UNIVERSITY OF CALIFORNIA
RIVERSIDE

Harnessing Large-Scale Quantum Calculations for Predicting Material and Chemical Properties

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Materials Science and Engineering

by

Anshuman Kumar

March 2023

Dissertation Committee:

Dr. Bryan M. Wong, Chairperson
Dr. P. Alex Greaney
Dr. Luat T. Vuong

The Dissertation of Anshuman Kumar is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am deeply grateful to my advisor, Prof. Bryan Wong, whose guidance, unwavering support, and endless patience have been instrumental during my Ph.D. I owe much of my success to his consistent encouragement, mentorship, and invaluable feedback, which have played a pivotal role in shaping my research projects and improving my writing skills. I would also like to thank him for countless in-depth discussions about our research projects. His insights and feedback have been crucial in honing my skills as a researcher.

I would also like to express my gratitude to my committee members - Prof. Alex Greaney and Prof. Luat Vuong, for their constant encouragement, valuable feedback, and support. I am very grateful to Prof. Alexander Korotkov and Prof. Roger Lake at UC, Riverside, for introducing me to quantum mechanics and encouraging me to pursue research in this area. I am greatly indebted to past members of our lab - Dr. Sarah Allec, Dr. Niranjan Ilawe, Dr. Freddy Aquino, and Dr. Sharma S. R. K. C. Yamijala for helping me understand the intricacies of density functional theory (DFT) during my early graduate years. Special thanks to Dr. Sarah for her help and guidance during my initial years. I am also thankful to Dr. Freddy for helping me compile various DFT and density functional tight binding (DFTB) codes on high-performance computing (HPC) clusters. I would also like to acknowledge the helpful discussions with previous postdocs at our lab - Dr. Chao Lian, Dr. Ravindra Shinde, and Dr. Kota Hanasaki. These discussions were very insightful and helped me further my research.

Special thanks to my collaborators - Prof. Reinhard Schweitzer-Stenner at Drexel University, Palermo lab and Haddon's group at UC Riverside, who have shared their valu-

able knowledge and insights with me. Without them, I would not have been able to make the progress I have thus far. Special thanks to Dr. Zulfikhar Ali, Mr. Xian Wang, Mr. Akber Reza, Dr. Lihua Xu, and all my colleagues for our collaborations and discussions. I am very grateful to Dr. Zulfikhar and others for their encouraging words and motivation, which helped me overcome some challenging times during my research. I am also thankful to the current group members of our lab - Dr. Mahmut Okyay, Dr. Min Choi, Dr. Sohag Biswas, Dr. Qiang Xu, and Dr. José M. Rodríguez-Borbón, for sharing their knowledge and insights with me. Their constructive feedback often helped me improve my research. Thanks to Mr. Yuan Chen, Mrs. Hyuna Kwon, Mr. Steve Yang, and Mr. Simon Sandhofer for discussing their projects. These were often rewarding and enriching with knowledge transfer. I would also like to thank Mr. Yijing Sun for helping me with the MAGMA library on GPUs in my early years. I am also thankful to the university staff at UC Riverside, especially Andrew Larratt-Smith, for their assistance during graduate school. I am immensely thankful to my undergraduate mentors at Manipal Institute of Technology - Dr. M. Sathish Kumar, Dr. Dayananda Nayak, and Dr. Shounak De, for their faith in me.

I am grateful to Dr. Scott Macmillian from NVIDIA for helping me create HPC containers on Microsoft's Azure. I would also like to thank the reviewers of my work for their thoughtful insights and critiques. Their feedback has been instrumental in helping me to improve my research and writing. I am also grateful to the agencies - National Science Foundation (NSF), Department of Energy (DOE), and Office of Naval Research (ONR), for their funding and support. I am deeply thankful to High-Performance Computing Center (HPCC) at UC Riverside and the NSF's Extreme Science and Engineering Discovery En-

vironment (XSEDE) supercomputers at UC San Diego. Special thanks to Microsoft and NVIDIA Corporation for providing computing resources during my research.

I am very grateful to my parents, Dr. Devendra Nath Akela and Mrs. Aruna Singh, for their countless blessings and unrelenting support. Their consistent encouragement, immense patience, and faith shown throughout my Ph.D. journey have been a great source of inspiration for me and played a crucial role in my success. I will always be grateful to my wife, Tanya Singh, for her immense love, support, patience, and constant encouragement in helping me overcome challenges during my Ph.D. Special thanks to my uncle Mr. Abhay Kumar Singh, for encouraging me to strive for excellence. Finally, I would like to express my sincere gratitude to all my family and friends for their unwavering support and constant encouragement throughout my journey.

Acknowledgment of previously published materials: The text of this dissertation, in part or in full, is a reprint of the material as appeared in previously published papers [1–8] that I authored or co-authored.

DEDICATED TO

All my teachers.

My parents, Dr. D. N. Akela and Mrs. Aruna Singh, for all the blessings, encouragement, and support.

My grandparents for sharing their experiences and life lessons.

My wife, Tanya Singh, for her love and sacrifices.

My extended family and friends in India and abroad.

ABSTRACT OF THE DISSERTATION

Harnessing Large-Scale Quantum Calculations for Predicting Material and Chemical Properties

by

Anshuman Kumar

Doctor of Philosophy, Graduate Program in Materials Science and Engineering
University of California, Riverside, March 2023
Dr. Bryan M. Wong, Chairperson

Density functional theory (DFT) is a powerful method for probing chemical and material properties and guiding the design of novel materials from first principles. However, the computational demands of large-scale and/or long-time DFT calculations can be challenging. To address this limitation, this thesis employs the approximate DFT method, density functional tight-binding (DFTB), as an alternative approach to DFT. The first part of the thesis demonstrates the accuracy and reliability of DFT methods in describing the electronic structure and properties of chemical and material systems. In the second part of the thesis, we transition from DFT to analyze the accuracy and efficiency of DFTB in performing large-scale electronic structure calculations. To achieve this goal, we have interfaced DFTB with the cluster approach to statistical mechanics (CASM) program, which allows for the efficient calculation of formation energies and convex hull. Furthermore, we have extended the DFTB approach to perform long-timescale metadynamics calculations on biochemical systems with the help of GPUs. GPU-enabled DFTB allows for an efficient and accurate description of the free energy surfaces and provides valuable insight into the

transition pathways. In summary, this thesis aims to accelerate *ab initio* computations by enabling accurate and efficient prediction of material and chemical properties with DFTB. GPU-enhanced DFTB is a powerful tool for exploring the electronic and thermodynamic properties of complex materials, chemical, and biochemical systems, with potential applications in materials science, physics, and chemistry.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

In the past decade, computational materials science has made great strides in developing tools to predict the properties of chemical, material, and biological systems from first principles. Density functional theory (DFT) based methods are among the most accurate and efficient quantum simulation techniques for predicting electronic properties in chemistry, physics, and materials science. These capabilities have enabled researchers to understand the behavior of materials at the atomic level and to design materials with specific functionalities. For example, they can be used to understand the structure-property relationships of materials and identify new materials with desirable physical, chemical, or electrical properties [9, 10]. This has broad implications for developing sustainable energy sources, novel materials, and biomedical therapeutics [11–14]. DFT has been successfully applied to various systems to compute properties, ranging from the formation energies of alloys [4] and electronic transport properties of doped carbon nanotubes [6] to the optical

response of peptides [1, 2], and polarizabilities of conjugated molecules [7]. However, these DFT-based methods are computationally costly, and these calculations are computationally prohibitive on a large length or long-time scale. To overcome this computational cost associated with DFT, DFTB provide an alternate cost-effective, efficient, and accurate approach for modeling the material and chemical systems.

The purpose of this thesis is to (i) assess the accuracy and effectiveness of DFT methods in correctly describing the electronic structure of chemical and material systems, (ii) extend DFTB methodology in performing large-scale electronic structure calculations by interfacing it with cluster approach to statistical mechanics (CASM) software program, and (iii) enable performance enhancement of metadynamics calculations with GPU-accelerated DFTB approach. This thesis is divided into two parts. Part I (Chapters 3-6) focuses on the application of DFT in predicting the electrical transport properties of doped carbon nanotube [6], the role of the exchange-correlation functional in correctly capturing the electronic structure of chemical systems [7], and using time-dependent DFT to calculate the UV absorption and circular dichroism (CD) spectra to reflect conformational dependence of peptides at high temperature [1, 2]. The second part of the thesis (Chapters 7 and 8) consists of the density functional tight binding (DFTB) approach, which is an approximate DFT method but computationally much cheaper and, thus, more efficient than DFT in performing large-scale quantum calculations. Specifically, in chapter 7, we have leveraged DFTB with the cluster approach to statistical mechanics (CASM) program to calculate the formation energies and convex hull of binary compounds [4]. Using large-scale DFTB calculations, we showed that DFTB could be used as an alternative approach (to DFT) to

enable fast and accurate prediction of formation energies/convex hulls of binary alloys. In chapter 8, we used GPU-enhanced DFTB to perform long-time-scale (~10ns) metadynamics simulations of alanine dipeptide and remdesivir [5]. Our DFTB metadynamics approach showed two orders of magnitude performance enhancement compared to the DFT approach. Also, our GPU-accelerated DFTB calculations qualitatively agree with computationally intensive hybrid DFT benchmarks.

While this thesis covers a selection of my research work, all of my research works at UC Riverside can be found at the Refs. [1–7, 15–20].

1.2 Outline

The first part of the thesis, Chapters 3-6, describes the application of DFT methods to study material and chemical systems. The goal of the section is to introduce the reader to the DFT and time-dependent DFT methodologies, which enables predictive capability in simulating the chemical and electrical properties of systems at the nanoscale. Specifically, we study the transport properties of doped carbon nanotubes (CNT), the polarizability of a π-conjugated chemical system, and the circular dichroism spectra of important biological peptides. More details on the specific topics are as follows:

- Chapter 2 opens with an introduction to the density functional theory (DFT) and time-dependent DFT to provide readers with a foundational understanding of these theories and their relevance to the related work presented in the subsequent chapters.
- Chapter 3 includes the application of DFT describing the effect of doping carbon nanotubes (CNT) with Lithium and Chromium atoms.

- Chapter 4 comprehensively analyzes the linear polarizability and second-order hyperpolarizability of conjugated oligomers. We compare the accuracy of various empirically tuned and conventional range-separated functionals against CCSD(T) benchmarks. We find that tuned range-separated functional is more accurate, especially when the system can relax to a lower-energy broken-symmetry configuration.

- Chapters 5 and 6 investigate the UV absorption spectra and UV-CD spectra of a series of cationic peptides in a vacuum and aqueous solution using time-dependent DFT.

The second part of the thesis (Chapters 7 and 8) focuses more on extending the DFTB approach to probe the electronic properties of material and chemical systems efficiently and accurately.

- Chapter 7 leverages the DFTB approach for efficiently predicting the formation energies and convex hulls of SiC and ZnO alloys. By interfacing DFTB with CASM program, we showed that the DFTB gives accurate results and can be used as an efficient computational approach for calculating and pre-screening alloy formation energies/convex hulls.

- Chapter 8 uses the GPU-enhanced DFTB approach for efficient prediction of free energies of biochemical systems. Specifically, we showed that converged free energy surfaces of complex systems are accurately predicted at the quantum level using the DFTB approach.

Chapter 2

Density Functional Theory

This chapter summarizes the theories used extensively in this thesis. The content of this chapter, in part or in full, is taken from my published works at Refs. [4] and [8].

2.1 Kohn-Sham Density Functional Theory

In Kohn-Sham DFT (KS-DFT), the electronic energy of a system is given by

$$E[\rho] = T_{\mathrm{KS}}[\rho] + E_{\mathrm{ext}}[\rho] + E_{\mathrm{H}}[\rho] + E_{\mathrm{xc}}[\rho]. \tag{2.1}$$

All the terms in Eq. 2.1 are a functional of the electron density ρ. $T_{\mathrm{KS}}[\rho]$ in Eq. 2.1 is the kinetic energy of the electrons, $E_{\mathrm{ext}}[\rho]$ is the interaction between the electrons and the nuclei, $E_{\mathrm{H}}[\rho]$ is the Hartree energy which is classical electrostatic energy of the electronic density charge distribution, and $E_{\mathrm{xc}}[\rho]$ is the exchange-correlation energy which describes exchange and correlation quantum interactions. The electron density, $\rho(\vec{r})$, is expressed as

the sum of one-electron wave functions, ψ_i, known as molecular orbitals (MOs).

$$\rho(\vec{r}) = \sum_{i=1}^{N_e} |\psi_i(\vec{r})|^2. \tag{2.2}$$

To construct the electron density, these MOs are obtained by solving the Kohn-Sham (KS) eigenvalue equation:

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\boldsymbol{r}) + V_{\text{H}}[\rho](\boldsymbol{r}) + V_{\text{xc}}[\rho](\boldsymbol{r})\right]\psi_i(\boldsymbol{r}) = \epsilon_i\psi_i(\boldsymbol{r}). \tag{2.3}$$

$V_{\text{H}}[\rho]$ and $V_{\text{xc}}[\rho]$ in Eq. 2.3 are themselves functionals of the density and, thus, also of the MOs, ψ_i, which are obtained after solving the KS eigenvalue equation (Eq. 2.3). However, the KS eigenvalue problem in Eq. 2.3 cannot be solved directly and must be calculated in an iterative fashion within a self-consistent field (SCF) method. The iterative process starts with a set of guess wavefunctions, $\{\psi_i\}$, which are used to compute the terms in Eq. 2.3. This then allows the calculation of a new set of MOs$\{\psi_i\}$ by solving Eq. 2.3. The resulting electron density is then used to calculate the electronic energy in Eq . 2.1. The calculated electronic energy is compared to the previous iteration's electronic energy, and if the difference is below a determined convergence threshold, the iteration stops. Otherwise, the guess wavefunctions are updated, and the process is repeated in a self-consistent way until $\{\psi_i\}$ and the energy in Eq. 2.1 are converged. The self-consistent field method ensures that the electronic energy is minimized with respect to the electron density, and the KS eigenvalue problem provides a set of one-electron wave functions that satisfy this condition. The convergence of the iterative process is important to ensure that the electronic energy and the electron density are accurately calculated, which in turn provides reliable predictions of electronic structure and molecular properties.

