. Phys. Chem. B* **2020**, acs.jpcc.0c00549.
- [216] Morris, J. A.; Martenson, R.; Deibler, G.; Cagan, R. H. *J. Biol. Chem.* **1973**, *248*, 534–539.
- [217] Hobbs, J. R.; Munger, S. D.; Conn, G. L. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 162–167.
- [218] Eakin, R. E.; Snell, E. E.; Williams, *The Discovery of Avidin by Esmond E. Snell A Constituent of Raw Egg White Capable of Inactivating Biotin in Vitro Classics 100 Years of Biochemistry and Molecular Biology Downloaded from*; 1940; Vol. 136; p 5.
- [219] Snell, E. E.; Eakin, R. E.; Williams, R. J. *J. Am. Chem. Soc.* **1940**, *62*, 175–178.
- [220] Miller, C. *Neuron* **1995**, *15*, 5–10.
- [221] Dames, S. A.; Mulet, J. M.; Rathgeb-Szabo, K.; Hall, M. N.; Grzesiek, S. *J. Biol. Chem.* **2005**, *280*, 20558–20564.
- [222] Stockert, J. C.; Blazquez-Castro, A. *Fluoresc. Microsc. Life Sci.*; Bentham Science Publishers, 2017.
- [223] de Wergifosse, M.; Grimme, S. *J. Chem. Phys.* **2018**, *149*, 024108.
- [224] de Wergifosse, M.; Botek, E.; De Meulenaere, E.; Clays, K.; Champagne, B. *J. Phys. Chem. B* **2018**, *122*, 4993–5005.
- [225] Duboisset, J.; Matar, G.; Besson, F.; Fichoux, D.; Benichou, E.; Russier-Antoine, I.; Jonin, C.; Brevet, P. F. *J. Phys. Chem. B* **2014**, *118*, 10413–10418.
- [226] De Meulenaere, E.; Asselberghs, I.; de Wergifosse, M.; Botek, E.; Spaepen, S.; Champagne, B.; Vanderleyden, J.; Clays, K. *J. Mater. Chem.* **2009**, *19*, 7514–7519.
- [227] De Meulenaere, E.; de Wergifosse, M.; Botek, E.; Spaepen, S.; Champagne, B. B.; Vanderleyden, J. J.; Clays, K. *J. Nonlinear Opt. Phys. Mater.* **2010**, *19*, 1–13.
- [228] De Meulenaere, E.; de Wergifosse, M.; Botek, E.; Vanderleyden, J.; Champagne, B.; Clays, K. **2015**, 522–525.

## Bibliography

- [229] De Meulenaere, E.; Nguyen Bich, N.; de Wergifosse, M.; Van Hecke, K.; Van Meervelt, L.; Vanderleyden, J.; Champagne, B.; Clays, K. *J. Am. Chem. Soc.* **2013**, *135*, 4061–4069.
- [230] Grimme, S. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
- [231] Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. *Angew. Chemie - Int. Ed.* **2017**, *56*, 14763–14769.
- [232] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.
- [233] Eckert, F.; Klamt, A. *AIChE J.* **2002**, *48*, 369–385.
- [234] COSMOtherm, C3.0, release 1601, COSMOlogic GmbH & Co KG, <http://www.cosmologic.de>.
- [235] Grimme, S. *Chem. - A Eur. J.* **2012**, *18*, 9955–9964.
- [236] Yu, J.; Zerner, M. C. *J. Chem. Phys.* **1994**, *100*, 7487–7494.
- [237] Zhu, W.; Wu, G. S. *Chem. Phys. Lett.* **2002**, *358*, 1–7.
- [238] Yamaguchi, Y.; Yokomichi, Y.; Yokoyama, S.; Mashiko, S. *J. Mol. Struct. THEOCHEM* **2002**, *578*, 35–45.
- [239] Cammi, R.; Cossi, M.; Mennucci, B.; Tomasi, J. *J. Mol. Struct.* **1997**, *436-437*, 567–575.
- [240] de Wergifosse, M.; Liégeois, V.; Champagne, B. *Int. J. Quantum Chem.* **2014**, *114*, 900–910.
- [241] Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
- [242] Cancès, E.; Mennucci, B. *J. Chem. Phys.* **2001**, *114*, 4744–4745.
- [243] Verbiest, T.; Clays, K.; Rodriguez, V. *Second-order nonlinear optical characterization techniques: an introduction*; CRC press, 2009.
- [244] Bersohn, R.; Pao, Y.; Frisch, H. L. *J. Chem. Phys.* **1966**, *45*, 3184–3198.
- [245] Castet, F. F.; Bogdan, E.; Plaquet, A.; Ducasse, L.; Champagne, B.; Rodriguez, V. *J. Chem. Phys.* **2012**, *136*, 24506.
- [246] Tuer, A.; Krouglov, S.; Cisek, R.; Tokarz, D.; Barzda, V. *J. Comput. Chem.* **2011**, *32*, 1128–1134.
- [247] Guthmuller, J.; Simon, D. *J. Phys. Chem. A* **2006**, *110*, 9967–9973.
- [248] Willetts, A.; Rice, J. E.; Burland, D. M.; Shelton, D. P. *J. Chem. Phys.* **1992**, *97*, 7590–7599.
- [249] Shao, Y. et al. *Mol. Phys.* **2015**, *113*, 184–215.
- [250] Frisch, M. J. et al. Gaussian 09. 2009.

- [251] Campo, J.; Wenseleers, W.; Goovaerts, E.; Szablewski, M.; Cross, G. H. *J. Phys. Chem. C* **2008**, *112*, 287–296.
- [252] Mançois, F.; Pozzo, J. L.; Pan, J.; Adamietz, F.; Rodriguez, V.; Ducasse, L.; Castet, F.; Plaquet, A.; Champagne, B. *Chem. - A Eur. J.* **2009**, *15*, 2560–2571.
- [253] Oudar, J. L.; Chemla, D. S. *J. Chem. Phys.* **1977**, *66*, 2664–2668.
- [254] Bannwarth, C.; Grimme, S.; Seibert, J.; Grimme, S. *Chirality* **2016**, *28*, 365–369.
- [255] Seibert, J.; Pisarek, J.; Schmitz, S.; Bannwarth, C.; Grimme, S. *Mol. Phys.* **2019**, *117*, 1104–1116.
- [256] Caricato, M. *Comput. Theor. Chem.* **2014**, *1040-1041*, 99–105.
- [257] Improta, R.; Scalmani, G.; Frisch, M. J.; Barone, V. *J. Chem. Phys.* **2007**, *127*, 074504.
- [258] Ponce-Vargas, M.; Azarias, C.; Jacquemin, D.; Le Guennic, B. *J. Phys. Chem. B* **2017**, *121*, 10850–10858.
- [259] Christiansen, O.; Koch, H.; Jørgensen, P. *Chem. Phys. Lett.* **1995**, *243*, 409–418.
- [260] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [261] McDonald, D.; Still, W. *Tetrahedron Lett.* **1992**, *33*, 7743–7746.
- [262] Wille, H.; Requena, J. R. *Pathogens* **2018**, *7*, 1–11.
- [263] Dill, K. A.; Maccallum, J. L. *Science (80-. )*. **2012**, *338*, 1042–1047.
- [264] Bian, Y.; Zhang, J.; Wang, J.; Wang, W. *Mol. Simul.* **2015**, *41*, 752–763.
- [265] Tonge, P. J.; Meech, S. R. *J. Photochem. Photobiol. A Chem.* **2009**, *205*, 1–11.
- [266] Jonasson, G.; Teuler, J. M.; Vallverdu, G.; Mérola, F.; Ridard, J.; Lévy, B.; Demachy, I. *J. Chem. Theory Comput.* **2011**, *7*, 1990–1997.
- [267] Hastings, G.; Reed, L. J.; Lin, S.; Blankenship, R. E. *Biophys. J.* **1995**, *69*, 2044–2055.
- [268] Vasil'ev, S.; Orth, P.; Zouni, A.; Owens, T. G.; Bruce, D. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8602–8607.
- [269] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- [270] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [271] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

## *Bibliography*

- [272] Peach, M. J.; Williamson, M. J.; Tozer, D. J. *J. Chem. Theo. Comp.* **2011**, *7*, 3578–3585.
- [273] Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2011**, *7*, 1814–1823.
- [274] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. *Chem. Rev.* **2016**, *116*, 5105–5154.

**Part V.**

**Appendix**

# A1. Supporting Information to Chapter 3

## Appendix A1 contains:

- Technical details of the calculations
- Supplementary figures
- Additional discussion

## A1.1. Computational details

One NMR structure (1L2Y) was taken from the Protein Data Bank (PDB)<sup>121,134</sup> and optimized at the HF-3c level of theory<sup>112</sup> applying the COSMO implicit solvation model ( $\epsilon = 78$ )<sup>127</sup>. The optimized structure was taken as input for the xTB ground-state calculation with the GBSA model for water.<sup>128</sup> All excitations up to 8 eV were calculated with the sTDA method<sup>76</sup>. The velocity formalism is used for the rotatory strengths. The starting structure of myoglobin was generated by taking the X-ray structure of a horse myoglobin (PDB code: 1DWR<sup>135</sup>) and adding the missing hydrogen atoms via a force-field as it is implemented in the UCSF chimera software (version 1.7<sup>201</sup>). With GFN-xTB<sup>74</sup> the geometry was optimized with implicit solvent (GBSA). A molecular dynamics (MD) simulation was carried out for 6 ps with a time step of 1 fs. From the resulting trajectory 100 snapshots were taken equidistantly and considered as input structures for the sTDA-xTB calculations. In order to save memory the default diffuse basis functions on all hydrogen atoms were neglected. All excitations up to 9 eV were calculated in the sTDA part. The resulting rotatory strengths are scaled by a factor of 3 and the length formalism is used for the rotatory strengths. As initial structure of the photoactive yellow protein (PYP) the X-ray diffraction structure 2PHY<sup>136</sup> from the PDB was used. Hydrogen atoms were added using Chimera. Two histidine residues, which are located on the water-accessible surface of the protein, were protonated as well, resulting in a total charge of  $-4$  of the protein. The full protein was then optimized with the AMBER ff12SB<sup>113</sup> force-field as implemented in Chimera (ANTECHAMBER<sup>269</sup> was used to determine the force-field parameters of the chromophore). Based on this structure, a cut-out of the chromophore and adjacent residues (including Y42, E46, T50, R52, P54, F62, F63, A67, P68, C69+HC4, T70, F75, Y94, F96, Y98, Q99, and M100) was reoptimized at the HF-3c level<sup>112</sup> including the implicit COSMO solvation model ( $\epsilon = 4$ ).<sup>127</sup> The ORCA program package (version 3.0)<sup>165</sup> was used for this purpose. In this cut-out, the backbone peptides were capped with methyl groups which were placed at the C $^{\alpha}$  atom position of the neighboring residues. The backbone carbon atoms were kept fixed during the QM



optimization. The resulting geometry (excluding capping groups) was then transferred back to the full protein structure and the remaining protein was allowed to relax using the AMBER ff12SB force field.<sup>113</sup> The optimized structure was taken as input for the xTB ground-state calculation with the GBSA model for water.<sup>128</sup> All excitations up to 9 eV were calculated with the sTDA method<sup>76</sup>. The length formalism is used for the rotatory strengths. The intensities are displayed in arbitrary units to match the intensity of the most intense band. The structure of poly[d(A)].poly[d(T)] is build up from scratch with the help of symmetry transformations available in the CRYSTAL<sup>137</sup> program and optimized at the 1D-periodic HF-3c level of theory. The optimization is carried out using the rotational-translational symmetry with 11 base pairs per unit cell and containing 72 water molecules in the grooves of the helix. After the optimization a fragment containing 12 base pair is prepared and saturated at the cut surface and the phosphate backbone, esulting a neutral system. This fragment geometry is used in the sTDA-xTB procedure. The convergence of the results with helix length has been tested carefully and no significant changes for chains longer than 8-10 repeating units are found See Figure A1.4). All excited-states up to 10 eV are computed in the sTDA step. The length formalism for the rotatory strengths is used and all rotatory strengths are scaled by a factor of 0.2. If not stated otherwise the TURBOMOLE<sup>138,139</sup> suite of programs (version 7.0) is used for all geometry optimizations with HF-3c while the CRYSTAL<sup>137</sup> program is used for the periodic optimization of the DNA. For geometry optimization, MD simulation, ground state calculation with the xTB method, the stand-alone program xtb<sup>140</sup> is used. All excited-state calculations are conducted with the stda stand alone program<sup>72</sup>.