One of the key computational bottlenecks in KS-DFT is the numerous three-dimensional integrals evaluated in each SCF step, which is extremely time-consuming. We will see later that density functional tight binding (DFTB) uses pre-computed integrals in each SCF cycle, and by doing so, its performance can be significantly faster (up to 2-3 orders of magnitude) than full DFT calculations.

2.2 Time-Dependent Density Functional Theory

As in ground-state density functional theory (DFT), TDDFT replaces the many-body Schrodinger equation with a set of single-particle equations whose orbitals yield the same density. However, in the case of TDDFT, one seeks the solution to the time-dependent Schrodinger equation:

$$i\frac{\partial}{\partial t}\psi_{el}(\vec{r};\vec{R}) = \hat{H}_{el}\psi_{el}(\vec{r},t;\vec{R}) \tag{2.4}$$

where $\vec{r}$ denotes the electronic coordinates, $\vec{R}$ denotes the nuclear coordinates, and t is time. The electronic Hamiltonian $\hat{H}_{el}$ is given by

$$\hat{H}_{el} = T_{el} + V_{ext} + V_{el-el} \tag{2.5}$$

where T_{el} is the kinetic energy of the electrons, V_{ext} is the external potential due to the nuclei and any other time-varying external potential (e.g., electric field), and V_{el-el} is the electron-electron interaction. The Kohn Sham (KS) Hamiltonian is given by,

$$\hat{H}_{KS} = T_{el} + V_{ext}(t) + V_H + V_{xc} \tag{2.6}$$

where V_H is the Hartree potential, and V_{xc} is the exchange-correlation potential, which together comprises the electron-electron interactions.

The corresponding time-dependent KS equations for a set of N orbitals are given by,

$$\hat{H}_{KS}\phi_i(t) = i\frac{\partial}{\partial t}\phi_i(t) \tag{2.7}$$

The set of orbitals $\phi_i(t)$ generates a time-dependent density that equals the density of the interacting system at all times. To study the linear response optical properties of the system, a small external perturbation is introduced ("small" implies that the ground-state structure is not completely destroyed). As electronic response properties such as excitation energies and oscillator strengths depend only on the density, these can be obtained exactly (in principle) within the KS formalism. Within the conventional linear-response approach, the full time-dependent KS equations are simplified to obtain excited-state properties in the frequency domain. Specifically, the excitation energy ω and corresponding excitation vectors $\mathbf{X}$ and $\mathbf{Y}$ are obtained by solving the following non-Hermitian eigenvalue equation, [21]

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}. \tag{2.8}$$

For a general hybrid exchange-correlation functional (discussed further below in this section), the elements of matrices $\mathbf{A}$ and $\mathbf{B}$ are:

$$A_{ai,bj} = \delta_{ab}\delta_{ij}(\epsilon_a - \epsilon_i) + K_{ai,bj}, \tag{2.9}$$

and

$$B_{ai,bj} = K_{ai,jb}, \tag{2.10}$$

where ϵ_p is the pth KS molecular orbital energy. The indices $i, j, \ldots$ and $a, b, \ldots$ correspond to occupied and virtual orbitals, respectively. $\mathbf{K}$ is the coupling matrix whose elements are

$$K_{ia\sigma,jb\tau} = \langle ia\sigma|jb\tau\rangle + \iint \phi^*_{i\sigma}(\vec{r}_1)\phi_{a\sigma}(\vec{r}_1)\frac{\partial^2 E_{XC}}{\partial\rho_\sigma(\vec{r}_1)\partial\rho_\tau(\vec{r}_1)}\phi_{j\tau}(\vec{r}_2)\phi^*_{b\tau}(\vec{r}_2)d^3\vec{r}_1 d^3\vec{r}_2 + K^{\mathrm{HF}}_{ia\sigma,jb\tau}. \tag{2.11}$$

The first term on the right side is the Hartree integral

$$\langle ia\sigma|jb\tau\rangle = \iint \phi^*_{i\sigma}(\vec{r}_1)\phi_{a\sigma}(\vec{r}_1)\frac{1}{\vec{r}_{12}}\phi_{j\tau}(\vec{r}_2)\phi^*_{b\tau}(\vec{r}_2)d^3\vec{r}_1 d^3\vec{r}_2, \tag{2.12}$$

and $K^{\mathrm{HF}}_{ia\sigma,jb\tau}$ is the mixed Hartree-Fock exchange integral, which is zero for pure functionals, and for hybrid functionals is given by

$$K^{\mathrm{HF}}_{ia\sigma,jb\tau} = c_x\delta_{\sigma\tau}\left\langle ja\sigma|ib\tau\right\rangle, \tag{2.13}$$

where c_x is a constant mixing rate. The oscillator strength of the I^{th} excited states is defined as,

$$f_I = \frac{2}{3}(E_I - E_0)\sum_{\nu=x,y,z}\left|\langle\psi_0|\,\vec{r_\nu}\,|\psi_I\rangle\right|^2, \tag{2.14}$$

where ψ_0 and ψ_I are the wavefunctions of the ground and the I^{th} excited states with energies E_0 and E_I, respectively. Within the linear-response TDDFT approach, this is implemented as follows:

$$f_I = \frac{2}{3}\sum_{\nu=x,y,z}\left[\sum_{ia\sigma} d^\nu_{ia\sigma}(\epsilon_{a\sigma} - \epsilon_{i\sigma})^{\frac{1}{2}}F^I_{ia\sigma}\right]^2, \tag{2.15}$$

where $d^\nu_{ia\sigma}$ is the transition dipole moment given by,

$$d^\nu_{ia\sigma} = \int \phi_{i\sigma(\vec{r})}\vec{r}_\nu\phi_{a\sigma}(\vec{r})d^3\vec{r}, \tag{2.16}$$

and $F^I_{ia\sigma}$ is the response function.

For a more comprehensive review of DFT and TD-DFT, readers are encouraged to review other literature [21–28].

Chapter 3

Covalent Atomic Bridges Enable Unidirectional Enhancement of Electronic Transport in Aligned Carbon Nanotubes

This chapter describes the effect of doping carbon nanotubes (CNT) with Lithium and Chromium atoms. To quantify the dopants' effect on transport properties of CNT, we performed *ab initio* calculations via density functional theory coupled to non-equilibrium Green's function approach. This work was performed in conjunction with our collaboration with experimentalists at Prof. Haddon's group at the University of California, Riverside. It was published as an article in the *ACS Applied Materials & Interfaces*, where the complete research is provided [6].

3.1 Introduction

Modern advances in understanding the chemical reactivity of carbon nanotubes and graphene have produced a myriad of functional structures and devices, all designed in a rational and predictive manner [29–35]. Chemistry-based approaches, in particular, have played a central role in improving the processing, bio- and chemical compatibility, separation, and alignment of carbon nanotubes, opening new opportunities in electronics, optoelectronics, and sensors. The design of 2D and 3D architectures often relies on enabling chemical techniques to interconnect carbon surfaces [36], which include e-beam welding [37], metal soldering [38], laser irradiation, and rapid Joule heating [39], all of which have been successfully used to create new bonds and connect carbon surfaces. Although the underlying mechanisms in these processes differ based on the approach, they all involve σ-bond formation, that is, conversion of carbon atoms from the conjugated surface lattice from sp^2 to sp^3 bonding. Thus, the introduced defects (sp^3 centers) ultimately compromise the electronic properties of the material.

A different approach to interconnect graphitic surfaces is to utilize transition-metal atoms with unfilled d orbitals that can efficiently overlap with the π-electrons of the benzenoid lattice and form a hexahapto (η^6) bond [40]. This type of covalent bond does not introduce sp^3 defects in the conjugated carbon network and, therefore, is an attractive approach for the design of electronic and optoelectronic materials with extended dimensionality.

A number of transition metals with unfilled d orbitals are capable of forming the η^6 bond to the benzenoid ring of the carbon nanotube walls. Thus, it has been shown

experimentally that transition metals from group 6 (Cr, Mo, W) [41] and the first row (Ti, V, Cr, Mn, Fe) [42] of the periodic table form η^6 bonds with SWNTs. Previous studies on the covalent attachment of transition-metal atoms to the sidewalls of carbon nanotubes revealed that Cr exerts the strongest effect on the SWNT film conductivity. This was assigned to the internanotube (van der Waals) distance in SWNTs (0.315 nm) [43], which closely matches the separation of 3.23 Å between the benzene rings in (η^6–C6H6)2Cr [44]. Prior work on classic organometallic compounds shows that the benzene–benzene distances are significantly larger in the case of other organometallic complexes including Mo and W [45, 46], resulting in a lower degree of functionalization of the carbon nanotubes.

In this study, we selected Cr as a classic 3d transition metal to form a hexahapto bond with the π-system of carbon nanotubes for electronically coupling aligned SWNTs. To further demonstrate that the organometallic hexahapto (η^6) bond covalently bridges the aligned SWNTs, we have conducted similar measurements with Li atoms deposited on the aligned SWNT films. The interaction of alkali metals with SWNTs has been extensively studied, and it is well understood that the transport mechanism involves electron transfer from the lithium atom to the nanotube, leading to the enhancement of the density of states near the Fermi level. We have selected Li as the reducing agent in our experiments because experimentally Li deposition by e-beam evaporation is easier to control compared to other alkali metals. This study demonstrates the effect of two types of chemical bonds on the electronic transport properties of aligned SWNTs: (i) covalent hexahapto bonding and (ii) ionic bonding (charge transfer, chemical doping). These two types of bonds induce conductivity enhancements in SWNT films that are governed by distinctly different mechanisms.

Chemical doping with alkali metals results in electron transfer to the conduction band of the SWNTs, and it has been shown to increase the conductivity of SWNT films by a factor of ~30 to 120 [47–50]. In contrast, bridging the benzenoid hexagonal surfaces of SWNTs with a hexahapto-metal bond improves the contacts between the individual SWNTs and can enhance the conductivity by a factor of 700 to 700 × 1000, depending on the SWNT film morphology (film thickness, SWNT electronic structure) [42, 51, 52].

To explore the unidirectional modulation of the electronic properties of SWNT films using covalent (organometallic) and ionic chemistry, the experimentalists have selected semiconductor-enriched arc discharge SWNT thin films as a model system in which the individual carbon nanotubes are aligned. As such, our theoretical model is also designed with semiconducting SWNTs. The layered structure of the aligned SWNTs produces films with electrical anisotropy—the resistivity values along the nanotube alignment direction are significantly lower than those measured normal to the tube axis [53, 54]. We examine the changes that hexahapto bonding of Cr and chemical doping with Li exert on the electrical transport properties of the aligned thin nanotube films in longitudinal and transverse directions. First-principles, non-equilibrium quantum mechanical calculations provide a further deeper understanding of the modifications in the electronic structure and mechanism of electron transport in perpendicular and parallel SWNT films upon the formation of ionic and covalent bonds with metal atoms. Our findings provide a rational path toward unidirectional modulation of the electrical transport in aligned carbon nanotube films and the design of crystalline carbon nanomaterials with novel electronic, magnetic, and possibly superconducting properties.

3.2 Theoretical Study and Methodology

All electronic structure calculations were performed with the FHI-aims electronic structure package [55] for comparing the electron transport between the intertube (perpendicular) and intratube (parallel) configurations of single-walled carbon nanotube (SWNT) networks. FHI-aims is an all-electron electronic structure code based on numeric atom-centered orbitals and provides an all-electron accuracy at a computational cost comparable to plane-wave/pseudopotential implementations. The numeric atom-centered orbitals, in conjunction with the grid localization used in the code, optimize the efficiency of the calculation so that most of the time-consuming computations scale linearly with time. We utilized the dispersion-corrected PBE-TS exchange-correlation functional [56, 57] in all of our calculations because van der Waals interactions play a significant role in SWNT junctions, and further details on the PBE-TS approach for carbon nanostructures are described in ref [57].