## A1.2. Rotatory strengths

Due to the origin independence, the velocity formalism is generally recommend to use for large systems. The respective intensities obtained in a Tamm-Dancoff approximated framework, as previously presented<sup>76</sup>, are improved. It is always preferential to compare both representations. Nevertheless, the underlying xTB method, in particular the basis set (minimal+diffuse functions) may affect the magnitudes for the velocity intensities, which are known to converge slower w.r.t. to the complete basis set.

### A1.3. Photoactive yellow protein

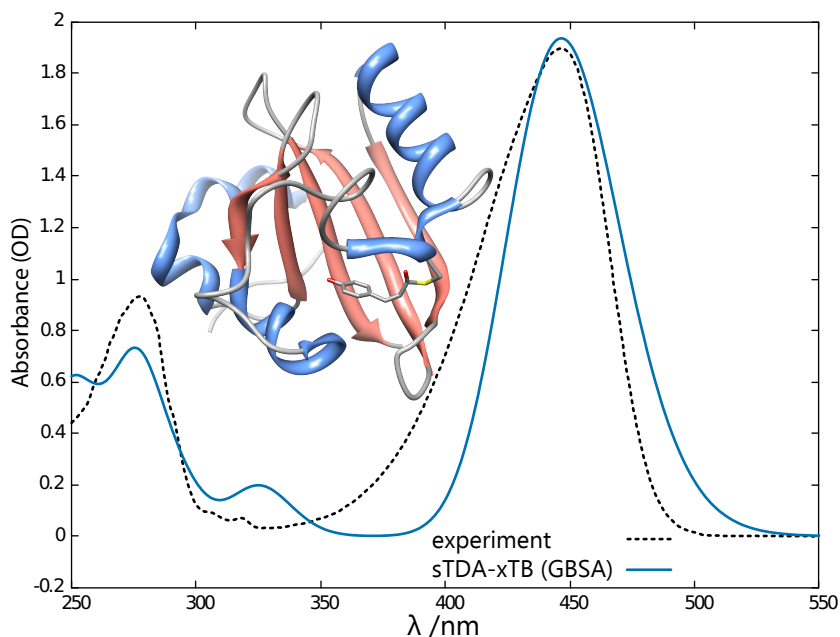


Figure A1.1.: Calculated absorption spectrum of photoactive yellow protein (PYP) (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV. The experimental spectrum (black dotted line) is taken from ref 125 and refer to a solution of PYP in water.

One can state that employing TD-DFT as underlying electronic structure method for spectra would generally be a worse choice than sTDA-xTB for the large and possibly charged systems. For the PYP system using the sTDA approach based on a B3LYP<sup>270,271</sup> and BHLYP<sup>92</sup> reference, the computed absorption and ECD spectrum (<6 eV) is shown in Figure A1.2. The energy of the visible state computed with B3LYP is red-shifted w.r.t. BHLYP, which is in agreement from the well-known results on small systems (i.e., more Fock exchange results in larger singlet excitation energies, see refs. 44,272). The well-known self-interaction error (SIE), which leads to artificially low-lying charge-transfer ('ghost') states in TD-DFT has a strong influence on the respective spectra. While for BHLYP (more Fock exchange, less SIE than B3LYP) the lowest-lying state corresponds to the visible one and the absorption spectrum is reproduced well by sTDA-BHLYP, the artificial mixing with only slightly higher-lying 'ghost' states results in the wrong ECD intensity of the first band. In the case of B3LYP, the low-lying energy range is entirely contaminated by 'ghost' states and the electrically allowed state is preceded by 175 non-visible 'ghost' states. Though some fortuitous error cancellation is observed for the ECD spectrum, the artificial mixing with 'ghost' states drains the intensity from the first absorption band. It is important to note, that this problem is not due to the simplified excited-state treatment but to the underlying ground-state method. Even a conventional

### A1.3. Photoactive yellow protein

TD-B3LYP calculation, results in non-visible ghost states as the lowest excitations (see intensities in SI of ref. 273). Our tight-binding approach is effectively corrected for this error due to the level-shift of the virtual orbitals. Hence, we consider it to provide more consistent results for small and large systems than DFT methods.

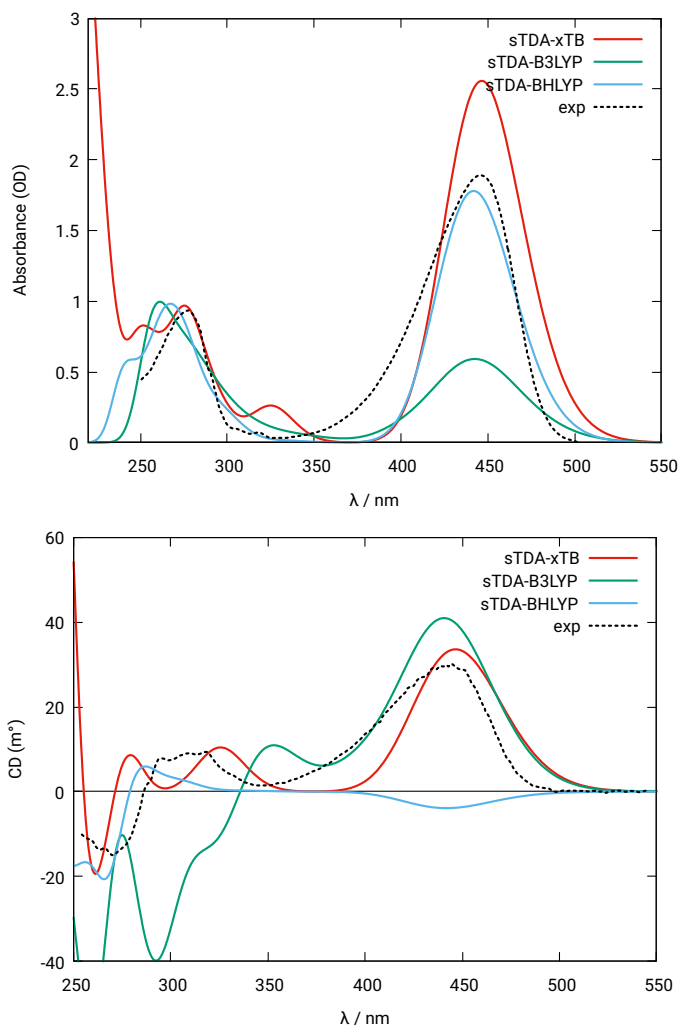


Figure A1.2.: Calculated absorption and CD spectrum of photoactive yellow protein (PYP) with sTDA-xTB (red solid line), sTDA-B3LYP (green solid line) and sTDA-BHLYP (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at  $1/e$  maximum of 0.5 eV and the spectra are red-shifted by 0.5 eV, 0.6 eV, 0.9 eV, respectively. The experimental curves (black dotted line) are taken from ref 125 and refer to a solution of PYP in water.

A1. Supporting Information to Chapter 3

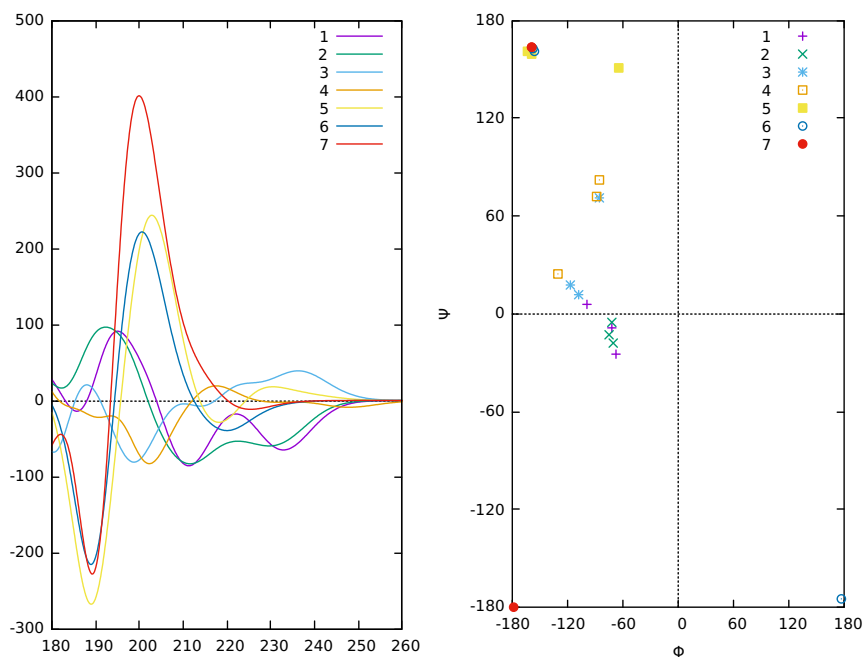


Figure A1.3.: Seven structures of a tripeptide (MeNH-Ala-Gly-Ala-Ac) representing geometry motifs when going from an folded helical to an unfolded structure. CD spectra of the seven structures (left) and corresponding Ramachandran-angles (right) are given. Structures taken from ref 274.

## A1.4. DNA

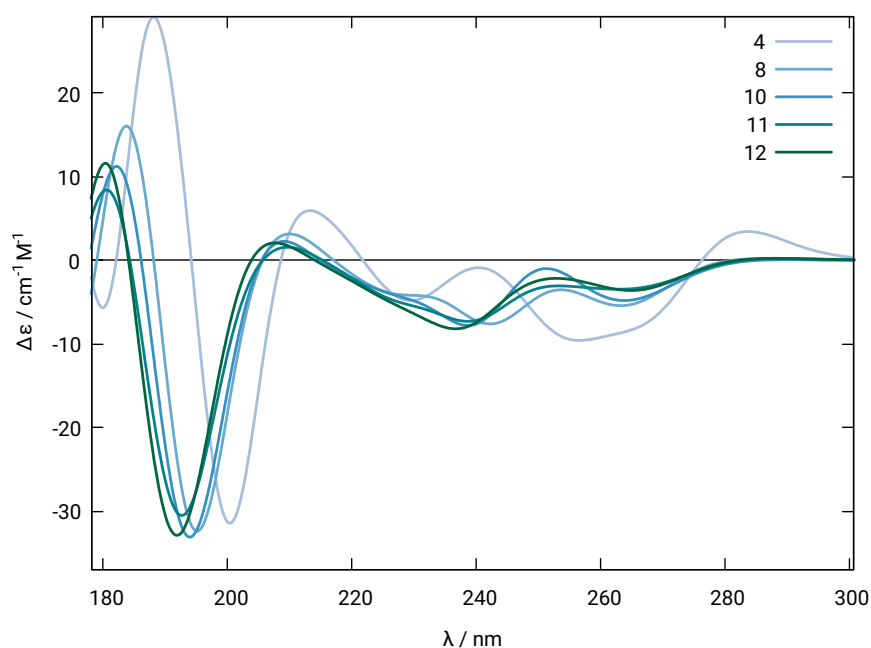


Figure A1.4.: Convergence test of the poly[d(A)].poly[d(T)] fragment. CD spectra, computed with sTDA-xTB for 4–12 base pairs. The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.6 eV.

## A1.5. 1L2Y Protein

As presented in Figure A1.5 the CD spectra based on the GFN-xTB optimized structures show a large deviation from each other. The direct comparison of all minimum structures shows that the main differences are present in the side chains, not in the secondary structure. Including non-equilibrium effects indicate that such side chain transitions are "washed out" over the sampling process. Short MD simulations were performed (200 ps) to get a qualitative picture how the sampling would affect the spectra. Already this crude sampling shows the convergence into the same spectrum. As mentioned before, this presented sTDA-xTB should give the ability to treat such systems in a reasonable time frame.

A1. Supporting Information to Chapter 3

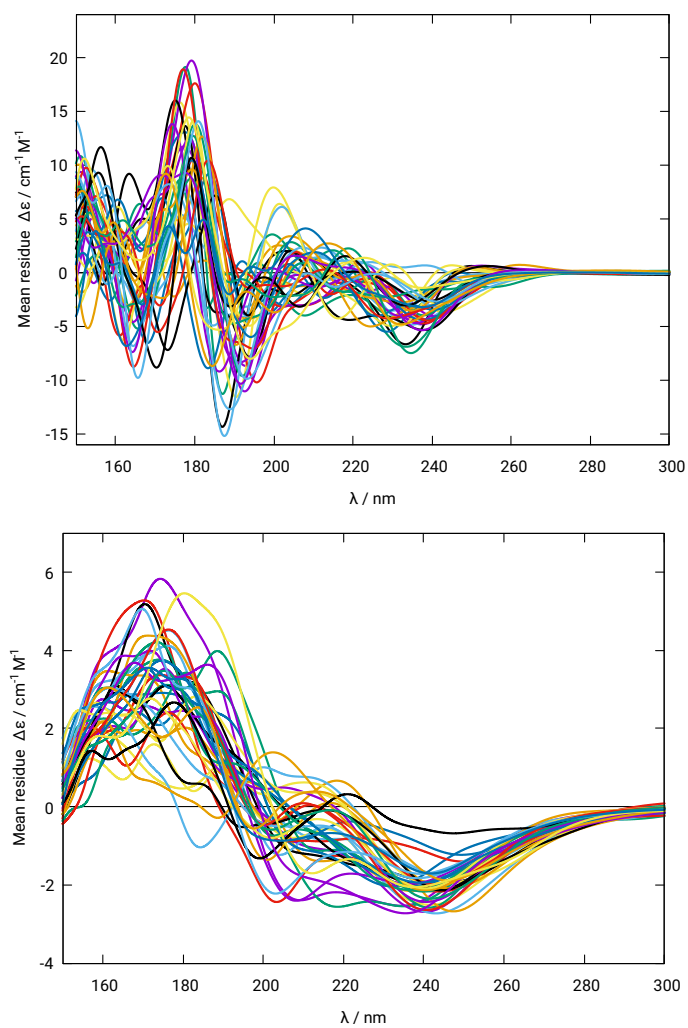


Figure A1.5.: CD spectra of 1L2Y peptide computed with sTDA-xTB on single structures (top) and 200 MD snapshots (bottom). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and the spectrum is red-shifted by 0.5 eV.

## A2. Supporting Information to Chapter 4

**Appendix A2 contains:**

- Additional technical details
- Supplementary figures
- Parameters

### A2.1. Additional technical details

If not noted otherwise, Geometries of all structures are optimized using the GFN-xTB<sup>74</sup> method including the implicit solvent model Generalized Born with an solvent accessible surface term (GB/SA).

**MD details for tellura-/selena-hydrindanes:**

After equilibration, MD simulations are performed with GFN-xTB for 1 ns with a timestep of 3 fs, applying SHAKE<sup>202,203</sup> algorithm. Following that, 100 equidistant MD snapshots are taken from the MD trajectories. The average of these snapshots (averaged peak positions as well as peak intensities) is then used for the calculation of absorption and CD spectra with sTDA-xTB.

**MD details for palladium complexes:**

After equilibration, MD simulations are performed with GFN-xTB for 10 ps with a timestep of 1 fs. Following that, 100 equidistant MD snapshots are taken from the MD trajectories. The average of these snapshots (averaged peak positions as well as peak intensities) is then used for the calculation of absorption and CD spectra with sTDA-xTB. For comparison, spectra on minimum structures are computed as well. These structures are optimized with HF-3c in the gas phase. Excitation energies are calculated up to 8 eV.

The calculation of absorption spectra is performed using the dipole length formalism whereas the corrected dipole velocity formalism<sup>76</sup> is applied to compute CD spectra.

## A2. Supporting Information to Chapter 4

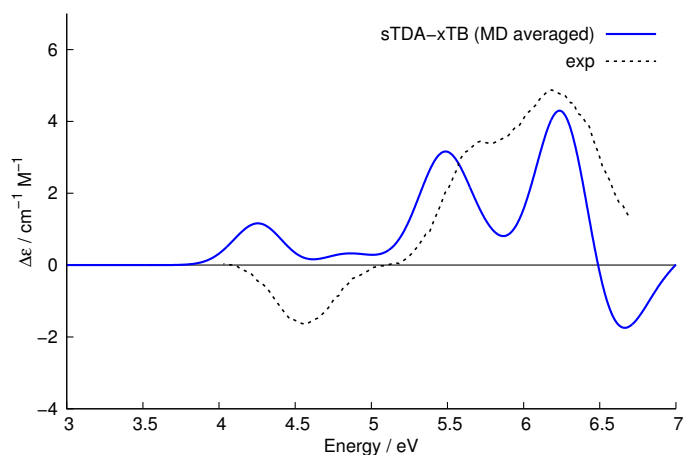


Figure A2.1.: Calculated CD spectrum of (2)-(3aS,7aS)-2-selena-trans-hydrindan (blue solid line) averaged along a GFN-xTB trajectory. The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV. The experimental curve (black dotted line) are taken from Ref. 191 and refer to a solution of (2)-(3aS,7aS)-2-tellura-trans-hydrindan in acetonitrile.

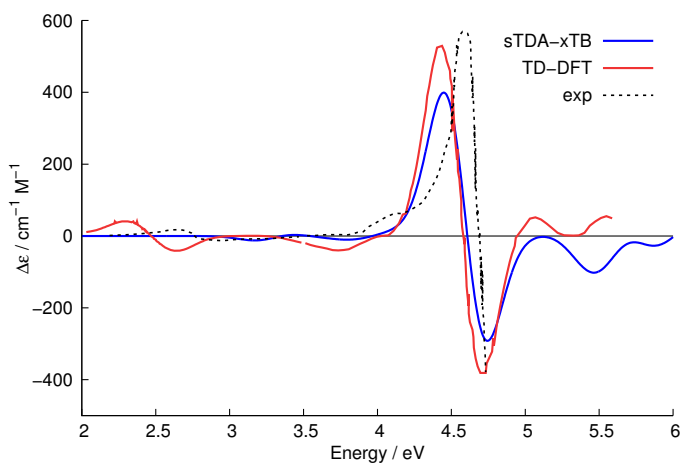


Figure A2.2.: Calculated CD spectrum of  $\Lambda$ -Os(phen)<sub>3</sub><sup>2+</sup> (blue solid line). The individual transition strengths are broadened by Gaussians with a full width at 1/e maximum of 0.5 eV and are red-shifted by 0.5 eV. The experimental (black dotted line) and TD-DFT (red solid line) spectra are taken from Ref. 192.



## A2.2. Parameters

### A2.2.1. VTB: element-specific parameters

Z=31  
 ao=4s4p4d  
 lev=-14.846373 -8.786294 6.870767  
 exp= 2.242350 2.065439 2.065439  
 EN= 2.212598

Z=32  
 ao=4s4p4d  
 lev= -16.317672 -11.984139 4.208333  
 exp= 2.576293 1.522858 1.522858  
 EN= 0.688632

Z=33  
 ao=4s4p4d  
 lev= -15.119432 -12.803798 1.676292  
 exp= 2.228765 1.867027 1.867027  
 EN= 2.132531

Z=34  
 ao=4s4p4d  
 lev= -16.871366 -15.977721 -0.853608  
 exp= 1.770725 1.935918 1.935918  
 EN= 1.315075

Z=37  
 ao=5s5p  
 lev= -7.135595 -3.953318  
 exp= 1.345821 1.446335  
 EN= 0.020648

Z=38  
 ao=5s5p  
 lev= -7.457055 -7.308730  
 exp= 1.218493 1.698185  
 EN= 1.086353

Z=39  
 ao=4d5s5p  
 lev= -7.051743 -11.449444 -12.099399

exp= 2.048162 0.438990 0.744938  
 EN= 0.336539

Z=40  
 ao=4d5s5p  
 lev= -7.883002 -11.370946 -10.393650  
 exp= 2.135242 0.579571 0.906228  
 EN= 0.976517

Z=41  
 ao=4d5s5p  
 lev= -8.714261 -11.292448 -8.687900  
 exp= 2.222323 0.720151 1.067517  
 EN= 0.617678

Z=42  
 ao=4d5s5p  
 lev= -9.545521 -11.213950 -6.982151  
 exp= 2.309403 0.860732 1.228807  
 EN= 1.760781

Z=43  
 ao=4d5s5p  
 lev= -10.376780 -11.135452 -5.276402  
 exp= 2.396483 1.001313 1.390096  
 EN= 0.871721

Z=44  
 ao=4d5s5p  
 lev= -11.208039 -11.056953 -3.570652  
 exp= 2.483563 1.141893 1.551386  
 EN= 1.316153

Z=45  
 ao=4d5s5p  
 lev= -12.039299 -10.978455 -1.864903  
 exp= 2.570644 1.282474 1.712676  
 EN= 2.320132

Z=46  
 ao=4d5s5p  
 lev= -12.870558 -10.899957 -0.159153  
 exp= 2.657724 1.423054 1.873965  
 EN= 0.125760

## A2. Supporting Information to Chapter 4

Z=47  
ao=4d5s5p  
lev= -13.701817 -10.821459 1.546596  
exp= 2.744804 1.563635 2.035255  
EN= 3.321728

Z=48  
ao=5s5p  
lev= -21.079306 -6.088509  
exp= 1.785289 0.612280  
EN= 1.159922

Z=49  
ao=5s5p5d  
lev= -19.376234 -7.197254 1.563211  
exp= 1.985939 2.242653 2.242653  
EN= 0.865035

Z=50  
ao=5s5p5d  
lev= -19.624845 -10.403894 8.275821  
exp= 2.361587 1.642049 1.642049  
EN= 0.078094

Z=51  
ao=5s5p5d  
lev= -19.750923 -13.664446 4.173547  
exp= 2.491235 1.617820 1.617820  
EN= 0.684316

Z=52  
ao=5s5p5d  
lev= -17.243039 -16.126007 0.019253  
exp= 2.608204 1.875164 1.875164  
EN= 0.420335