Before performing the transport calculations, we carried out geometry relaxations for a network of (14,0) SWNTs (arranged in a 2D array in the xy plane) until the largest force component on all the atoms was below 10^{-4} eV/Å. Periodic boundary conditions were used in all calculations, and a vacuum layer of ~100 Å was used in the z direction to prevent spurious interaction of the 2-D SWNT network with its periodic image. Because of the size of the large supercell used in our calculations, Fig. 3.1 we used a single Γ-point sampling scheme for all our calculations. We neglected the spin dependence in both the ground-state and transport calculations.

The electronic transmission functions were computed for the two-terminal SWNT network using the non-equilibrium Green's function (NEGF) approach [58]. Each SWNT was constructed from 12 unit cells, yielding a total of 678 atoms in the supercell for the Li/Cr-SWNTs and 672 atoms for the pristine SWNTs. The first and last four unit cells of the SWNTs are contained in Green's function transport calculations of the semi-infinite leads. To use preconstructed Green's functions, the leads must maintain their shape in the transport calculations, and therefore, relaxations of the lead atoms of the total system were not carried out. With the assumption that the boundary between the leads and the scattering region is non-reflecting, the Green's function of the device was calculated from the equation [59, 60]

$$\left[E - \hat{H}_0 - \sum_{\mathrm{i}} \sum_{\mathrm{i}}^{\mathrm{r}} (E)\right] G^r(\mathbf{r}, \mathbf{r}'; E) = \delta(\mathbf{r} - \mathbf{r}'), \tag{3.1}$$

where E is the energy, $\hat{H}_0$ is the Hamiltonian of the scattering region, and $\sum_{\mathrm{i}}^{\mathrm{r}}$ is the self-energy of the i^{th} lead.

The Green's functions G^r and other operators in Eq. 3.1 are represented by matrices in the basis of the numerical atomic orbitals. The solution of Eq. 3.1 for the Green's function requires knowledge of the Hamiltonian of the scattering region and the self-energy of the semi-infinite leads. The Hamiltonian is obtained via a DFT calculation for the periodic structure as described above. To compute the self-energies of the leads, a standard DFT calculation of the periodic infinite leads is performed, and Green's function for the semi-infinite leads is solved iteratively [61].

As described above, a ground state DFT calculation with periodic boundary conditions must be carried out for the scattering region and the semi-infinite leads to obtain the retarded Green's function G^r. To compute the transmission functions for the entire system, the energy levels of the ground-state solutions must be correctly aligned, and the carbon 1s energy eigenvalue of the atoms in the lead region is used as the energy reference. The transmission function for an SWNT junction between leads i and j is given by the equation [59, 60]:

$$T(E) = \int_{\partial\Omega_i}\int_{\partial\Omega_i}\int_{\partial\Omega_j}\int_{\partial\Omega_j} d\mathbf{r}_i d\mathbf{r}_i' d\mathbf{r}_j d\mathbf{r}_j' \Gamma_1(\mathbf{r}_i, \mathbf{r}_i') G^r(\mathbf{r}_i', \mathbf{r}_j) \Gamma_2(\mathbf{r}_j, \mathbf{r}_j') G^a(\mathbf{r}_j', \mathbf{r}_i) \quad (3.2)$$

In Eq. 3.2, retarded Green's function G^r is a solution to Eq. 3.1, and the advanced Green's function G^a is the Hermitian conjugate of G^r. The Γ functions are given by the relation $\Gamma_i = i\left(\sum_i^r - \sum_i^a\right)$, where $\sum_i^a$ is the Hermitian conjugate of $\sum_i^r$, and $\sum_i^a$ describes the coupling of the semi-infinite leads to the scattering region.

At zero temperature with no bias between the leads and the system, the conductance, G, of the nanotube junction in the Landauer formulation reduces to [58]

$$G = \frac{2e^2}{h} T(E_F), \quad (3.3)$$

where e is the elementary charge, h is Planck's constant, and $T(E_F)$ is the value of the transmission function at the Fermi energy, E_F. In this work, the transport calculations are performed with the system at zero bias with the same number of k-points as used in the periodic boundary DFT calculations. To make a connection to our experimental results, we calculate the linear response conductance (G) given by Eq. 3.3 as described above.

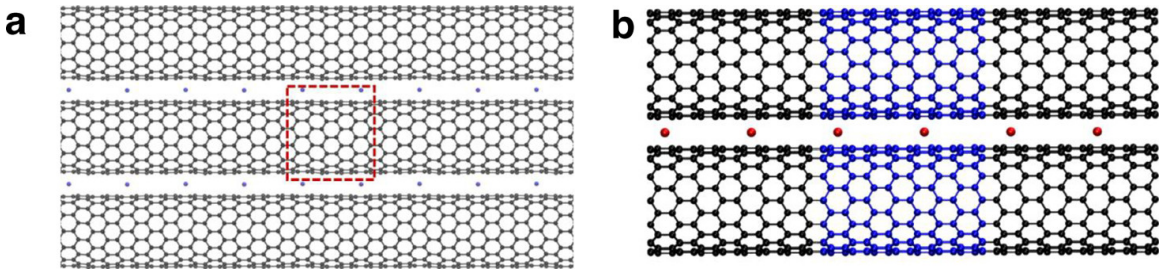


Figure 3.1: (a) Optimized (14,0) SWNT with two metal atoms (Cr/Li) per computational unit cell (denoted by the dotted red lines). The metal atoms are shown in blue. (b) (14,0) SWNT junctions with Cr/Li considered in this work. The blue atoms show the scattering region, and the black atoms represent the leads.

3.3 Results and Discussion

To understand our collaborators' experimental findings, we used theoretical non-equilibrium Green's function calculations to describe the changes in the electronic structure and properties of aligned Sc-SWNTs upon addition of chromium and lithium atoms. We specifically chose (14,0) semiconducting SWNTs for our theoretical study because these nanotubes possess a diameter of ~1.1 nm, which is close to the SWNT diameters used in our experiments. The computational unit cell used in the calculations is shown in Figure 3.1(a). The electronic transmission functions were computed for the two-terminal SWNT network using the non-equilibrium Green's function (NEGF) approach [59, 60]. This approach approximates the electron transmission by dividing the entire system into a central scattering region and semi-infinite leads. An array of SWNTs connected via metal atoms with two semi-infinite leads is illustrated in Figure 3.1(b).

The transmission function for intratube (parallel) and intertube (perpendicular) transport within the 2D SWNT network is shown in Figure 3.2. When metal atoms are added between the SWNTs, the transmission for both intratube and intertube transport

increases. In the case of intratube electron transport, the transmission is more enhanced with Li than for Cr, whereas for intertube transport, the opposite trend is observed, in agreement with the experimental results. Figure 3.2 (a and b) shows the intertube and intratube transport for the Cr- and Li-doped SWNTs. The large number of peaks near the Fermi energy in the Cr-SWNT directly corresponds to openings of new transmission channels.

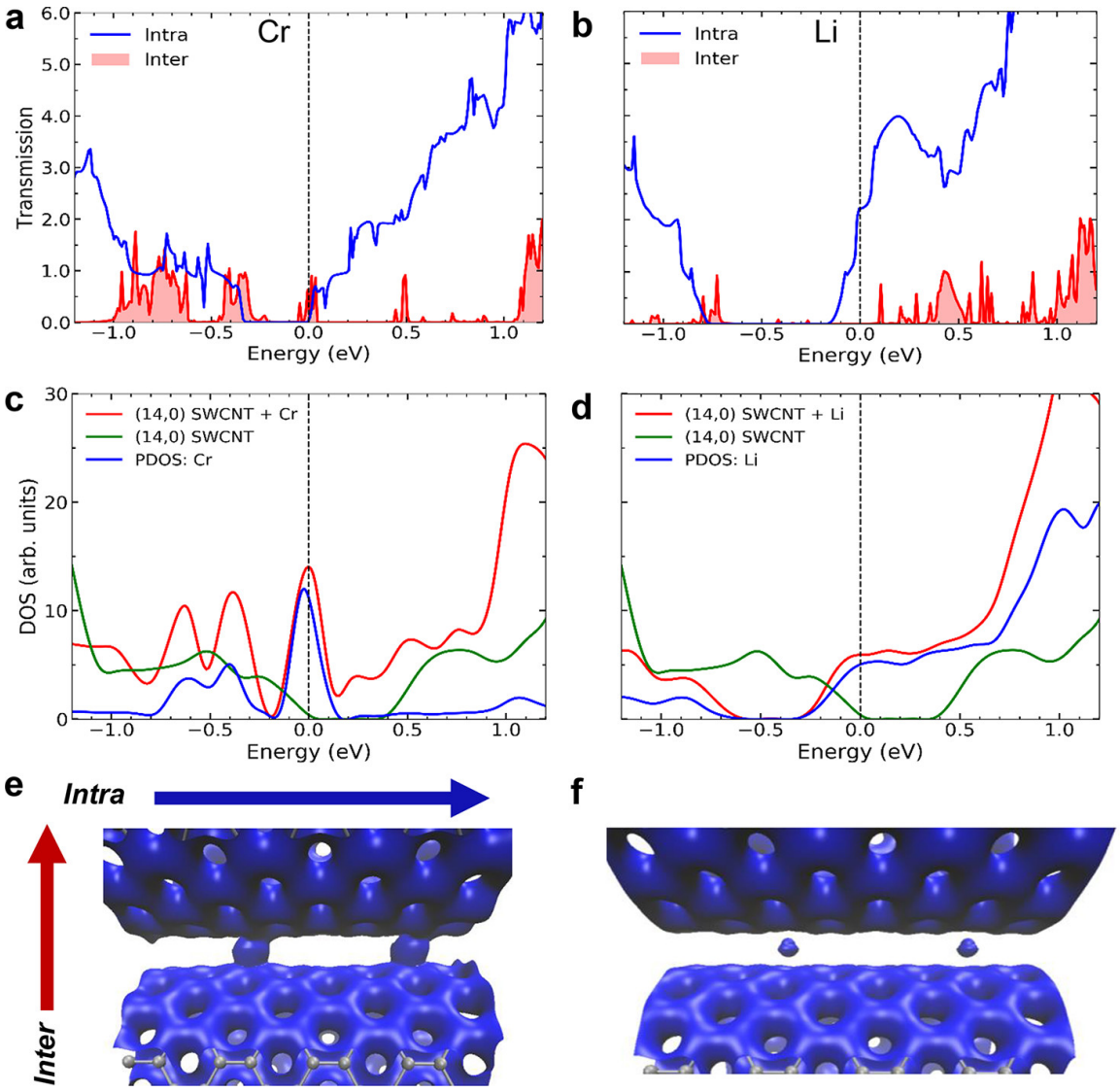


Figure 3.2: Intertube and intratube electronic transmission functions for (14,0) SWNT junctions with Cr and Li dopant atoms compared to pristine SWNTs. The zero energy corresponds to the Fermi energy. Electronic transmission functions for (14,0) SWNT junctions with (a) Cr atoms and (b) Li atoms. Electronic density of states for (14,0) (c) Cr-SWNTs and (d) Li-SWNTs along with the projected density of states (PDOS) for each dopant atom. (e, f) Total electron density (depicted by the blue isosurface) of (e) SWNTs + Cr and (f) SWNTs + Li. The carbon atoms are shown in silver, and the same isosurface value is used to depict the electron density for both systems. The arrows show the direction of inter and intratube transport.

These findings are consistent with the high density of states (DOS) near the Fermi level (cf. Fig. 3.2c) of the Cr-SWNT that further enhances the intertube electron tunneling between the SWNTs. Indeed, Fig. 3.2c shows that the total DOS near the Fermi level can be traced to localized electronic states from the Cr atom itself. These localized states hybridize with the C atoms on the SWNT and form conducting channels across the tube.

In this work, the transport calculations are performed with the system at zero bias with the same number of k-points as used in the periodic boundary DFT calculations. To make a connection to our experimental results, we calculate the linear response conductance (G) given by the equation [59, 60]:

$$G = \frac{2e^2}{h} \int \mathrm{d}E \; T(E) \left(-\frac{\partial f}{\partial E} \right), \tag{3.4}$$

The conductance for all configurations is summarized in Table 3.1. For intratube (parallel) transport, the conductance for the Li-doped SWNT is an order of magnitude larger than that of the Cr-SWNT configuration.

Orientation	SWNT + Cr	SWNT + Li
intratube transport	1.75×10^{-5}	1.52×10^{-4}
intertube transport	2.05×10^{-5}	4.97×10^{-7}

Table 3.1: Conductance (S) for the Intratube and Intertube Configurations

In stark contrast to the intratube case, the intertube (perpendicular) conductance for the Cr-SWNT is two orders of magnitude higher than that for the Li-doped system. These different trends in intratube versus intertube conductance are related to forming bonds between the Cr atoms and the SWNTs. Specifically, the hybridization of the orbitals

of the Cr atom and C atoms on the SWNT leads to the formation of conducting channels across the two SWNTs. This can be seen in Fig. 3.2e,f where the charge density around the Cr atoms significantly overlaps with the adjacent two SWNTs, thus enhancing intertube transport. In contrast, a smaller charge density is observed near the Li atoms, which hampers intertube transmission. We carried out a Hirshfeld charge analysis to further analyze this more quantitatively. We found that Li transfers a more significant amount of charge (0.175 e) to the SWNT compared to the Cr dopant atom (0.091 e), thus enhancing intratube transmission.