Z=55  
ao=6s6p  
lev= -6.867561 -4.938313  
exp= 1.555359 1.658885  
EN= 0.016939

Z=56  
ao=6s6p  
lev= -8.696186 -7.822302

exp= 1.395024 1.822277  
EN= 0.483521

Z=57  
ao=5d6s6p  
lev= -9.629399 -10.250765 0.391868  
exp= 1.809168 0.317399 0.358573  
EN= 0.015693

Z=72  
ao=5d6s6p  
lev= -11.744438 -10.160739 0.361513  
exp= 1.901027 0.541660 0.723636  
EN= 1.234912

Z=73  
ao=5d6s6p  
lev= -13.859478 -10.070712 0.331159  
exp= 1.992886 0.765921 1.088698  
EN= 1.089737

Z=74  
ao=5d6s6p  
lev= -15.974517 -9.980686 0.300804  
exp= 2.084745 0.990182 1.453760  
EN= 1.526869

Z=75  
ao=5d6s6p  
lev= -18.089557 -9.890659 0.270449  
exp= 2.176604 1.214444 1.818823  
EN= 1.365010

Z=76  
ao=5d6s6p  
lev= -20.204597 -9.800633 0.240094  
exp= 2.268464 1.438705 2.183886  
EN= 2.271872

Z=77  
ao=5d6s6p  
lev= -22.319636 -9.710607 0.209740  
exp= 2.360323 1.662966 2.548948  
EN= 2.312089

Z=78  
 ao=5d6s6p  
 lev= -24.434675 -9.620580 0.179385  
 exp= 2.452182 1.887227 2.914010  
 EN= 2.343199

Z=79  
 ao=5d6s6p  
 lev= -26.549715 -9.530554 0.149030  
 exp= 2.544041 2.111488 3.279073  
 EN= 2.745512

Z=80  
 ao=6s6p  
 lev= -22.481684 -4.404783  
 exp= 2.472997 0.594810  
 EN= 1.993152

Z=81  
 ao=6s6p5d  
 lev= -17.851551 -7.671243 3.309653  
 exp= 2.673133 2.378773 2.378773  
 EN= 2.394022

Z=82  
 ao=6s6p5d  
 lev= -20.237407 -8.995878 0.219954  
 exp= 2.541206 1.867835 1.867835  
 EN= 0.118079

Z=83  
 ao=6s6p5d  
 lev= -23.081861 -12.616958 2.223805  
 exp= 2.857106 1.688527 1.688527  
 EN= 0.709619

Z=84  
 ao=6s6p5d  
 lev= -16.883971 -14.132950 -5.979790  
 exp= 2.897842 2.121365 2.121365  
 EN= 0.438103

Z=85  
 ao=6s6p5d  
 lev= -17.000000 -14.447474 -5.981650

exp= 2.700000 2.645541 2.645541  
 EN= 1.062702

### A2.2.2. XTB: element-specific parameters

Z=31  
 ao=4s4p  
 lev= -15.434720 -9.328277  
 exp= 1.211204 1.935872

Z=32  
 ao=4s4p5sp  
 lev= -16.761771 -10.938012 -4.085576  
 exp= 1.435626 1.971698 0.676600

Z=33  
 ao=4s4p5sp  
 lev= -22.227421 -12.586141 -3.947731  
 exp= 1.400519 2.089501 0.746700

Z=34  
 ao=4s4p5sp  
 lev= -24.924300 -14.275287 -3.986999  
 exp= 2.166243 1.915023 0.809100

Z=39  
 ao=4d5s5p  
 lev= -10.141034 -9.376363 -2.581727  
 exp= 1.977311 1.325319 0.975122

Z=40  
 ao=4d5s5p  
 lev= -10.723506 -9.709181 -3.045962  
 exp= 2.144307 1.430613 1.085118

Z=41  
 ao=4d5s5p  
 lev= -11.305979 -10.042000 -3.510197  
 exp= 2.311303 1.535908 1.195113

Z=42

## A2. Supporting Information to Chapter 4

ao=4d5s5p

lev= -11.888451 -10.374818 -3.974433

exp= 2.478299 1.641202 1.305109

Z=43

ao=4d5s5p

lev= -12.470923 -10.707636 -4.438668

exp= 2.645295 1.746496 1.415105

Z=44

ao=4d5s5p

lev= -13.053396 -11.040455 -4.902903

exp= 2.812290 1.851790 1.525100

Z=45

ao=4d5s5p

lev= -13.635868 -11.373273 -5.367138

exp= 2.979286 1.957085 1.635096

Z=46

ao=4d5s5p

lev= -14.218341 -11.706092 -5.831374

exp= 3.146282 2.062379 1.745091

Z=47

ao=4d5s5p

lev= -14.800813 -12.038910 -6.295609

exp= 3.313278 2.167673 1.855087

Z=48

ao=5s5p

lev= -13.949240 -8.425540

exp= 1.930502 2.542935

Z=49

ao=5s5p

lev= -15.616760 -8.195664

exp= 1.624682 1.094100

Z=50

ao=5s5p

lev= -16.929130 -9.611976

exp= 1.428280 2.363711

Z=51

ao=5s5p6sp

lev= -18.790829 -11.966088 -4.353164

exp= 1.556710 2.169621 0.717727

Z=52

ao=5s5p6sp

lev= -24.051353 -13.570469 -4.975528

exp= 2.268192 1.815329 0.974796

Z=57

ao=5d6s6p

lev= -7.940169 -9.490222 -1.909889

exp= 1.998119 1.476280 0.983580

Z=72

ao=5d6s6p

lev= -8.892959 -10.010039 -2.165240

exp= 2.170330 1.569600 1.083197

Z=73

ao=5d6s6p

lev= -9.845749 -10.529856 -2.420592

exp= 2.342542 1.662920 1.182814

Z=74

ao=5d6s6p

lev= -10.798538 -11.049673 -2.675943

exp= 2.514753 1.756240 1.282431

Z=75

ao=5d6s6p

lev= -11.751328 -11.569489 -2.931294

exp= 2.686964 1.849560 1.382048

Z=76

ao=5d6s6p

lev= -12.704118 -12.089306 -3.186645

exp= 2.859175 1.942880 1.481665

Z=77

ao=5d6s6p

lev= -13.656908 -12.609123 -3.441997

exp= 3.031386 2.036200 1.581282

Z=78

ao=5d6s6p  
lev= -14.609697 -13.128940 -3.697348  
exp= 3.203598 2.129520 1.680899

Z=79  
ao=5d6s6p  
lev= -15.562487 -13.648757 -3.952699  
exp= 3.375809 2.222840 1.780516

Z=80  
ao=6s6p  
lev= -12.933314 -5.775036  
exp= 2.034551 2.203538

Z=81  
ao=6s6p  
lev= -14.839499 -8.586386  
exp= 1.802645 1.107880

Z=82  
ao=6s6p  
lev= -28.906345 -11.540444  
exp= 2.368286 2.767297

Z=83  
ao=6s6p  
lev= -17.073904 -11.716014  
exp= 2.555552 2.740773

Z=84  
ao=6s6p5sp  
lev= -17.551795 -13.384357 -6.696071  
exp= 3.167994 1.629991 1.257796

Z=85  
ao=6s6p5sp  
lev= -24.500000 -14.257558 -7.377690  
exp= 1.227343 3.105721 0.780900

A2. Supporting Information to Chapter 4

Table A2.1.: Molecules and reference excitations of the fit set for XTB parameters.

Molecule	Excitation energy [eV]	Oscillator strength [au]
Arsabenzol	4.3151	0.001300
Arsenobetain	6.1636	0.000200
Arsenocholin	7.5176	0.001600
Arsol_C4H4AsH	5.2171	0.035300
As2H2Me2	5.5581	0.282300
As4	4.7885	0.000000
As4O6(Adamantan)	5.9439	0.000000
As4S4	3.8935	0.005700
AsCl3	6.0368	0.055900
AsF3	8.0214	0.195300
AsF5	9.2371	0.000000
AsH3	6.8936	0.158900
AsH3O	5.7319	0.039800
AsHMeEt	6.0712	0.055300
AsMe3	6.1851	0.003600
biimidazol	4.0061	0.042800
bithiazol	4.7477	0.111600
bitriazin	3.1584	0.000000
CH2AsH	4.7944	0.013100
dipen	5.2385	0.000000
glyphos	6.2886	0.013600
h3po4	6.5519	0.062800
p2me2	3.1502	0.000000
pbc2h8	5.9824	0.091100
pnh2me2	5.4158	0.000500
ppyr	5.2177	0.035500
Arsabenzol.2	5.0029	0.204100
CH2AsH.2	5.7703	0.224800
at2	2.1904	0.004000
at2.2	6.4018	0.696800
atcl	2.6167	0.001700
atcl.2	7.5317	0.575100
ath	5.0763	0.014600
ati	2.3854	0.003000
atme	4.4995	0.001900
atme.2	6.1672	0.122100
atoh	2.7953	0.000900
phat	4.3767	0.001300
bi2	2.6416	0.000000
bi2.2	4.3809	0.067300
bi2cl6	4.3127	0.036000
bi2h2me2	4.5814	0.188500
bi4	3.4838	0.000000
bibc2h8	5.3615	0.085600
bibetaine	4.3678	0.068800
bicl3	5.6612	0.071800
bicl5	2.8686	0.000000
bien	5.1407	0.075800
biet2_2	4.4043	0.028400
bif3	7.3855	0.137600
bih2ch3	5.8896	0.104400
bih3	6.6347	0.074000
bih3h2o	6.2200	0.156600
bih3hcl	6.4787	0.214300
biimidazol	3.3518	0.013200
bime3	5.7422	0.149500
bime_3_ring	3.6795	0.004400
bimefe_co_4_2_ring	2.1247	0.000000
binh2me2	4.7102	0.038900

## A2.2. Parameters

bipH2nh2sh	4.2724	0.035700
bipyridin	3.8236	0.001600
bithiazol	3.8916	0.158900
bitriazin	2.1530	0.000000
a	2.4874	0.000300
bi4o6	4.8771	0.000000
bi4s6	3.6586	0.000000
bime_4_ring	2.7449	0.000000
al2me4	6.9783	0.000000
al2me4.2	7.1362	0.312500
al2o4	3.3606	0.000000
alcl3nh3	5.0894	0.000900
alh3ph3	7.1839	0.000000
alh3ph3.2	7.7918	0.143100
ch3alh2	5.4965	0.001500
ga15	4.3169	0.129700
ga2cl2h4	6.9904	0.000000
ga2cl6	7.2987	0.026900
ga2cl6.2	7.6328	0.381300
ga2h6	7.1732	0.000000
ga2h6.2	8.1870	0.290500
ga2me4h2	6.9749	0.000000
gacl3	6.9812	0.000000
ga_cn_3	6.2435	0.000000
ga_cycle	3.7313	0.141700
gaf3	8.6615	0.000000
gaf3.2	8.7638	0.134800
gah2me	5.4958	0.001500
gah3	5.7223	0.000000
gah3.2	8.5550	0.212400
gah3nh3	6.7908	0.021300
game3	5.6702	0.000000
game3.2	6.6046	0.100100
ga_nc_3	6.4620	0.000000
ga_nc_3.2	7.8335	0.138800
ga_oh_3	6.6192	0.000000
ga_oh_3.2	7.0928	0.150600
gapH2nh2sh	5.6648	0.009300
ch2ge	2.1807	0.000000
dtfs	6.3384	0.006700
ge2h2	1.4620	0.005100
ge2h4	4.0465	0.355700
ge2h6	7.5755	0.000000
ge2me6	6.4453	0.000000
ge2o6h4	7.0597	0.015200
ge2oh6	7.1705	0.013800
ge37	3.4720	0.171800
ge38	3.3432	0.007500
ge3h8	7.1052	0.118800
gebr2h2	6.6117	0.054200
gec2h4cooh_oh_3	6.1371	0.000100
gecl2me2	7.7678	0.196600
gecl3h	7.4735	0.000000
gecl3me	7.5267	0.000000
gecl4	7.0971	0.000000
gef3ch3	9.0779	0.001100
gef3_onme2	6.3138	0.006700
geh2	2.5573	0.015700
geh2c2h4_3ring	5.7908	0.000000
geh2c4h6_ring	5.0996	0.069900
geh2ch2	5.2558	0.299600