3.4 Conclusion

We demonstrate that hexahapto bonds and doping have a dramatically different effects on the unidirectional electrical transport in aligned Sc-SWNTs. The covalent η^6 bonds formed by the deposition of transition-metal atoms on aligned Sc-SWNTs increase the transverse (intertube) conductivity of the films, in contrast to the doping effect of alkali metals, which mainly affects the conductivity along the SWNTs. The distinctly different effects that e-beam deposition of chromium and lithium brings about in the conductivity of aligned SWNT films provide strong evidence for the formation of bis-hexahapto (η^6) bonds between the chromium atoms and the sidewalls of adjacent aligned SWNTs. Theoretical calculations reveal a high density of states (DOS) near the Fermi level of the SWNTs upon the addition of Cr atoms, which results in the opening of new transmission channels. The increased DOS near the Fermi level is due to localized electronic states from the Cr atoms, which hybridize with the carbon atoms of the SWNTs and form conducting channels across

the nanotube. These studies have important implications for applications of carbon nanotubes in electronics and optoelectronics and provide a controlled means to design ordered conjugated carbon structures with increased dimensionality of the electronic structure that is expected to exhibit unique electronic, magnetic, and possibly superconducting properties.

Our findings have implications in the field of anisotropic conduction, which is of great scientific and technological interest. Aligned SWNTs have been used as the building blocks of high-performance transistors [62–66], light-emitting diodes [67], logic gates [68], radio-frequency devices [69], transparent conductive films [70], sensors [71, 72] photodetectors and photoswitches [73], and synaptic transistors [74]. The ability to modulate the transport properties of aligned SWNTs in a controlled manner offers a route to expand the functionality of these technologies. In fact, the combination of the unique anisotropic properties of aligned SWNTs with well-defined arrays of metal nanoparticles can open new frontiers in nanophotonics, plasmonics, and spintronics. Ordered gold nanoparticle arrays coupled with aligned SWNTs have been successfully fabricated using standard e-beam lithography processes [75].

Furthermore, ordered carbon nanotube structures with one-dimensional trigonal channels are of interest as a carbon host for intercalated metal atoms [50]. The insertion of transition metals between pairs of aligned SWNTs would generate compounds analogous to graphite-intercalated compounds (GICs) in which the ionic bond characteristic of the GICs is replaced by covalent bis-hexahapto bonds [76]. Apart from the degree of charge transfer, the interplanar separation between the covalently complexed benzene rings of the SWNTs is much smaller in comparison to the distances reported in the charge transfer

(ionic) GICs [77], some of which are superconductors [78,79]. Our study demonstrates that the organometallic bond is able to function as an interconnect between graphene surfaces without inducing structural rehybridization or greatly perturbing their electronic structure. These findings have important implications for the design of ohmic contacts for carbon-based and organic electronics. Additionally, a variety of ligands can be used to stabilize the electronic properties of the organometallic functionalized carbon surfaces; thus, a recent study suggests the (η^6-graphene) $Cr(CO)_3$ complex as a robust quantum Hall resistance standard [80].

Chapter 4

Linear Polarizabilities and Second Hyperpolarizabilities of Streptocyanines: Results from Broken-Symmetry DFT and New CCSD(T) Benchmarks

This chapter's work resulted from collaborating with a former lab colleague, Lihua Xu. We contributed equally to performing the DFT calculation, analyzing the results, creating tables, and addressing the reviewers' comments. Lihua Xu wrote the original manuscript. I contributed to the review of the manuscript before publication. This is published as a full article in the *Journal of Computational Chemistry* [7].

4.1 Introduction

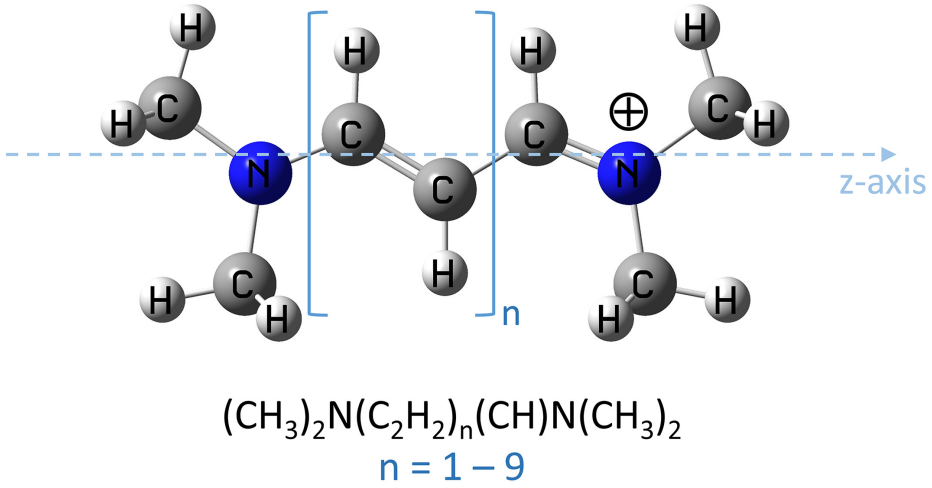
Cyanine dyes comprise a subset of π-conjugated organic systems that exhibit remarkably strong linear and nonlinear optical (NLO) response electronic properties [81–90]. In the past 50 years, these NLO properties have been harnessed in a variety of advanced materials and technologies such as electro-optic waveguide devices [91, 92], memory devices [93–95], three-dimension fluorescence microscopy [96,97], and nanofabrication [98,99]. To further enhance their usage in next-generation optical technologies, significant effort has been focused on achieving a detailed understanding of the electronic structure–property relationships20 that govern their performance in realistic materials. As such, predictive computational methods, particularly quantum chemical methods, offer a rational approach for guiding future experimental efforts to obtain tailored NLO properties in these functional π-conjugated systems [100].

Historically, wavefunction-based techniques, particularly coupled-cluster methods and Moller–Plesset perturbation theory (MP2 or MP4), have been used to provide accurate predictions for both the linear polarizability (α) and second hyperpolarizability (γ) in various NLO materials [101]. While it is generally accepted that the coupled-cluster method with single and double excitations and perturbative triple excitations (CCSD(T)) is widely viewed as the "gold standard" in quantum chemistry, its usage as a routine computational screening tool is significantly limited due to its immense computational cost (particularly for very large systems) [102–105]. Conversely, methods such as Kohn–Sham density functional theory (DFT)26 have garnered significant attention because of their favorable balance between computational cost and accuracy in large organic and inorganic

NLO materials [103, 106]. However, the specific choice of the exchange-correlation (XC) functional in DFT still poses a practical problem since the accurate prediction of both α and γ is particularly challenging for most DFT methods [103, 107]. In particular, XC functionals such as the local density approximation (LDA), generalized gradient approximation (GGA), and even conventional hybrid functionals that contain a fixed percentage of exact-exchange fail to provide an accurate description of polarizabilities in one dimensional π-conjugated polymers [103, 108]. These difficulties can be traced to the lack of a field-induced counteracting term in the response part of the XC functional [109].

In a very recent paper by Wang et al. [110], the authors examined the second hyperpolarizability (γ) in a series of streptocyanine systems (Fig. 4.1) using non-empirically tuned range-separated functionals and second-order Moller–Plesset perturbation theory (MP2) as benchmark standards. Based on their MP2 benchmarks, the authors concluded that hyperpolarizabilities and other electronic structure–property relationships, such as bond order alternation parameters (which are used in the design of π-conjugated molecular and polymer active materials), were predicted poorly with DFT methods.

To provide further insight into the surprising DFT trends reported by Wang et al. [110], we present new calculations using a variety of range-separated DFT functionals with new CCSD(T) and CCSD(T)-F12 calculations. Contrary to their previous studies, we find that the electronic ground states for several of the streptocyanine oligomers are not closed-shell singlets, which can be obtained by allowing the system to relax to a lower-energy broken-symmetry solution (this effect is similar to a recent paper published by our group which also found that enhanced accuracy with range-separated DFT could be obtained



$(CH_3)_2N(C_2H_2)_n(CH)N(CH_3)_2$

$n = 1 - 9$

Figure 4.1: Chemical structure of streptocyanine oligomers (n = 1–9) studied in this work. The linear polarizability (α) and second hyperpolarizability (γ) are computed along the z-axis shown in figure.

for other conjugated systems with a broken-symmetry solution [103]). In short, broken-symmetry effects arise when the restricted (closed-shell) wavefunction becomes unstable toward an unrestricted (open-shell) solution [111, 112]. To provide a stringent verification of our results, we carried out new CCSD(T) and explicitly correlated CCSD(T)-F12 calculations as benchmark standards in our study. It is important to note that Wang et al. [110] commented that their MP2 benchmarks likely underestimated γ by about 30%; as such, our CCSD(T) and CCSD(T)-F12 calculations permit a new assessment of the MP2 benchmarks by comprising the most complete and accurate calculations of polarizabilities and hyperpolarizabilities for the streptocyanine systems to date. We present a detailed analysis for all of these aforementioned effects and discuss the implications of using broken-symmetry effects and high-level CCSD(T) benchmarks for calculating polarizabilities and bond order alternation properties in these challenging streptocyanine systems.

4.2 Theory and Methodology

The primary goals of this study are to assess the linear polarizabilities (α) and second hyperpolarizabilities (γ) for streptocyanines using broken-symmetry range-separated density functional methods as well as to provide new high-level CCSD(T) and CCSD(T)-F12 benchmarks for these systems. The various quantum chemical methods utilized in this study are briefly reviewed below.

In contrast to conventional hybrid functionals [113, 114], the range-separated approach mixes short-range density functional exchange with long-range Hartree–Fork (HF) exchange by partitioning the electron–electron repulsion operator into short- and long-range parts (i.e., the mixing parameter is a function of electron coordinates). In the Coulomb-Attenuating Method (CAM) proposed by Yanai et al. [114–116], the most general form of the range-separated approach is given by the following partitioning

$$\frac{1}{r_{12}} = \frac{1 - [a + b \cdot erf(\omega \cdot r_{12})]}{r_{12}} + \frac{[a + b \cdot erf(\omega \cdot r_{12})]}{r_{12}}, \tag{4.1}$$

where erf stands for the standard error function, r_{12} is the interelectronic distance between electrons 1 and 2, and ω is the range-separated parameter in units of Bohr^{-1}. The parameters, a and b, satisfy the following inequalities: $0 \leq a + b \leq 1, 0 \leq a \leq 1$, and $0 \leq b \leq 1$. The partitioning in Eq. 4.1 allows a contribution of HF exchange over the entire range by a factor of a, while the parameter b allows us to incorporate long-range asymptotic HF exchange by a factor of $(a + b)$. Previously, we [117–123] and others [124, 125] have shown that maintaining a full 100% contribution of asymptotic HF exchange (i.e., constraining $a + b = 1$), was essential for accurately describing valence excitations in relatively simple molecular systems. Therefore, in the present work on the streptocyanine system, we fix the parameters

a = 0.0 (no short-range exchange) and b = 1.0 in conjunction with self-consistently tuning the range-separation parameter ω in the LC-ωPBE functional to satisfy DFT-Koopmans' theorem [126–128]. Briefly, this theorem states that the energy of the highest occupied molecular orbital, ϵ_{HOMO}, is equal to the negative of the ionization potential, –IP, for the exact exchange-correlation functional. The ionization potential is typically computed via a ΔSCF energy difference between the neutral and cationic electron system. Although several numerical schemes exist for self-consistently tuning ω to satisfy DFT-Koopmans' theorem, one widely used approach is to numerically minimize the following objective function, J^2:

$$J^2 = [\epsilon^{\omega}_{HOMO}(N) + IP^{\omega}(N)]^2 + [\epsilon^{\omega}_{HOMO}(N+1) + IP^{\omega}(N+1)]^2, \tag{4.2}$$

where $\epsilon^{\omega}_{HOMO}(N)$ and $\epsilon^{\omega}_{HOMO}(N+1)$ are the $HOMO$ energies of the N and $N+1$ electron systems, respectively. Similarly, $IP^{\omega}(N)]$ and $IP^{\omega}(N+1)]$ are the ionization potentials of the N and $N+1$ electron systems computed with the same ω value as respectively. Since DFT-Koopmans' theorem does not directly relate the negative of the LUMO energy to the electron affinity (EA), the second term in Eq. 4.2 takes into account the $N+1$ electron system to indirectly tune the LUMO energy of the N electron system.