A2. Supporting Information to Chapter 4

geh2meet	7.5119	0.016400
geh3br	6.6789	0.026400
geh4	9.3196	0.214200
geh4h2o	7.5079	0.085500
geme4	7.5657	0.109300
geo	4.8545	0.035600
geo2	4.3893	0.000000
gepsclh	6.0171	0.028200
germanium_adamantan	6.1860	0.000000
germanium_heptene	6.3533	0.115300
germin	4.2080	0.044100
sichdien	5.0978	0.069900
sih42	9.2121	0.000000
silacarben	4.0968	0.206700
sioh4	7.6036	0.033200
germin.2	4.6873	0.226000
sih42.2	9.2410	0.370600
c2h4hg	6.5208	0.180500
hg2cl2	5.6439	0.001100
hg2cl2.2	5.8223	0.830600
hg2f2	6.2242	0.035600
hg4	5.7586	0.000000
hg4s4	1.8313	0.000000
hg8	5.1440	0.000000
hga	4.9018	0.000000
hga.2	5.6211	1.113200
hgb	4.5729	0.043600
hgbr2	4.8502	0.000000
hgc	4.7730	0.005400
hgcl2	5.4874	0.000000
hgclme	6.5961	0.020200
hgcp2	4.4761	0.007600
hgd	5.7010	0.000300
hge	4.8169	0.009400
hgf2	6.3321	0.000000
hghi	5.4295	0.020900
hgme2	5.6826	0.000000
hgme2nh32	5.5878	0.140900
hgs	0.7851	0.001400
hg_scn_2	4.4479	0.002500
hg_scn_2.2	5.5377	0.361200
al2cl2h4	6.6321	0.000000
al2cl6	6.2163	0.000000
al2h6	6.6753	0.000000
al2me4	6.2859	0.000000
al2o4	2.6118	0.000000
alcl3nh3	4.3164	0.008500
alh3nh3	6.3232	0.017000
alh3ph3	6.5124	0.000400
aloh3	6.5124	0.000400
C2H5OInC2H5_2	4.7397	0.001300
ch3alh2	4.9153	0.000600
HOInCl2	5.7228	0.011900
HOInMe2	5.4817	0.000000
InCl	4.6916	0.147200
InH3	5.2832	0.000000
InI3	4.4977	0.000000
InMe	2.9247	0.061100
InMe3	4.9706	0.000000
InOH	4.5146	0.161800
In_OH_3	5.5960	0.000100



## A2.2. Parameters

In_OTf_3	7.1678	0.004100
O=InCl	3.2119	0.030700
dtfs	2.6107	0.096200
geh4h2o	7.3027	0.266900
pbbrsh	3.5234	0.007900
pbcl2	4.1230	0.017700
pbcl2ph2	4.9853	0.001400
pbcl4	4.5743	0.000000
pbcloh	4.3926	0.021700
pb_coome_2	3.3192	0.036200
pbet3h	6.1581	0.075400
pbet4	5.9528	0.078600
pbf2	5.1026	0.073600
pbh2	2.7112	0.009300
pbme2i2	4.3215	0.060300
pbme3cl	6.3098	0.193000
pb_oac_2	5.5700	0.039800
pb_ocome_2	5.5640	0.039800
pb_oh_2	4.8616	0.000000
pbpH2nh2	3.1090	0.012200
pbs	3.1630	0.000000
pb_sme_2	3.1142	0.028700
ring	3.8085	0.015200
si2h4	2.9972	0.251200
si2me6	5.3690	0.943600
si2oh6	4.7236	0.032800
si3h8	5.0282	1.119300
sichdien	5.0629	0.046600
sif4	6.1596	0.000000
sih4	7.8257	0.244400
sih42	7.7220	0.000000
siladamantan	3.1763	0.000000
silacarben	2.9482	0.210800
siof2	2.9313	0.018400
sn2o6h4	3.6537	0.000100
ch2s	1.3968	0.000000
ch2s.2	4.9480	0.148800
ch3sh	3.7876	0.001100
disen	3.7712	0.000000
dmsso	3.9662	0.024800
h2so4	3.1241	0.000200
hcssh	1.3714	0.000100
hcssh.2	2.8359	0.152900
po2me2	2.0231	0.005800
poc4h4	3.8549	0.000000
pocl2	2.5554	0.000000
pocl2_oh_2	5.2093	0.011200
pocl4	3.8634	0.007900
pocpo	2.1883	0.000000
pocpo.2	3.3924	0.832100
poh2	4.1739	0.000000
poh4	3.4327	0.000500
pome2	3.6121	0.000000
pome3i	3.9907	0.007300
pome4	3.5099	0.002800
po_nh2_4	4.6628	0.002100
pooh4	5.3735	0.008900
popH2no2hbr	2.5991	0.006300
sulfanil	3.4964	0.000100
target22	3.0356	0.003100
thiophen	3.8549	0.000000

A2. Supporting Information to Chapter 4

thiophen.2	4.4558	0.123900
asbetaine	5.5303	0.026300
biimidazol	3.4515	0.029100
bithiazol	4.1644	0.165700
bitriazin	2.4782	0.000000
ch2ph	4.3319	0.016500
glyphos	4.8467	0.086100
h3po4	4.7241	0.053000
h4sb2o7	4.5378	0.046700
p2h2me2	4.7533	0.285600
pbc2h8	5.4999	0.143200
ph3o	4.7438	0.045100
PhSbO	1.9482	0.003500
pnh2me2	4.9814	0.016100
ppyr	4.9577	0.010900
sb4o6	5.3379	0.000000
sb73	4.5576	0.217800
sb74	4.7814	0.220700
sbcl3	5.8112	0.084400
sbf3	7.3355	0.227000
sbf5	7.8668	0.000000
sbform	3.5590	0.006400
sbform.2	5.1360	0.158800
sbh3	6.4158	0.106100
sbh3bh3	7.1276	0.038800
sbh3s	4.3821	0.027900
sbme5	4.7433	0.000000
sbme5.2	5.3178	0.027300
sboh_o_2	3.8321	0.000000
sbpH2nh2sh	4.5866	0.049400
sb_ring1	4.8461	0.006100
sb_ring2	4.5005	0.012200
sb_triazin	2.4781	0.000000
C3H6Se_4ring	4.0861	0.000100
C3H6Se_kette	4.9625	0.000500
CCl2Se	2.3472	0.000000
CCl2Se.2	4.8464	0.193500
CH2Se	2.0707	0.000000
CH2Se.2	6.1634	0.207300
CH3SeH	5.0430	0.004500
CSe2	3.4744	0.000000
CSe2.2	5.3090	1.138800
disen	4.2811	0.000000
h2s2	4.3367	0.010300
H2Se	5.6178	0.000000
h2se6	5.6908	0.002700
h2so4	6.4887	0.005000
HcSe2H	2.9575	0.002100
HSeCN	5.3759	0.000400
me2s2	4.3551	0.010200
Me2SeO	5.3546	0.050400
methylthiirane	4.4214	0.000100
se6zn6	3.4682	0.000000
Se8	3.7762	0.000000
sea1	4.5116	0.032800
sea2	5.2138	0.524900
sea2.2	4.6548	0.006300
sea3	5.2454	0.441600
sea3.2	4.7052	0.011100
se_anthracen	4.6506	0.000000
SeCl4	3.7630	0.001800

## A2.2. Parameters

secl6	2.4577	0.000000
se_cystein	5.2592	0.004300
SeF4	6.6255	0.055900
SeF6	8.4736	0.000000
Selenazol1	5.3254	0.000200
Selenazol2	5.0984	0.125100
Selenophen	5.4053	0.124100
Selenophen.2	5.5986	0.000000
SeO2	3.4630	0.002700
sulfanil	5.0462	0.006200
sulfanilme2	5.0419	0.006300
ch2ge	1.7156	0.000000
dtfs	5.0928	0.001900
ge2h2	0.8934	0.003200
geh4h2o	7.5459	0.132200
H2Sn_OCMe_2	6.5924	0.002100
MeOOCsnBu3	5.2351	0.010300
O_SnEt3_2	5.7724	0.080200
ring	4.8074	0.014000
ring2	4.8762	0.014900
si2h4	3.3110	0.357200
si2me6	6.0147	0.121000
si2oh6	5.6921	0.019200
si3h8	6.2061	0.849800
sichdien	5.0373	0.067100
sih2c2h4	4.9595	0.000500
sih2ch2	5.2557	0.299600
sih42	8.5769	0.000000
sih42.2	8.6514	0.450500
silacarben	3.2353	0.214000
siof2	3.5126	0.027200
sn2o6h4	5.2934	0.002100
sn30	4.0581	0.000000
sn55	3.1005	0.027600
SnCl2	3.9973	0.042100
SnCl4	6.1706	0.000000
SnClOH	4.5122	0.066200
SnF2	5.0560	0.154400
SnF2_OH_2	6.6938	0.036300
SnH3Cl	6.8646	0.024700
SnH4	8.6873	0.238400
SnMe2_6	4.4407	0.000000
SnMe2Cl2	7.0173	0.243900
SnMe3I	5.7509	0.043300
sno2	3.5164	0.000000
SnPh2Cl2	6.2290	0.308700
SnPH2NH2O	3.6099	0.001800
S_SnMe3_2	5.5584	0.111200
C3H3NTe	4.2802	0.000200
C3H6Te	3.0596	0.000000
C3H6Te_kette	4.1147	0.000700
C4H4Te	4.7319	0.138700
CCl2Te	1.6662	0.000000
CCl2Te.2	4.0674	0.200800
ch2s	1.5738	0.000000
CH3TeH	4.0406	0.001500
CTe2	2.4845	0.000000
disen	3.9124	0.000000
dmsO	4.4503	0.036600
h2te6	4.4499	0.000200
hcssh	1.5769	0.000100

A2. Supporting Information to Chapter 4

HTeCN	4.1627	0.000100
me2s2	3.5563	0.009400
Me2TeO	4.4520	0.036700
methylthiirane	3.4418	0.000100
OTeMeOH_2	4.9157	0.015500
se6zn6	2.8234	0.000000
sulfanil	5.0431	0.005100
sulfanilme2	4.9414	0.058500
target22	3.3311	0.019700
Te4P2Me2	2.0761	0.000000
te_anthracen	3.8849	0.036500
TeBr2Ph2	4.6891	0.175400
tef6	8.0439	0.000000
TeH2	4.4308	0.000000
TeH2Cl2	5.7694	0.129000
TeH6	5.0462	0.000000
TeH6.2	7.3200	0.226000
TeMe2	3.8692	0.000000
TeMe2F2	6.4374	0.165300
TeMe3I	4.3999	0.116600
TeMe4	4.1539	0.000900
TeNH2OH	3.7901	0.002100
TeSHPH2	3.6980	0.001400
thiophen	4.2085	0.000000
a	3.3748	0.000000
al2cl2h4	6.4589	0.000000
al2h6	5.7215	0.000000
al2me4	5.5923	0.000000
al2o4	3.2514	0.000000
alcl3nh3	4.1883	0.008800
alh3nh3	6.0529	0.010800
alh3ph3	6.1531	0.000500
b	4.1624	0.001500
c	4.9684	0.440200
ch3alh2	4.6275	0.001000
f	4.8381	0.010700
O=C_OTI_2	4.6020	0.000000
TI2S	3.4708	0.000000
TICl	4.8387	0.053900
TICN	4.4731	0.046800
TICp	4.4416	0.000000
TICp.2	4.9790	0.365700
TIF	5.2486	0.111000
TIH	3.0073	0.013900
TIH3	5.0153	0.000000
TIH3.2	7.3735	0.340200
TIMe3	4.6078	0.000000
TINC	5.0451	0.045700
TIOEt	3.8946	0.034500
TIOH	4.2553	0.051900
TIPH2SHNH2	4.6789	0.044600
ag2	3.0869	0.416817
ag83d	2.5935	0.000000
agbursch2	4.3505	0.106105
agc2h	3.2776	0.002334
agch3	3.4750	0.149216
agcl	3.0125	0.009455
agcn	3.9143	0.060021
agcu	2.8946	0.131945
agesser9	4.9354	0.015962
agh	3.7233	0.098253

## A2.2. Parameters

agli	2.4857	0.271246
agoh	2.4669	0.009417
ag+triflat	3.1432	0.001883
azmomo	2.3531	0.003194
azuleneRuRu	3.3116	0.006349
cp2zrme2	4.1290	0.000594
cp2zrmecl	4.1781	0.005768
cpyich2sime3	3.9157	0.020487
dhh	2.8512	0.015248
fac-rhcl3_h2o_3	2.3805	0.000000
mobipy	0.4579	0.002045
moc6h62	2.0696	0.000000
moc6h6co3	3.3010	0.000000
moco5ch2	1.9125	0.000000
moco6	3.8918	0.000000
mof6	5.2473	0.000000
mo_g	1.5795	0.000105
mo_h	2.4002	0.001267
mo_i	3.3625	0.000000
mo_j	2.6184	0.000052
mo_k	5.0415	0.000984
mo_nh2_6	3.1971	0.015356
moo2cl2	4.5867	0.003247
nb2o7h4	6.2873	0.005151
nb2s3oh4	2.4348	0.000178
nbcl3	0.3466	0.000000
nbcl3shnh2	2.9126	0.000986
nbcl5	3.4430	0.000000
nbcl52	3.1859	0.000000
nbcl5opcl3	3.1883	0.000060
nbco5no	2.2375	0.005239
nbcpo4	3.1200	0.004718
nbf54	6.6097	0.000000
nbf5oet2	5.2371	0.000027
nbh5dmpe	3.3883	0.011597
nbname25	3.5123	0.001017
nbocl3	4.9451	0.000000
nb_oh_3	1.3303	0.000000
nbooh	6.8127	0.000175
nbph2nh2sh	0.5689	0.001436
pdacac2	3.0507	0.000000
pdBC	1.1120	0.000000
pdcl2	1.0725	0.000347
pdclmedicarben	4.2850	0.008636
pd_cn_2	2.6574	0.009547
pdco	2.9016	0.010327
pdco3	3.5746	0.000000
pdco4	4.8686	0.000000
pddimer	3.2726	0.000144
pddjukic2	1.0530	0.014412
pd_dppe_cl2	1.4057	0.029447
pdf2	0.9396	0.000481
pdh2	3.0959	0.032925
pdh2cl2	1.1521	0.000000
pdi2	1.0346	0.000464
pdme2	2.7290	0.024368
pdme2_pme3_2	4.4824	0.000000
pd_no3_2	2.5206	0.000000
pd_oh_4	1.3483	0.001674
pdph34	4.7232	0.000000
pd_scn_2	1.7081	0.000000

A2. Supporting Information to Chapter 4

rhbr_pme3_3	2.6621	0.010904
rhcl2hpy3	3.1986	0.010714
rhcl3_co_3	2.8667	0.312604
rhcl3me2co	1.5160	0.000011
rhcl_co_2_2	1.8640	0.000187
rhclco4	3.3768	0.000330
rhclet_pme3_2	3.1736	0.016246
rhcloh3ch2co	1.9404	0.007229
rhco2co8	3.6988	0.000000
rhcpethen2	3.5276	0.000057
rhjiwyak	2.5521	0.000018
rhmarek	2.7494	0.024578
rhmpy3	3.3280	0.005133
rhn33cl3	2.5262	0.001333
rhnoco3	2.9606	0.000000
rhph3hco	4.0468	0.027641
rhshnh2i	1.3295	0.001991
rhtm09	3.4671	0.001927
rhtm48	2.2108	0.000170
ru2co9	2.9504	0.000000
ru3co12	2.9774	0.001644
rucl2_pme3_2	1.8259	0.006992
rucl4	1.1460	0.000000
rucl4_h2o_2	0.3175	0.000073
ruco2no2	2.3593	0.000000
ruco3tmm	5.4210	0.000000
ruco5	4.2175	0.000000
ru_cod_cot	2.1906	0.001561
rucp2	3.9750	0.000000
rugrubbs1	1.6500	0.001108
ruheme	2.5720	0.000950
ruo2	1.5093	0.002566
tc_a	1.8037	0.000121
tc_d	0.4286	0.000552
tc_f	0.1226	0.000001
tc_g	1.4847	0.000054
tch7	2.4813	0.000777
tc_i	4.3140	0.000431
tc_m	2.8026	0.000707
tc_n	1.6449	0.012846
tripd	2.5917	0.000000
y2br6	4.6096	0.000686
y2cl6	5.3800	0.000148
y2i6	3.6531	0.001758
y2o4	2.7952	0.000900
y2ome6	5.1656	0.000504
y4s6	1.5379	0.000000
yacetat3	6.2151	0.000210
ycl3	4.9739	0.000000
yf3	6.7093	0.003471
yh3	3.4218	0.000018
yhmecl	3.3125	0.000795
yme3	3.2769	0.006242
y_oh_3	5.2277	0.000000
yoxalat3	5.8477	0.000147
yph2nh2sh	2.9919	0.010150
zr2_cl2cp2ome2	4.3647	0.004913
zr3o6	2.9423	0.000132
zr3s6	1.8538	0.000115
zrbr4	3.7293	0.000000
zrcl4	5.5844	0.000000

## A2.2. Parameters

zrcp2cl2	3.9228	0.007350
zr_dipy_3	0.9765	0.101093
zrkerallen	1.7244	0.000219
zrh2cl2	4.1096	0.000022
zrh2me2	4.1974	0.000046
zrh4	4.0330	0.000000
zrmecl3	4.5533	0.000156
zro2	0.5702	0.000038
zro2n	2.5392	0.003916
zr_oh_4	7.2993	0.000004
zrph2nh2shbr	3.3532	0.018445
zrschartz	4.1064	0.001877
au2	2.8766	0.000000
au2br6	1.9744	0.000332
au2cl6	2.3993	0.000000
au83d	2.4823	0.000000
au8p	2.2586	0.000000
auakzeptor1	5.1293	0.000000
aubursch2	4.6705	0.000000
auc2h	3.1873	0.002135
auch3	4.1764	0.000063
aucl	2.6499	0.007108
aucl3h2o	2.4051	0.000068
auclme2	3.5011	0.000004
aucn	3.6091	0.049437
auh	3.6902	0.070004
auli	3.1885	0.148899
auph2shoh	2.2656	0.020691
azuleneOsOs	3.0297	0.003740
cisplatin	3.3129	0.000071
cp2hfcl2	4.2532	0.008390
cp2hfcl	4.3809	0.002314
cp2hfme2	4.4265	0.000186
cp2hfmecl	4.4928	0.006911
cp2hfse2me2	3.5381	0.070244
cp2os	4.4820	0.000000
cpau	0.7557	0.000748
erkerallen	1.6534	0.000279
fac-ircl3_h2o_3	2.7930	0.000001
hf3o6	2.3438	0.000001
hf3s6	1.7188	0.000009
hfbr4	4.2208	0.000000
hfcl4	6.1027	0.000000
hfdipy3	0.9545	0.111613
hff4	9.6693	0.000000
hfh2cl2	4.4760	0.001063
hfh2me2	4.5832	0.000016
hfh4	4.4141	0.000000
hfmecl3	5.1478	0.000569
hfo2	1.7300	0.000396
hfo2n	1.9031	0.002270
hf_oh_4	7.2747	0.000282
hfph2nh2shbr	3.8159	0.024224
ir2co8	3.4521	0.000000
irbr_pme3_3	3.1600	0.019669
irburger	1.6330	0.001069
irbursch1	3.4624	0.000011
ircl2hpy3	2.9489	0.023349
ircl3_co_3	3.4655	0.003394
ircl3me2co	0.8646	0.000380
ircl3_osme2_3	2.5549	0.000065

A2. Supporting Information to Chapter 4

ircl_co_2_pph3_2	3.2288	0.040391
irclco4	3.7493	0.000179
irclet_pme3_2	3.1947	0.018387
irclph3ch2co	2.1952	0.010954
irf5	0.8341	0.000000
irh3	3.1489	0.000000
irjiwyak	2.9063	0.000039
irmppy3	3.0148	0.009264
irnoco3	3.2686	0.000000
irtm09	3.1398	0.001521
irtm48	2.7053	0.000382
irtm62	2.6752	0.003335
la2br6	4.5119	0.000000
la2cl6	5.5585	0.001028
la2i6	3.6304	0.001752
la2o4	2.8321	0.000085
la2ome6	4.8272	0.000000
laacetat3	5.6753	0.000137
lacl3	4.8407	0.004937
lACP3	3.2271	0.000000
laf3	6.5062	0.001934
lah3	3.3083	0.000050
lahmecl	3.2568	0.000448
lame3	3.2841	0.005113
laoh3	5.1633	0.000860
laoxalat3	5.8080	0.000176
laph2nh2sh	2.7224	0.004799
o2oso2c2me2	2.0867	0.000390
os2co9	2.6876	0.000000
os3co12	3.4858	0.057751
osbr4	1.0551	0.000000
oscl2n2_pph3_2	2.3542	0.000031
oscl3_nhcn_bpy	1.1069	0.000351
oscl4	1.1692	0.000000
osco2no2	2.2763	0.000000
osco3tmm	5.3388	0.000000
osco5	4.0364	0.000000
oscp2	4.5509	0.000000
oso4	1.7534	0.000262
os_si_komplex	1.4472	0.015247
pme3auc2me	4.6332	0.000000
ptacac2	3.0518	0.000000
ptcl2	0.9287	0.000327
ptclmedicarben	4.2456	0.008466
ptco	2.5843	0.000000
ptco3	3.4410	0.000000
ptco4	5.2252	0.000000
ptdimer	3.0468	0.000079
ptf2	0.5807	0.000396
pth2	2.7417	0.019024
pth2cl2	2.9675	0.000000
ptkomplex1	2.7844	0.000009
ptkomplex4	2.9638	0.000006
ptme2	2.6717	0.013819
ptoxalip	3.9603	0.000016
ptoxalip2	4.1364	0.000045
pt_pf3_4	5.8024	0.000252
ptph34	4.7986	0.000000
ptsi-komplex	3.4867	0.008727
rea	1.8724	0.000084
reco3no	1.6163	0.014077