To determine the optimal range-separation ω values for each oligomer, we carried out several single-point energy calculations by varying ω from 0.0 to 0.4 (in increments of 0.02) for each of the N, $N+1$, and $N-1$ electronic states. A spline interpolation was then used to refine these optimal ω values, which were then subsequently used in final LC-ωPBE calculations for each streptocyanine oligomer. For all of the oligomers, we also carried out a DFT stability analysis to converge (if possible) toward a lower-energy broken-symmetry solution, which allows for an unrestricted spin state and a reduction in the

symmetry of the orbitals. This additional stability analysis is necessary since very recent work by us had shown that large conjugated systems will favor a lower energy broken-symmetry configuration with DFT methods that inherently contain a large percentage of HF exchange (such as range-separated methods) [103].

To maintain a consistent comparison with the previous study of Wang et al., the same geometry optimization procedure from Ref. [110] was used throughout this work (i.e., all molecular geometries were optimized at the ωb97x/cc-pVTZ level of theory with C_{2v} symmetry constraints). The Cartesian coordinates for all of the oligomers investigated in this study are tabulated in the Supporting Information for completeness. Since all of the streptocyanine oligomers possess C_{2v} symmetry, the linear polarizability (α), and second hyperpolarizability (γ) are dominated by the longitudinal tensor component along the z-axis (cf. Fig. 4.1). As such, we only report the z-component of the linear polarizability and second hyperpolarizability in this study, which we calculated with a custom-developed code as second- and fourth-order numerical derivatives of the energy with respect to the external electric field strength. In these calculations, the following finite field amplitudes in atomic units ($1\ au = 5.14226 \times 10^{11}\ V/m$) were selected: $F = 0.0$, 1×10^{-4}, 2×10^{-4}, 4×10^{-4}, 8×10^{-4}, 16×10^{-4}, and $32 \times 10^{-4}\ au$. To assess the accuracy of the various DFT calculations, we also carried out large CCSD(T)/6-311G(d,p) and explicitly correlated CCSD(T)-F12 calculations as benchmark standards. We specifically chose to use the 6-311G(d,p) basis for all of our calculations since MP2/6-311G(d,p) benchmark calculations by us (see Results and Discussion section) demonstrated that this basis set was large enough to closely reproduce the MP2/Def2-TZVPD benchmarks originally used by

Wang et al. [110] (while, at the same time, the 6-311G(d,p) basis is small enough to use for our computationally-expensive CCSD(T) calculations). Extremely accurate electronic energies can be obtained with CCSD(T)-F12 approaches (typically at a higher level of accuracy than conventional CCSD(T) with the same basis) since these methods construct a wave function that depends explicitly on the interelectronic coordinates. As such, the explicitly correlated CCSD(T)-F12 methods exhibit dramatic improvements in basis set convergence, and results of quintuple-zeta quality that were obtained with smaller triple-ζ basis sets have previously been shown [129–131]. All linear polarizabilities (α) and second hyperpolarizabilities (γ) were obtained from self-consistent CCSD(T)-F12a energies since extensive benchmarks have indicated that the CCSD(T)-F12a variant gives better results than CCSD(T)-F12b, particularly for smaller basis sets24 (the F12b variant differs from the F12a method by the inclusion of an additional energy correction that approximately doubles the magnitude of the coupling between the conventional and explicitly correlated pieces of calculation [104,129]). Taken together, the α and γ values resulting from the CCSD(T)-F12 method serve as an additional check on both the DFT and CCSD(T) calculations as well as provide new high-quality benchmarks for the streptocyanine system. It is also worth mentioning that the previous study by Wang et al. [110] utilized lower-level MP2 calculations as benchmark standards, and the present work significantly extends this previous study by providing new, high-level CCSD(T) and CCSD(T)-F12 benchmarks to more stringently assess the merits/deficiencies of DFT methods for the various streptocyanine oligomers.

Finally, we also computed Mayer bond order and bond order alternation (BOA) parameters [132] to assess the electronic mixing between resonance structures for selected

streptocyanine oligomers. The atomic Mayer bond order, I_{AB}, between two adjacent atoms, A and B, is defined by the following:

$$I_{AB} = \sum_{a \in A} \sum_{b \in B} \left[(PS)_{ba}(PS)_{ab} + (P^s S)_{ba}(P^s S)_{ab} \right], \tag{4.3}$$

where P, P^S, and S correspond to the total density matrix, spin density matrix, and overlap matrix, respectively. The BOA is the average difference in bond orders between adjacent carbon–carbon bonds in the π-electron conjugated bridge in the geometry-optimized ground state [133]. As such, the BOA provides a good measure and predictor of linear and nonlinear optical (NLO) properties. Since all of the streptocyanine oligomers in our calculations were optimized under C_{2v} symmetry constraints, we calculated the BOA as the mean value of the absolute bond order difference of adjacent atoms in the π-conjugation bridge (i.e., the carbon-carbon backbone depicted in Fig. 1) according to the following expression:

$$\text{BOA} = \text{avg}(|I_{12} - I_{23}| + |I_{34} - I_{45}| + |I_{56} - I_{67}| + \cdots), \tag{4.4}$$

All CCSD(T) and CCSD(T)-F12 calculations were carried out with the MOLPRO package [134], and all DFT calculations were performed with the Gaussian 09 software package [135]. The numerical differentiations required to obtain α and γ were calculated with custom-developed codes, and the Multiwfn 3.3.9 package [136] was utilized to calculate the various bond order parameters.

4.3 Results and Discussion

Figure 4.2 shows the objective function, J^2 [cf. Eq. 4.2], as a function of the range-separated parameter, ω, for the various streptocyanine oligomers obtained with (a) the

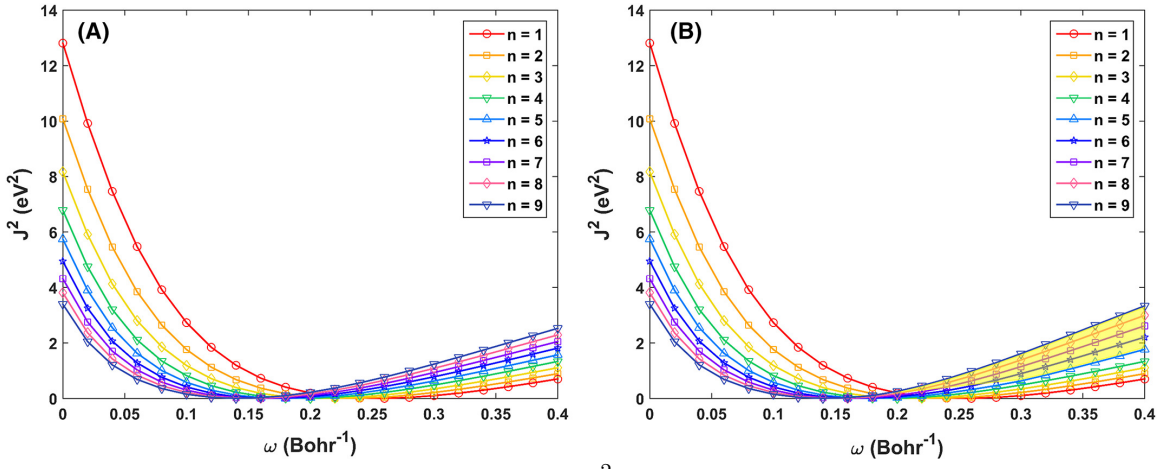


Figure 4.2: Plots of the objective function, J^2 [cf. Eq. 4.2], as a function of the range-separated parameter, ω, for streptocyanines using the non-empirically tuned LC-ωPBE functional (A) without broken-symmetry solutions and (B) with symmetry-broken solutions. The yellow-shaded region denotes regions (primarily for the n = 5–9 oligomers) where a broken-symmetry solution is obtained.

conventional LC-ωPBE functional and (b) an LC-ωPBE approach that allows for a broken-symmetry (BS) solution. A broken-symmetry solution was obtained by performing a full DFT stability analysis to give a lower-energy broken-symmetry solution (if possible). Figure 4.2b shows that a broken-symmetry solution is obtained at the LC-ωPBE/6-311G(d,p) level of theory for large values of ω. This result is in line with a previous study by our group which showed that π-conjugated systems undergo an electronic symmetry breaking as a function of backbone length [103]. Since the chemical structure of the streptocyanine system is composed of successive single-double carbon–carbon bonds, the π orbitals are much more strongly conjugated along the backbone. Furthermore, from a more theoretical standpoint, a lower-energy broken-symmetry configuration is more favored if the DFT method contains a higher percentage of HF exchange [112, 137–140], (such as range-separated functionals with large values of ω). It is also interesting to note that only the streptocyanine oligomers with more than 4 monomer units will exhibit broken-symmetry

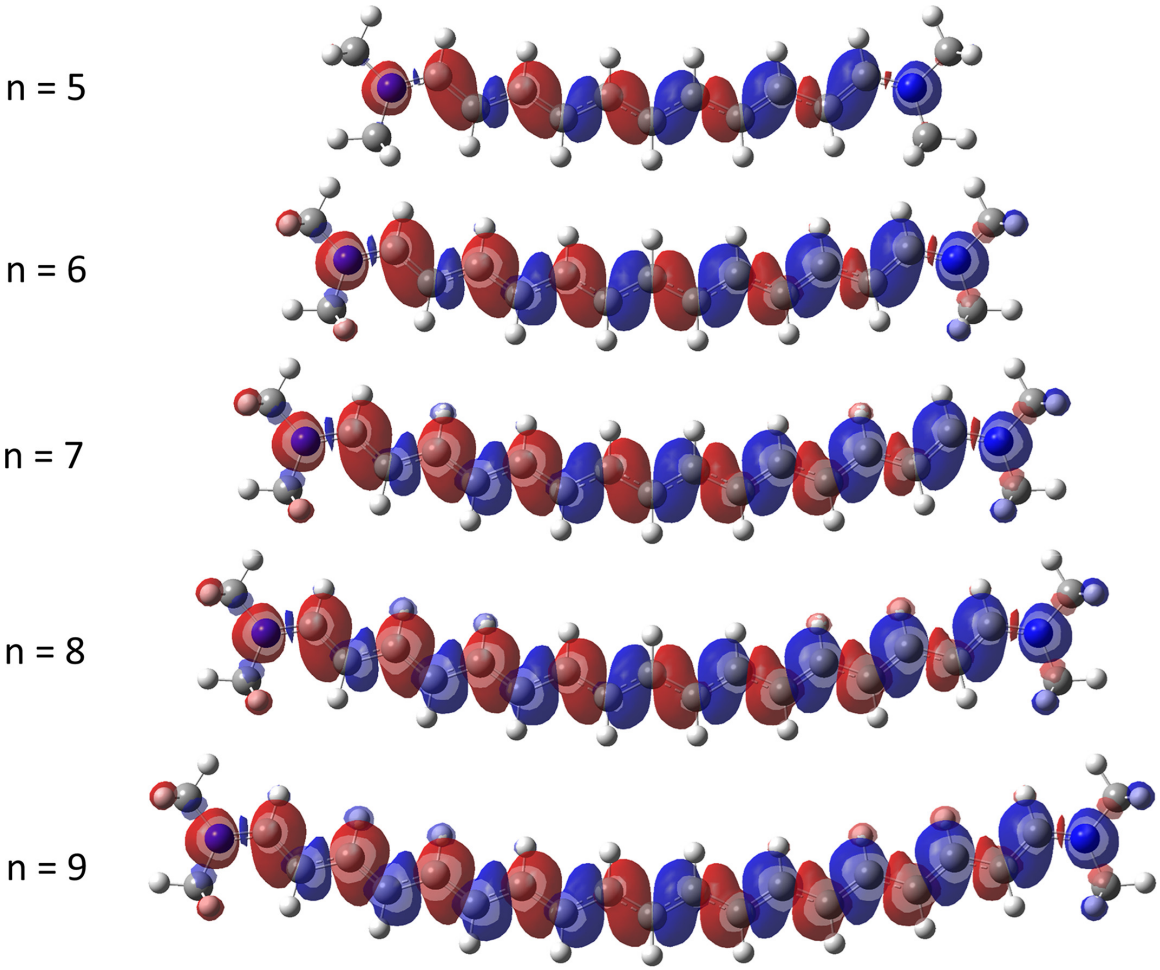


Figure 4.3: Spin density difference (blue = positive spin density and red = negative spin density) obtained with LC-ωPBE $\omega = 0.33$ for the various streptocyanine oligomers.

configurations where the alpha and beta spin densities alternate through the whole backbone of the molecule when ω is sufficiently large. Moreover, the long-range ordering of the spin density still persists when the length of the streptocyanine chain is increased. This can be ascribed to electron localization of the p orbitals on the carbon atoms, which results in an antiferromagnetic pattern that can be visualized as alternating spin density differences in Figure 4.3. Only streptocyanines with monomer units from n = 5 to 9 are shown since the ground states for the smaller streptocyanines are closed-shell singlets and do not exhibit an