## A2.2. Parameters

reco5cl	3.4937	0.012227
recpco3	4.4938	0.000001
red	0.4669	0.000888
ref	3.0398	0.000002
reh	1.6224	0.000025
rei	4.9488	0.000700
rej	4.7829	0.000059
rek	4.3526	0.000000
rel	3.3788	0.004471
ren	1.7173	0.000174
reo3f	4.8728	0.000391
ta2o7h4	6.7125	0.005823
ta2s3oh4	2.1552	0.000013
tacl3shnh2	3.8717	0.003423
tacl5	4.0430	0.000000
tacl5_2	3.7929	0.000000
tacl5opcl3	3.7980	0.000077
tacl5sme2	1.9754	0.000017
taco5no	2.1295	0.005217
tacpco4	3.0485	0.002586
taf5_4	7.4523	0.000003
taf5oet2	6.2501	0.000029
tah5dmpe	3.9751	0.001796
ta_nme2_5	3.9607	0.001279
taocl3	5.3699	0.000000
taooh	7.2087	0.000000
taph2nh2shme2	2.9534	0.004328
vaskas	3.4207	0.030482
w_a	3.5102	0.001032
wc6h6co3	3.1432	0.000000
wco5ch2	1.8064	0.000000
wco6	3.6835	0.000000
w_d	2.5865	0.000000
wdundul	2.2671	0.003280
w_e	6.2854	0.000000
w_g	2.9917	0.000000
w_h	3.6752	0.000000
w_i	2.1183	0.002064
w_j	2.4876	0.018809
w_k	3.5364	0.000047
w_l	0.4289	0.000295
w_m	4.8474	0.003656
w_n	4.7347	0.002032
wvelwis	2.9661	0.010494

## A3. Supporting Information to Chapter 6

Appendix A3 contains:

- Supplementary figures
- Additional discussion

### A3.1. TDHF computed first hyperpolarizabilities

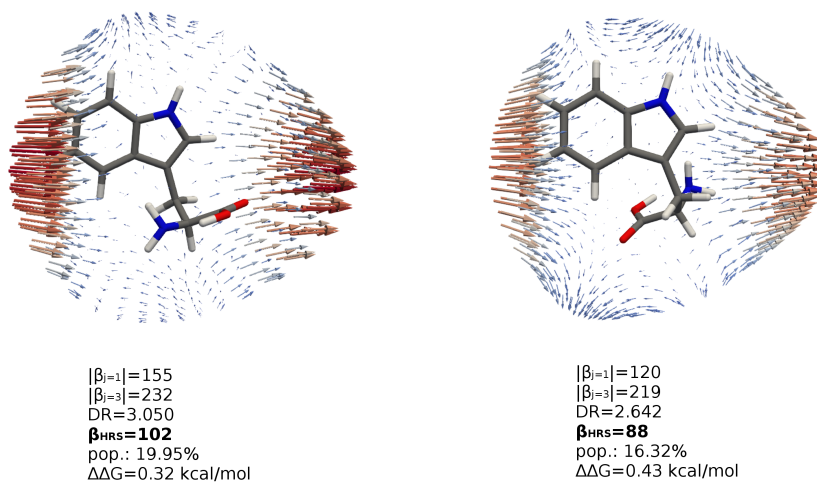
Table A3.1.:  $\beta_{\text{HRS}}$  values for tryptophan (W) and KWK, computed with TDHF/aug-cc-pVDZ. Results are in a.u. while wavelengths are in nm.

$\lambda$	W	W/PCM(water)	KWK	KWK/PCM(water)
$\infty$	51	196	188	173
1900	53	123	194	91
1500	55	128	198	94
1064	58	140	210	102
929	60	150	218	108
794	64	166	232	119
713	68	184	247	132
632	75	216	272	157

Table A3.2.:  $\beta_{\text{HRS}}$  values for tryptophan (W) and KWK, computed with TDHF/6-31+G(d).

$\lambda$	W	W/PCM(water)	KWK	KWK/PCM(water)
$\infty$	56	200	193	173
1900	58	129	199	94
1500	59	123	203	97
1064	62	144	215	106
929	64	152	223	113
794	67	166	238	124
713	70	181	253	138
632	76	208	278	162

### A3.2. Unit-sphere representation

Figure A3.1.: Unit-sphere representation for tryptophan conformers **2** and **4**. The same scaling as for Figure A3.2 is applied.

A3. Supporting Information to Chapter 6

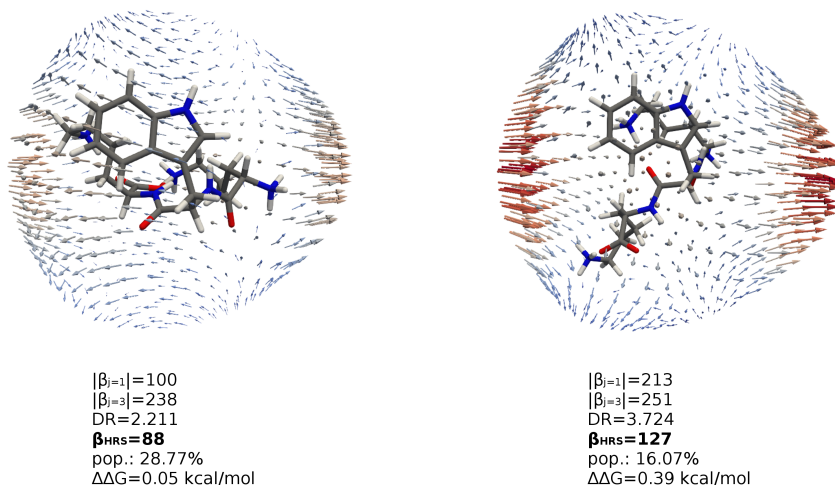


Figure A3.2.: Unit-sphere representation for KWK conformers **2** and **4**. The same scaling as for Figure A3.1 is applied.

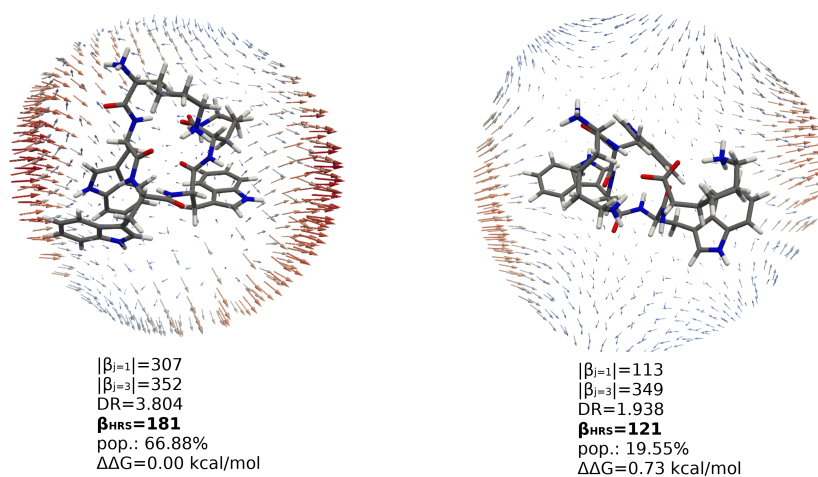


Figure A3.3.: Unit-sphere representation for KWWWK conformers **1** and **2**.

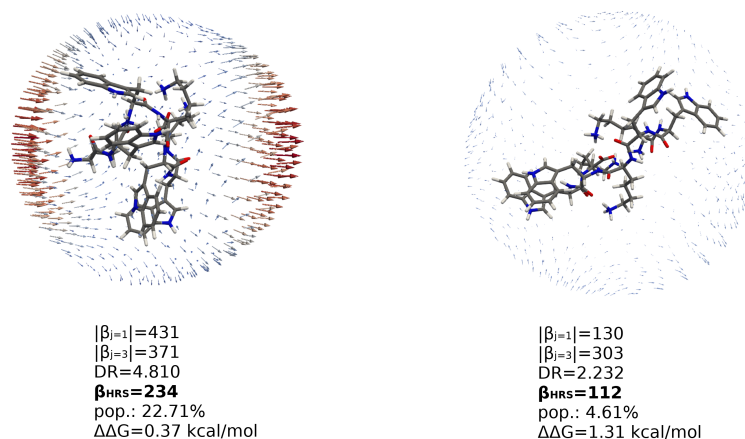


Figure A3.4.: Unit-sphere representation for KWWKWWK conformers **3** and **4**.

### A3.3. Molecular dipole and $\beta$ vectors

In order to test the hypothesis that the alignment of indole moieties is responsible for the significant change of the first hyperpolarizability, model systems including only indole units were investigated. For educational purpose, Figure A3.5 displays the dipole and  $\beta$  vectors of a single indole unit. Figure A3.6 shows for model systems of KWWK (resulting from cuts of the conformers 2 and 5 and saturation with hydrogen atoms), the dipole and the  $\beta$  vectors for the monomers as well as for the dimer. In the case of the  $\beta$  vector, the sum of the two monomer  $\beta$  vectors is plotted as well.

Figure A3.6 shows the additive behavior of the molecular dipole vector has a purely additive behavior. This way, two antiparallel indole units will cancel each others local dipole moment. More interestingly and non-trivially is that we observed a similar (but not pure) additivity for the  $\beta$  vector. In conformer 2, an enhancement of the total  $\beta$  vector is observed due to additive contributions from parallel indole units. Thus, the response of conformer 2 is larger than the one of 5 (having antiparallel indole units that almost cancel each other). These model systems give a very good insight in the structure-property relationship for these specific chromophores.

### A3. Supporting Information to Chapter 6

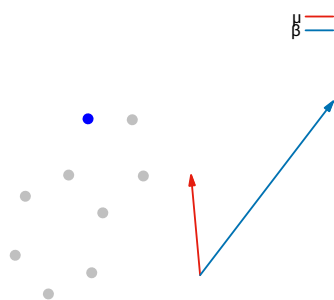


Figure A3.5.: Dipole and  $\beta$  vector of indole.

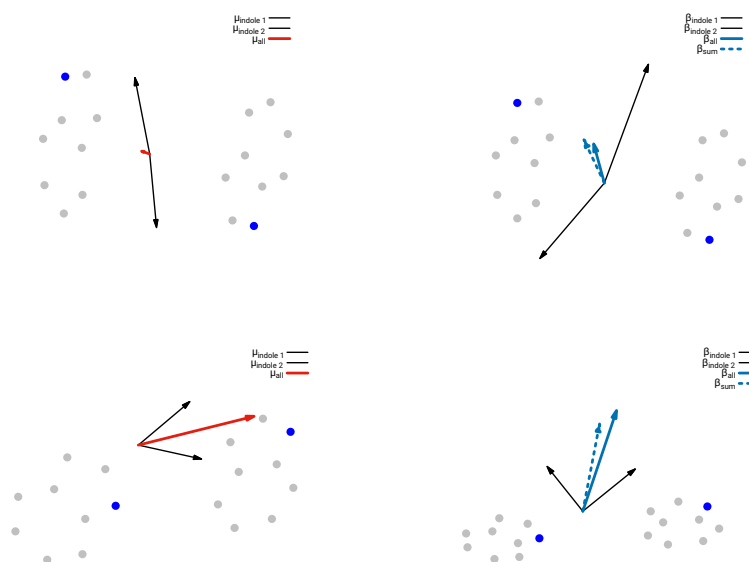


Figure A3.6.: Dipole (left) and  $\beta$  (right) vector for model systems based on conformer 2 (top) and 5 (bottom) of KWWK.