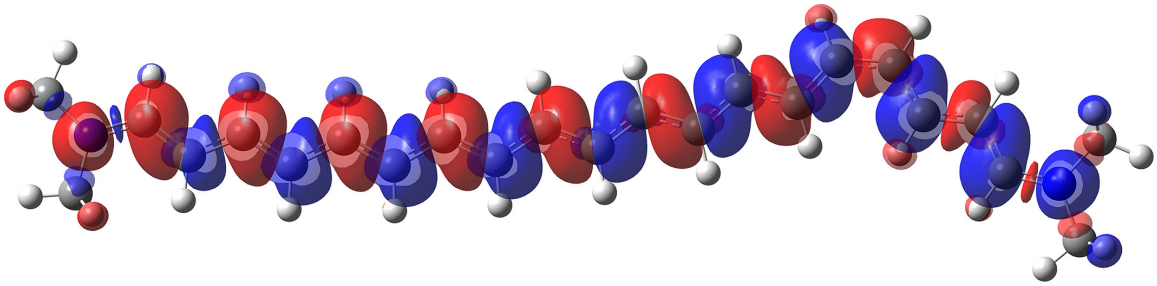


Figure 4.4: Spin density difference (blue = positive spin density and red = negative spin density) obtained with LC-ωPBE $\omega = 0.33$ for the lowest-energy electronic state of a distorted n = 9 streptocyanine oligomer.

antiferromagnetic pattern. While the present work focuses on streptocyanine oligomers possessing a C_{2v} symmetry (i.e., to permit a consistent and fair comparison with the previous study by Wang et al. [110]), it is natural to ask whether a broken-symmetry solution is still obtained if the molecular symmetry is significantly reduced. To test this effect, we distorted the largest streptocyanine oligomer by intentionally introducing a cis-conformational "kink" in the original all-trans-conformation backbone, as shown in Figure 4.4. On convergence, we found that the broken-symmetry solution is more stable and 0.08 eV lower in energy than the closed-shell (restricted) solution at the LC-ωPBE DFT ($\omega = 0.33$) level of theory. As such, these calculations show that a broken-symmetry solution can still persist in the midst of disorder, and one should still check for these broken-symmetry effects when the molecular symmetry is reduced.

Table 4.1 lists the optimally tuned ω values as a function of size for the various streptocyanine oligomers. As the optimally tuned ω values are the same for both the LC-ωPBE and LC-ωPBE (BS) levels of theory, we only list the values for the former in Table 4.1. Interestingly, both the optimal ω numerical values and their electronic properties (such as the HOMO or IP energies) are not affected by the broken-symmetry solutions since the

shaded regions (where these broken-symmetry solutions occur) are located to the right of the minima of all the J^2 curves, as shown in Figure 4.2. From Table 4.1, the optimal ω values reflect the trend that larger-sized streptocyanine oligomers have smaller optimal ω values compared to smaller-sized oligomers, and for all cases, these are smaller than the "default" ω value of 0.33. These findings are in line with the nonempirically tuned ω values (shown in the parentheses in Table 4.1) reported by Wang et al [110]. Moreover, a similar size-dependence has already been reported in several other chemical systems [103,119,120,128,141,142]. All the subsequent linear polarizability (α) and second hyperpolarizability (γ) calculations were evaluated using the optimum ω values summarized in Table 4.1.

Table 4.1: Nonempirically tuned ω values for streptocyanine at the LC-ωPBE/6-311G(d,p) level of theory. Values in parentheses in the second column denote the nonemperically tuned LC-ωPBE/Def2-TZVPD values by Wong et al. [110]

n	**LC-ωPBE**
1	0.254 (0.244)
2	0.230 (0.222)
3	0.210 (0.203)
4	0.194 (0.188)
5	0.180 (0.175)
6	0.168 (0.164)
7	0.158 (0.154)
8	0.149 (0.146)
9	0.142 (0.138)

Table 4.2 gives a concise summary and analysis of α and γ for streptocyanine systems containing up to nine monomer units as obtained by the LC-ωPBE DFT (ω = 0.33, 0.40, and nonempirically tuned values) and MP2 methods in comparison to the new high-level CCSD(T) and explicitly correlated CCSD(T)-F12 benchmarks. Table 4.2 also includes the broken-symmetry (BS) HF and DFT solutions to allow for a direct comparison with the closed-shell DFT values. It is worth noting that while the broken-symmetry HF and DFT polarizabilities are numerically quite different from each other, the spatial spin density differences are similar, and the broken-symmetry HF densities resemble those shown in Figure 4.3. Additionally, the results in the parentheses for the MP2 and LC-ωPBE (ω = tuned) functionals were taken from Ref. [110]. As can be seen in Table 4.2, there are only minor discrepancies between our calculations and the prior MP2 benchmarks by Wang et al. [110], which justifies our use of the 6-311G(d,p) basis set for the DFT, MP2, and CCSD(T) calculations in the present study. Another minor discrepancy is the finite-field expressions used in the current work; specifically, Wang et al., calculated hyperpolarizabilities via finite differences of the dipole moment using the expression $\gamma = [\mu_z(2\Delta_z) - 2\mu_z(\Delta_z) + 2\mu_z(-\Delta_z) - \mu_z(2\Delta_z)]/[2(\Delta_z^3)]$, where μ_z is the dipole moment component along the z-axis and Δ_z (= 0.0003 au) is the variation of the electric field. In our study, however, we directly fitted the energies (as a function of the external electric field) to second- and fourth-order polynomials to evaluate α and γ, respectively. We specifically chose to calculate γ using finite differences of energies since the CCSD(T) and CCSD(T)-F12 dipole moments from the correlated coupled-cluster density matrix are not available in most

Table 4.2: Linear polarizability (α) and hyperpolarizability (γ) for increasingly large streptocyanine oligomers at various levels of theories. All HF, MP2, CCSD(T), and CCSD(T)-F12 calculations utilized 6-311G(d,p) basis set. The abbreviation "BS" and ω=tuned indicate broken-symmetry calculation and a nonemperically tuned value of ω, respectively. Values in parentheses denote the nonemperically tuned LC-ωPBE and MP2 values by Wang et al. [110]

n	HF	LC-ωPBE ω=0.33	LC-ωPBE (BS) ω=0.33	LC-ωPBE ω=0.40	LC-ωPBE (BS) ω=0.40	LC-ωPBE ω=tuned	LC-ωPBE (BS) ω=tuned	MP2	CCSD (T)	CCSD (T)-F12
α (10^{-24}esu)										
1	24.7	26.6	26.6	26.3	26.3	26.9	26.9	28.4	28.8	29.2
2	44.5	48.9	48.9	48.5	48.5	49.1	49.1	54.3	57.3	57.5
3	71.9	81.3	81.3	81.3	81.3	80.7	80.7	93.7	103.6	103.5
4	365.4	125.8	125.8	126.1	125.4	122.6	122.6	149.0	172.6	171.8
5	1536.8	183.4	192.1	184.7	213.8	175.6	175.6	222.2	269.5	267.9
6	2051.2	255.4	300.8	257.8	357.1	240.4	240.4	315.5	399.2	397.1
7	2533.9	341.3	455.8	345.4	579.4	316.6	316.6	429.3	563.8	562.9
8	3000.6	441.6	670.9	447.5	897.8	404.5	404.3	563.8	763.0	765.5
9	3462.8	556.3	955.9	563.8	1306.6	505.1	502.7	719.6	996.6	1005.6
MAE	1092.7	143.8	55.2	141.5	71.0	159.2	159.5	86.5		
γ (10^{-36}esu)										
1	0	208.5	208.5	129.0	129.6	98.8	98.8	-48.2 (-63.6)	-53.8	-32.2
2	0	-361.5	-361.5	-239.9	-239.9	-413.5 (-206.6)	-413.5	-355.6 (-502.2)	-757.7	697.2
3	20.5	-326.0	-326.0	-776.0	-776.0	-486.2 (-438.8)	-486.2	-1629.1 (-1595.5)	-4336.0	-4788.4
4	114.6	-2088.8	-2088.8	-2338.1	13952.8	1189.5 (-1331.6)	-1189.5	-5395.4 (-5386.8)	-16427.9	-16783.0
5	-3128.9	-6545.4	17551.8	-76629.3	22887.6	-3057.1 (-3155.1)	-3057.1	-14351.3c(-13834.6)	-49420.9	-49101.1
6	-4289.5	-15727.9	5523.2	-17676.6	-28440.0	-6942.6 (-6689.8)	-6942.6	-32797.0 (-32391.4)	-121840.3	-122917.0
7	-5366.4	-32338.8	-71322.0	-36739.0	-252693.8	-12526.5 (-12992.4)	-12526.5	-65867.2 (-67417.6)	-255482.3	-263970.6
8	-6398.0	-60167.2	-279902.8	-68500.7	-752385.5	-21414.7 (-23237.3)	-16974.7	-118522.2 (-123962.1)	-464455.9	-488312.3
9	-7415.7	-101779.1	-676593.4	-115022.6	-1536932.8	-33744.6 (-38333.2)	3522.8	-194389.2 (-213884.2)	-758199.8	-807627.0
MAE	182723.5	161316.5	73740.4	158020.2	141089.1	176811.0	181445.1	137513.3		

standard quantum chemistry codes. All of our DFT and CCSD(T)/CCSD(T)-F12 results for α and γ were computed using the polynomial fitting procedure discussed above, which we found in previous work [103] to be more stable than other numerical techniques such as the Romberg differentiation procedure. All mean absolute errors (MAE) were computed with respect to the CCSD(T)/6-311G(d,p) benchmarks to allow for a consistent comparison with DFT calculations.

At this point, it is natural to ask whether our CCSD(T) values are accurate benchmarks for our DFT calculations. Particularly, in light of our previous discussion on broken-symmetry effects in DFT, one can naturally inquire whether broken-symmetry CCSD(T) calculations should be used as reference values instead. First, to check for possible non-dynamical correlation effects in our restricted HF-reference CCSD(T) calculations, the T1 diagnostic for all streptocyanine oligomers was also computed. We obtained T1 values ranging from 0.0148 to 0.0207 for the n = 1 to n = 9 streptocyanine oligomers, which indicate that these conjugated systems are relatively well described by a single-reference electron correlation method. Second, we also carried out new unrestricted CCSD(T) benchmark calculations obtained from a broken-symmetry HF reference determinant (UHF-CCSD(T)). Surprisingly, we found that all of the UHF-CCSD(T) energies were larger (i.e., less stable) than their restricted CCSD(T) counterparts, further verifying (in conjunction with the T1 diagnostic discussed previously) that correlation effects are properly handled at the single-reference restricted CCSD(T) level of theory. The failure of UHF-CCSD(T) to yield accurate energies for these systems seems paradoxical, particularly since our DFT calculations seem to imply that the symmetry-broken reference state is a more correct zeroth-order

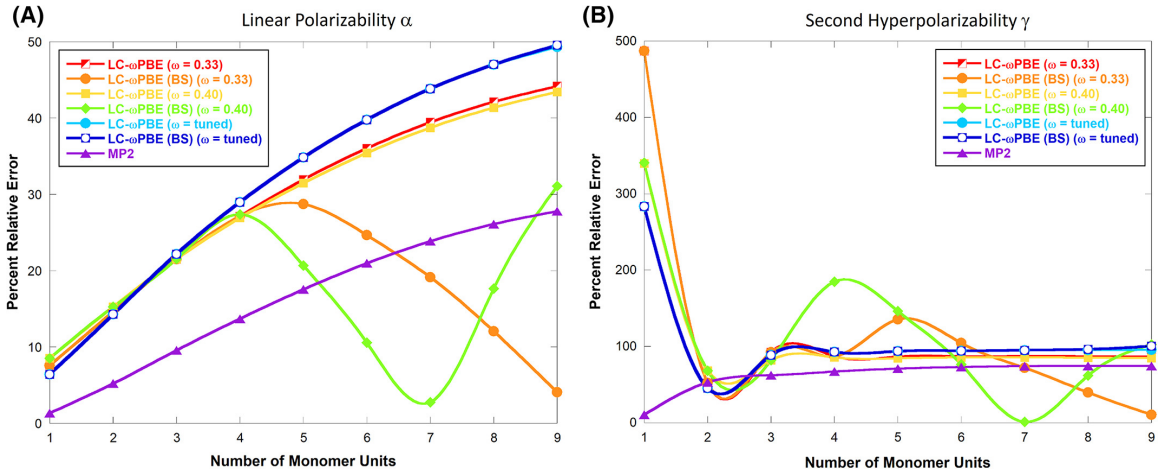


Figure 4.5: Percent relative error (compared to CCSD(T) benchmarks) of (A) linear polarizability and (B) second hyperpolarizability as a function of streptocyanine length for different approximation levels of theory.

description of the system (compared to the restricted reference state). In a seminal review on coupled-cluster methods [143], Lee and Scuseria stated,"when performing approximate electron correlation calculations, it is better to use a symmetry constrained reference function rather than a symmetry broken reference function... *This is in spite of the fact that the symmetry broken reference function possesses lower energy, and these conclusions appear to be valid whether discussing spin symmetry or spatial symmetry.*" Our CCSD(T) and UHF-CCSD(T) trends are in agreement with Lee and Scuseria's observations, giving us additional confidence that the restricted HF-reference CCSD(T) calculations are accurate benchmarks for our DFT calculations (instead of the broken-symmetry CCSD(T) results).