Considering KWWKWWK (Figure A3.7), conformer 3 includes three aligned indole units that enhance the first hyperpolarizability response. Conformer 4 presents an almost complete cancellation of the local vectors resulting in a low static first hyperpolarizability value. The vector sum of conformer 4 (KWWKWWK) is barely visible. The deviation of the sum of all monomer vectors with respect to the computed vector for the quadruplex is larger than for KWWK. Since the  $\beta$  vector is a condensed representation of a rank three tensor, a full additivity could not be expected. However, to a first approximation the orientation of the indole units can still be used to analyze the first hyperpolarizability.

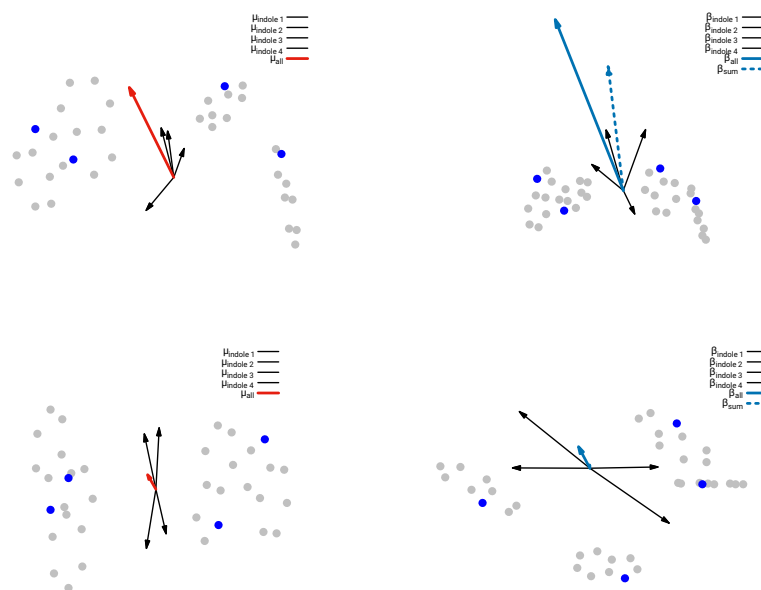


Figure A3.7.: Dipole (left) and  $\beta$  (right) vector of model systems based on conformer 3 (top) and 4 (bottom) of KWWKWWK.

### A3.4. Molecular dynamics sampling

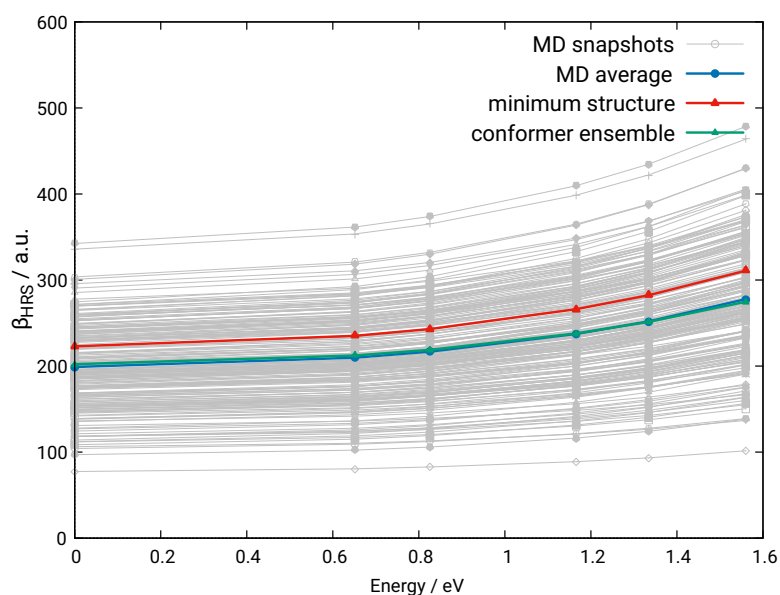


Figure A3.8.: Frequency dispersion computed with sTD-DFT-xTB for 200 snapshots (grey), the MD average (blue), the Boltzman weighted conformer ensemble (green), and for the optimized minimum structure (red) for the KWWKWWK peptide.

### A3.5. Comparison of sTD-DFT-xTB with TDHF

A comparison of first hyperpolarizabilities computed with sTD-DFT-xTB/GBSA and TDHF/IEF-PCM is provided. The experimental trend arranged by increasing SHG response is: KWK, KWWK, W, KWWWK, gramicidin A, KWWKWWK (see Table 1). Due to the computational cost of the TDHF method, it is not feasible to undergo the same number of calculations as at the sTD-DFT-xTB level. Therefore, we performed TDHF reference calculations for the lowest energy conformers. The frequency dispersion for all systems computed with TDHF/IEF-PCM and sTD-DFT-xTB/GBSA is depicted in Figure A3.9.

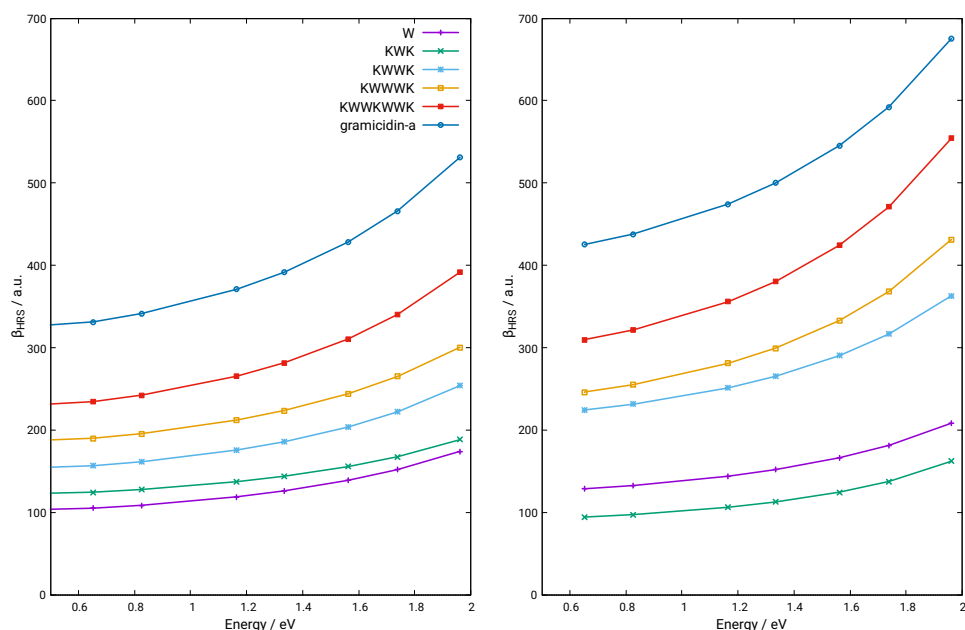


Figure A3.9.: Frequency dispersion computed with sTD-DFT-xTB/GBSA (left) and with TDHF/IEF-PCM/6-31+G(d) (right).

The qualitative trend from experiment for the model peptides are more or less well reproduced at both levels of theory. The TDHF/IEF-PCM level of theory qualitatively reproduces the difference between tryptophan and KWK. The sTD-DFT-xTB/GBSA method cannot describe this difference because of missing non-equilibrium solvent effects. In contrast, the experimental ordering of gramicidin A with respect to the other compounds is not reproduced correctly using both methods.



### A3.6. Quantitative comparison with experimental values

systems	TDHF	TDHF/PCM	xTB/GBSA
W	49 min	55 min	< 1s
KWK	2d 1h 28min	2d 2h 45min	4s
KWWK	6d 16h 4min	5d 20h 21min	6s
KWWWK	14d 19h 0min	14d 16h 7min	9s
KWWKWWK	28d 21h 52min	29d 11h 43min	15s

Table A3.3.: Timings for first hyperpolarizability calculations at TDHF/6-31+G(d), TDHF/IEF-PCM/6-31+G(d) and sTD-DFT-xTB/GBSA level of theory (wall time 8 thread processor).

### A3.6. Quantitative comparison with experimental values

Table A3.4.: Static first hyperpolarizabilities for tryptophan, KWK, KWWK, KWWWK, KWWKWWK, gramicidin A computed with TDHF/6-31+G(d), TDHF/6-31+G(d)/PCM, sTD-DFT-xTB, sTD-DFT-xTB/GBSA and Boltzmann weighted ensemble with sTD-DFT-xTB/GBSA.

system	exp	TDHF	TDHF/PCM	xTB	xTB/GBSA	xTB(confs)
W	240	56	128	85	100	97
KWK	44	193	94	229	119	113
KWWK	175	212	224	304	149	154
KWWWK	365	201	246	247	181	167
KWWKWWK	536	281	309	347	222	202
gramicidin A	384	249	425	314	316	

A3. Supporting Information to Chapter 6

Table A3.5.: Relative errors and MAE for static first hyperpolarizabilities for tryptophan, KWK, KWWK, KWWWK, KWWKWWK, gramicidin A computed with TDHF/6-31+G(d), TDHF/6-31+G(d)/PCM, sTD-DFT-xTB, sTD-DFT-xTB/GBSA and Boltzmann weighted ensemble with sTD-DFT-xTB/GBSA.

system	TDHF	TDHF/PCM	xTB	xTB/GBSA	xTB(confs)	<b>MAE</b>
W	-0.77	-0.47	-0.64	-0.59	-0.60	<b>0.61</b>
KWK	3.39	1.14	4.20	1.71	1.57	<b>2.40</b>
KWWK	0.21	0.28	0.74	-0.15	-0.12	<b>0.30</b>
KWWWK	-0.45	-0.33	-0.32	-0.50	-0.54	<b>0.43</b>
KWWKWWK	-0.48	-0.42	-0.35	-0.59	-0.62	<b>0.49</b>
gramicidin A	-0.35	0.11	-0.18	-0.18		<b>0.20</b>
<b>MAE</b>	<b>0.94</b>	<b>0.46</b>	<b>1.07</b>	<b>0.62</b>	<b>0.57</b>	

## D. Acronyms

### Acronyms

**BJ** Becke–Johnson

**BO** Born–Oppenheimer

**BSE** bovine spongiform encephalopathy

**CAMM** cumulative atomic multipole moments

**CIS** configuration interaction singles

**DFA** density functional approximation

**DFT** density functional theory

**DNA** deoxyribonucleic acid

**CD** circular dichroism

**FF** force field

**GFN** geometries, frequencies, and non-covalent interactions

**GGA** generalized gradient approximation

**HF** Hartree–Fock

**KS** Kohn–Sham

**LCAO** linear combination of atomic orbitals

**LSDA** local spin density approximation

**MD** molecular dynamics

**MO** molecular orbital

**MTD** meta-dynamics

**NLO** nonlinear optical

**NMR** nuclear magnetic resonance

**PES** potential energy surface

**QM** quantum mechanical

**RMSD** root-mean-square deviation

**SCF** self-consistent field

**SE** Schrödinger equation

**SHG** second harmonic generation

**SHIM** second harmonic imaging technique

**SQM** semiempirical quantum mechanics

**sTDA** simplified Tamm–Dancoff approximation

**TD** time-dependent

**TDA** Tamm–Dancoff approximation

**TFD** Thomas–Fermi–Dirac

**UEG** uniform electron gas

**UV** ultraviolet

**UV-Vis** ultraviolet and visible

**WF** wave function

**WFT** wave function theory

**xTB** extended tight-binding