Figure 4.5 depicts the percent relative errors in the linear polarizability ($= |[\alpha_{comp.} - \alpha_{CCSD(T)}]/\alpha_{CCSD(T)}| \times 100\%$) and the second hyperpolarizability ($= |[\gamma_{comp.} - \gamma_{CCSD(T)}]/\gamma_{CCSD(T)}| \times 100\%$) with respect to oligomer chain length, where the subscript "comp." denotes the various computational methods surveyed and CCSD(T)

is the benchmark reference. For conciseness, we do not plot the broken-symmetry HF results in Fig. 4.5 since the relative errors are extremely large and off-scale. In Fig. 4.5A, the DFT results overshoot the CCSD(T) reference values for the linear polarizability, and the overestimation is expected to further increase with size as the curves have not attained a characteristic plateau in the asymptotic limit. When n = 9, most of the LC-ωPBE variants yield a fairly large relative error (> 40%) for the linear polarizability, except for the broken-symmetry LC-ωPBE $\omega = 0.33$ and LC-ωPBE $\omega = 0.40$ functionals. For $1 \leq n \leq 4$, the relative errors are very similar among the DFT methods irrespective of whether they include broken-symmetry effects or whether ω is nonempirically tuned or not. A remarkably different behavior can be observed from both the LC-ωPBE $\omega = 0.33$ (BS) and LC-ωPBE $\omega = 0.40$ (BS) functionals in which the relative error surprisingly decreases after n = 5 and n = 4, respectively. As such, allowing the system to converge toward a lower-energy broken-symmetry solution improves the accuracy in the linear polarizability (although the relative error increases after n = 7 for the LC-ω PBE $\omega = 0.40$ (BS) functional). In their previous work, Wang et al. [110] showed that these errors could be reduced when electron correlation is perturbatively included at the MP2 level of approximation; however, our CCSD(T) benchmarks indicate that the MP2 polarizabilities still show significant errors (25%) that increase as a function of size. As such, our results indicate that the MP2 method should not be used as reliable benchmarks for polarizabilities, in contrast to the previous MP2 study by Wang et al [110]. It is interesting to note that the LC-ωPBE $\omega = 0.33$ (BS) functional gives the best agreement with the lowest MAE value of 55.16×10^{-24} *esu* compared to the CCSD(T) benchmarks. As far as α is concerned, the MAEs can be summarized in the fol-

lowing order: LC-ωPBE $\omega = 0.33$ (BS) < LC-ωPBE $\omega = 0.40$ (BS) < MP2 < LC-ωPBE $\omega =$ 0.40 < LC-ωPBE $\omega = 0.33$ < LC-ωPBE $\omega =$ tuned < LC-ωPBE $\omega =$ tuned (BS). According to this sequence, the LC-ωPBE $\omega = 0.33$ (BS) and LC-ωPBE $\omega = 0.40$ (BS) functionals are in better agreement with the CCSD(T) benchmarks than their counterparts without broken-symmetry solutions (a reduction in MAE from 143.75×10^{-24} to 55.16×10^{-24} esu and from 141.45×10^{-24} to 71.00×10^{-24} esu, respectively).

Turning to the second hyperpolarizability (γ) in Fig. 4.5b, we observe significantly larger deviations—for the smallest oligomer (n = 1), all levels of theory except MP2 show extremely large relative errors. In addition, the curves resulting from the LC-ωPBE ω = tuned, LC-ωPBE ω = tuned (BS), LC-ωPBE $\omega = 0.33$, and the LC-ωPBE $\omega = 0.40$ functionals reach a characteristic plateau in the asymptotic limit at around 100% relative error. Similar to the linear polarizability, the LC-ωPBE $\omega = 0.33$ (BS) and LC-ωPBE $\omega =$ 0.40 (BS) functionals exhibit a nonmonotonic behavior as a function of size. Specifically, the relative errors for both of these functionals fluctuate near $2 \leq n \leq 5$ and subsequently decrease for $n \geq 5$. Overall, the overestimation is more severe for γ than α for both the DFT and MP2 methods. In our study, the γ values resulting from the LC-ωPBE $\omega =$ tuned functional closely mirror those from Wang et al. [110], with the exception of the n = 1 and 2 monomer units which they mentioned were underestimated compared to the default LC-ωPBE functional. Similarly, our MP2 results are in relatively good agreement with those reported by Wang et al., as shown in Table 4.2. Nevertheless, the relative error of the MP2 method compared to the CCSD(T) benchmarks is still quite sizeable and reaches an asymptotic limit of around 80%. These trends in the relative error are consistent with

that observed for the linear polarizabilities depicted in Figure 4.5A, emphasizing the fact that electron correlation effects beyond MP2 (i.e., at the CCSD(T) level of theory) play a major role in π-conjugated systems. For γ, we noticed that the LC-ωPBE ω = 0.33 (BS) method still gives the most accurate results due to its lowest MAE value of 73,740.40 (10-36 esu) among all other methods. To summarize, the MAEs for γ can be summarized as follows: LC-ωPBE ω = 0.33 (BS) < MP2 < LC-ωPBE ω = 0.40 (BS) < LC-ωPBE ω = 0.40 < LC-ωPBE ω = 0.33 < LC-ωPBE ω = tuned < LC-ωPBE ω = tuned (BS). Among all the DFT methods examined here, the LC-ωPBE ω = 0.33 (BS) functional is the most accurate for calculating both α and γ for the streptocyanine system (although none of the DFT methods still predict the correct sign of γ for the shortest oligomer), indicating the importance of including broken-symmetry effects when calculating polarizabilities and hyperpolarizabilities for this chemical system.

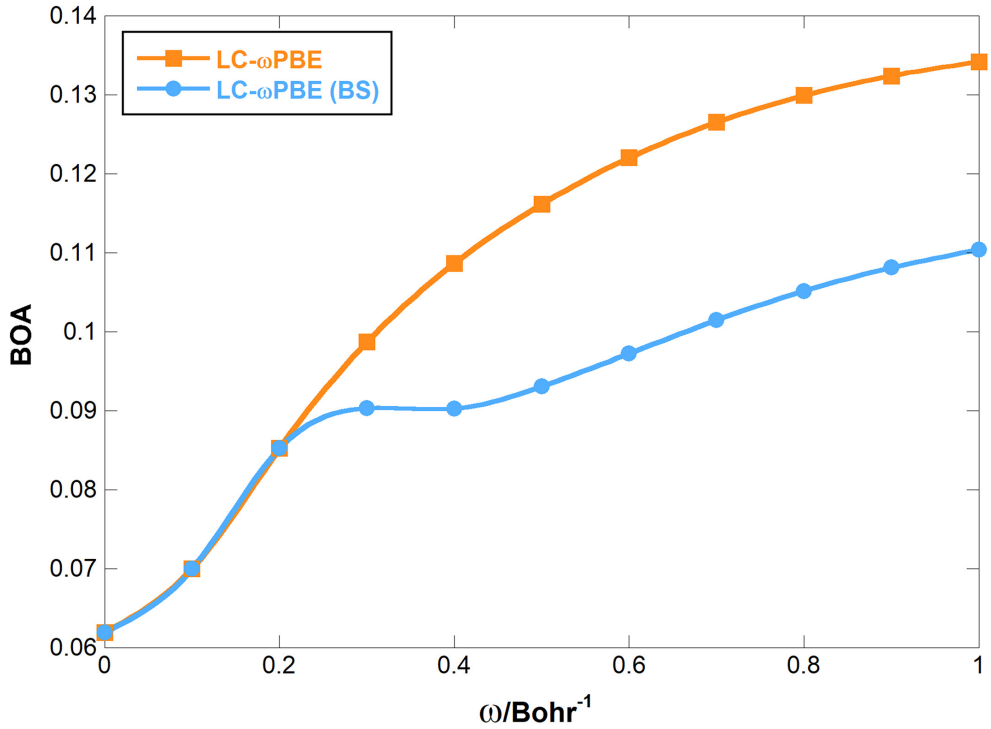


Figure 4.6: BOA for the n = 6 streptocyanine oligomer as a function of the range-separated parameter ω for LC-ωPBE and LC-ωPBE (BS), respectively

The bond order alternation (BOA) gives a predictor of linear and nonlinear optical (NLO) properties since it provides a measure of the electronic mixing between resonance structures in a π-conjugated system. Figure 4.6 demonstrates the relationship between the BOA and ω for the streptocyanine hexamer (n = 6) computed with the LC-ωPBE and LC-ωPBE (BS) functionals. Within the $0 \leq \omega \leq 0.2$ range, both LC-ωPBE and LC-ωPBE (BS) exhibit similar trends for the BOA (which is not surprising, since broken-symmetry effects come into play for larger values of ω). The BOA increases monotonically with ω for the conventional LC-ωPBE functional; however, for the LC-ωPBE (BS) functional, an inflection point occurs near $\omega = 0.3$, at which the BOA decreases initially then moderately increases once again.

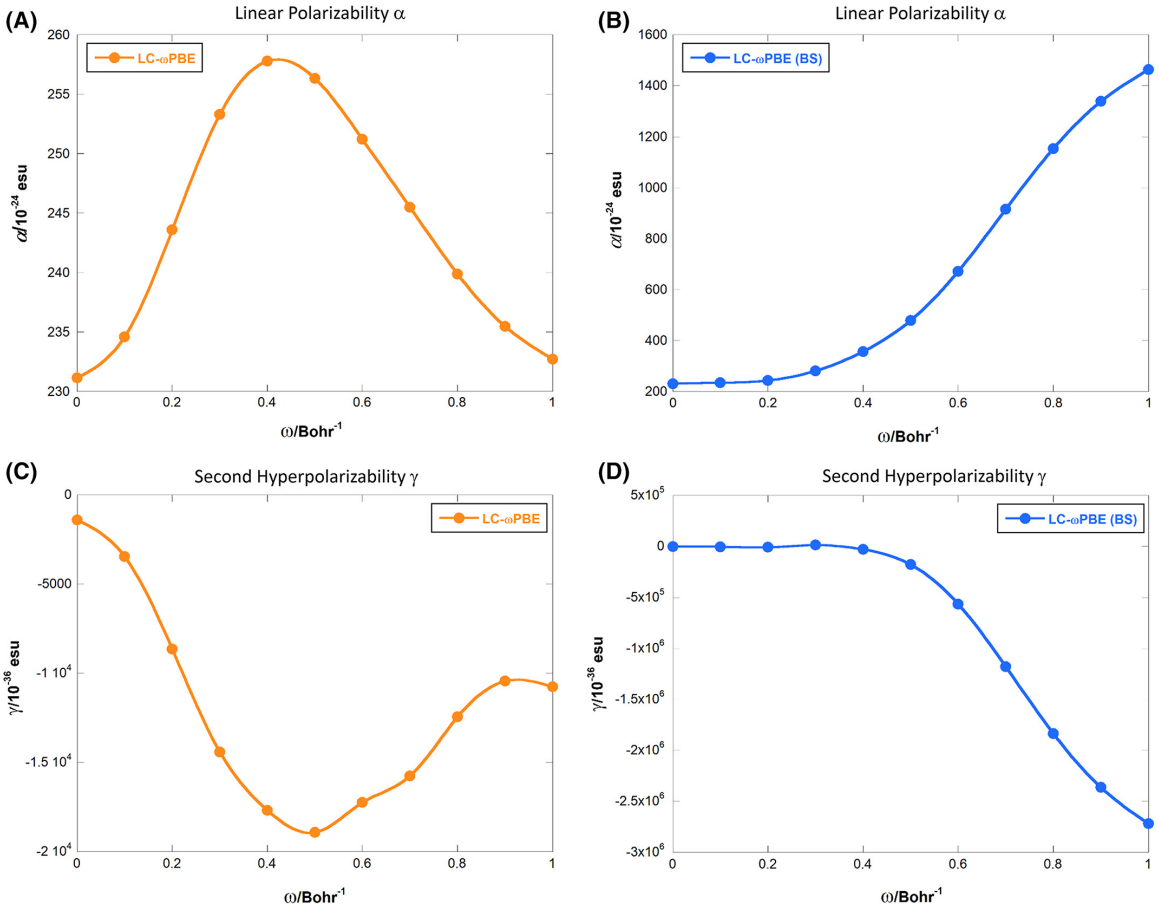


Figure 4.7: Structure–property relationships as a function of the range-separation parameter, ω, for the n = 6 streptocyanine oligomer: (A) LC-ωPBE α values, (B) broken-symmetry LC-ωPBE α values, (C) LC-ωPBE γ values, and (D) broken-symmetry LC-ωPBE γ values.

Figure 4.7 depicts the trends in α and γ as a function of ω for the n = 6 streptocyanine oligomer. Our trends in α and γ are similar to the closed-shell calculations for the n = 7 streptocyanine oligomer reported by Wang et al [110]. However, when we allow the system to relax to a lower-energy broken-symmetry configuration, we obtain extremely different trends: α increases incrementally from $\omega = 0$ to 0.3 and subsequently grows substantially, whereas γ decreases moderately until $\omega = 5$, at which point it then diminishes rapidly. These qualitatively different trends are due to the extended/delocalized antiferromagnetic

electron pattern (cf. Fig. 4.3) resulting from the broken-symmetry formalism (which, in turn, arises from a large amount of Hartree–Fock exchange present at large ω values). In other words, increasing ω in a conventional closed-shell range-separated DFT calculation leads to an over-localization of electron [119,120] and, hence, decreasing linear polarizability (cf. Fig. 4.7a). However, allowing the system to relax to a lower-energy broken-symmetry configuration will instead delocalize electrons into the extended patterns seen in Fig. 4.3, yielding a linear polarizability that increases as a function of ω, as seen in Fig. 4.7b.

4.4 Conclusion

Within this extensive theoretical study, we have analyzed a variety of range-separated DFT, MP2, CCSD(T), and CCSD(T)-F12 methods for predicting the linear polarizability (α) and the second hyperpolarizability (γ) in a series of π-conjugated streptocyanine oligomers. Contrary to previous studies, we have found that the electronic ground states for several of the streptocyanine oligomers are not closed-shell singlets, which can be obtained by allowing the system to relax to a lower-energy broken-symmetry solution. Our extensive analyses are complemented by new large-scale CCSD(T) and explicitly correlated CCSD(T)-F12 calculations that comprise the most complete and accurate benchmarks of α and γ for the streptocyanine systems to date. Based on these new benchmarks, we find that the broken-symmetry LC-ωPBE $\omega = 0.33$ functional yields the most accurate results among the various DFT methods for predicting both α and γ. In addition, our CCSD(T) calculations indicate that the MP2 benchmarks used in previous studies still exhibit significant errors (~25% for α and ~100% for γ) and, therefore, the MP2 calculations should

not be used as reliable benchmarks for polarizabilities or hyperpolarizabilities. Finally, we explored other structure–property relationships, such as bond order alternation (BOA) parameters, to assess the electronic mixing between resonance structures in the π-conjugated streptocyanine system. From our BOA analysis, we observe qualitatively different trends (compared to the conventional closed-shell calculations) for both α and γ that result from the broken-symmetry formalism. Taken together, our new CCSD(T) and CCST(T)-F12 benchmarks in conjunction with our broken-symmetry range-separated DFT calculations (1) highlight the importance of incorporating electron correlation effects beyond MP2 (i.e., at the CCSD(T) level of theory) for benchmarking both α and γ in π-conjugated streptocyanines and (2) emphasize the importance of testing for a lower-energy open-shell configuration when calculating nonlinear optical properties for these systems.

Chapter 5

A New Interpretation of the Structure and Solvent Dependence of the Far UV Circular Dichroism Spectrum of Short Oligopeptides

This work resulted from collaboration with Prof. Reinhard Schweitzer-Stenner in the Department of Chemistry at Drexel University. Prof. Reinhard assisted with the initial manuscript preparation. This work is published in the *Chemical Communication* journal. [1]

5.1 Introduction

Far UVCD spectroscopy is one of the most employed approaches for determining the secondary structure of proteins and peptides in solution [144, 145]. Each secondary structure gives rise to a fingerprint with a characteristic sequence of positive and negative maxima. The structural sensitivity of the UVCD spectra results from the orientational dependence of couplings between electronic transitions in different peptide groups [146].

Unfolded and disordered proteins, polypeptides, and even very short tripeptides such as glycylalanylglycine (GAG) frequently exhibit UVCD spectra with a pronounced negative maximum at ca. 190 nm. Additionally, poly- and oligopeptides with a substantial alanine or proline content exhibit a weak positive maximum at ca. 214 nm (Fig. 1) [147,148]. The conventional understanding of such CD spectra is that they arise from a random coil state in which the backbone of the peptide samples the entire sterically allowed region of the Ramachandran plot.1 However, this notion was already challenged ca. fifty years ago when Tiffany and Krimm assigned it to a polyproline II (pPII) conformation that exhibits dihedral angles of ϕ = -70° and ψ = 140.6° [149]. Their interpretation was challenged because it was generally believed that short peptides cannot adopt any well-defined conformation in aqueous solution. However, thanks to the work of Woody and colleagues [150–153], Tiffany and Krimm were eventually proven to be correct. Today, we know that the very pronounced pPII signal in the spectrum of alanine-containing peptides reflects the high propensity of the corresponding amino acid residue for this conformation [154, 155].

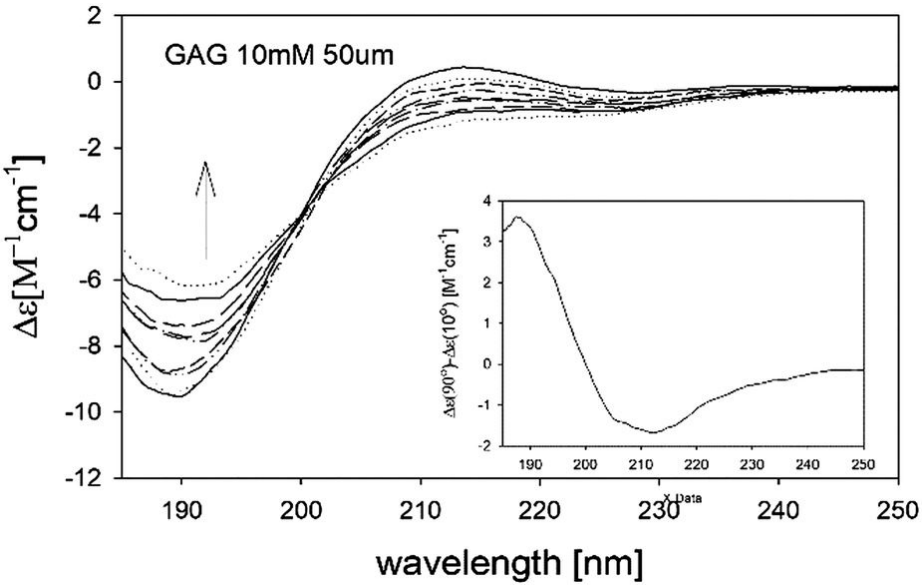


Figure 5.1: UV ECD spectrum of cationic GAG in water measured as a function of temperature. The arrow indicates the direction of the changes below 200 nm. The inset shows the difference spectrum calculated by subtracting the spectrum measured at 10°C from the spectrum taken at 90°C. Taken from ref. [146] and modified.

While the relationship between the strongly negatively-biased positive couplet in Fig. 5.1 and the pPII conformation has been firmly established over time, an explanation of the underlying physics turned out to be challenging. Generally, the UVCD spectra of secondary structures are explained in terms of excitonic coupling models, which describe the rotational strength of electronic transitions in terms of either dipole moments or transition charge densities associated with individual electronic transitions between the HOMO and the lowest unoccupied MOs of individual peptide groups [150, 156]. The canonical UVCD spectra were thus described by considering interactions between the π(HOMO) $\rightarrow$ π^*(LUMO) (NV1) and n $\rightarrow$ π^* transitions, which were shown to produce considerable rotational strength at the wavelength position of the latter. The n $\rightarrow$ π^* transition is barely visible in the corresponding absorption spectra owing to its rather small electronic transi-

tion dipole moment. The rotational strength, however, is written as the dot product of the electric and magnetic transition dipole moment so that it can become detectable even in the absence of a visible contribution to the absorption spectrum.

The above model can only account for symmetric CD signals where the integrated intensities of the negative and positive bands are identical. Apparently, this notion does not apply to the pPII signal in Fig. 5.1. To remedy the situation, several attempts have been made to include polarization effects in the formalism for the rotational strength [150, 157]. Empirically, the polarizability tensor is treated as a ground state property with empirical values for chemical bonds listed in the literature.

We wondered whether a more definite physical assessment of the UVCD spectra of short peptides could be provided by time-dependent density functional theory (TDDFT) calculations. Previous studies [158–160] have utilized TDDFT calculations to examine the underlying physics of the CD spectra of model peptides and proteins in their folded state. A similar investigation of unfolded peptides is still outstanding. We recently performed DFT calculations on four different cationic GxG peptides (x = A, V, I, L) in implicit and explicit water (10 H2O molecules) in order to obtain the energetics of their pPII and β-strand conformations [161]. For the present study, we used the optimized geometry of these two conformations in TDDFT calculations at the ωB97X-D/cc-pVTZ level of theory to calculate the UVCD and corresponding absorption spectra of GAG in implicit and explicit water [161]. We specifically chose the ωB97X-D functional for our studies since it contains an asymptotically-correct (range-separated) portion of nonlocal exchange in conjunction with dispersion corrections, which are essential for accurately predicting charge-transfer ex-

citations and hydrogen-bonding interactions [162], respectively. It is also important to note that prior work by Neto et al. [163] has shown that the optimal range-separated parameter, ω, required to accurately predict excited states in the explicit solvent is 0.25, which is already close to the default value of $\omega = 0.2$ used in the ωB97X-D functional. To test the effects of (1) changing the range-separation value, (2) using different exchange–correlation functionals, (3) utilizing a larger basis set (such as aug-cc-pVTZ), and (4) altering the positions of explicit water molecules, we spot-checked our calculations by comparing the computed ECD spectra obtained separately with each of these effects. As shown in Figs. A1-A.4 in Appendix A, we did not observe any significant change in the ECD spectra within the relevant wavelength range, which further validates our choice of the ωB97X-D/cc-pVTZ level of theory used in our studies.

5.2 Results and Discussion

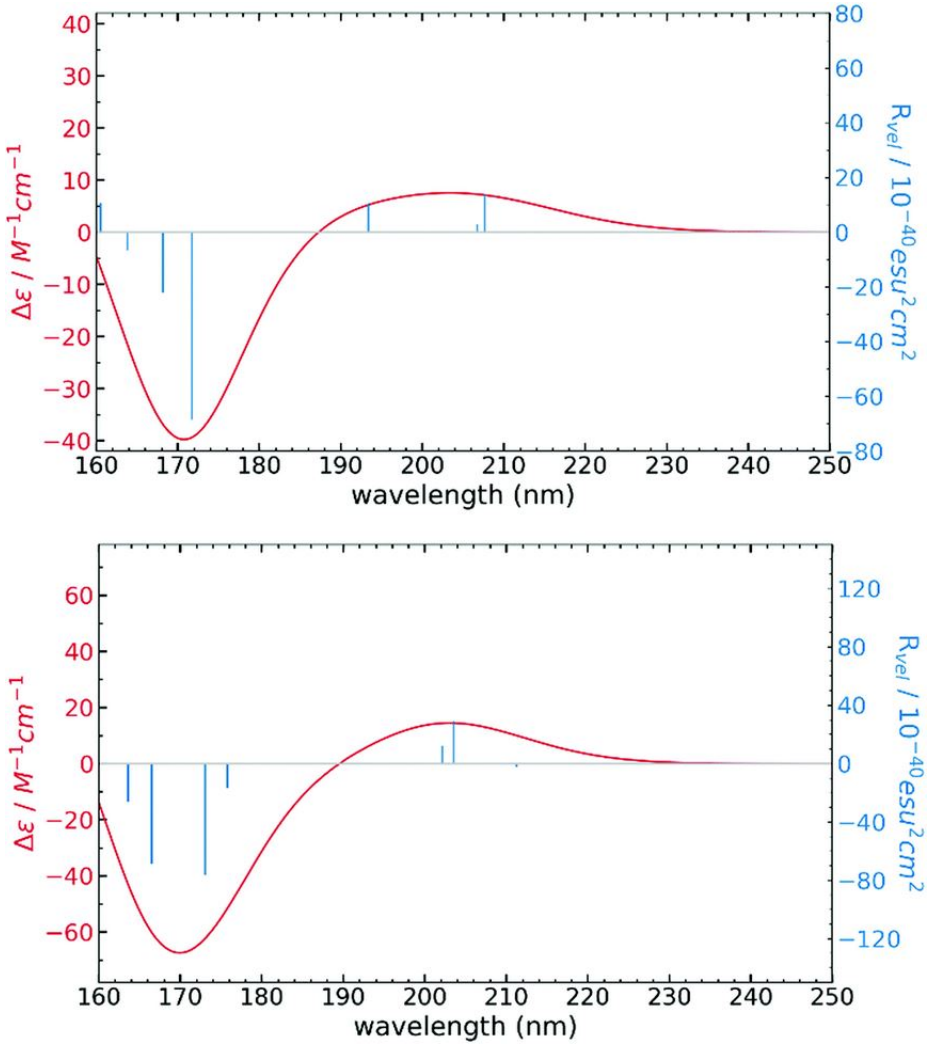


Figure 5.2: ECD spectra of GAG in implicit (upper panel) and explicit water (lower panel) calculated for the optimized pPII conformation,

Fig. 5.2 shows the calculated CD spectra and the underlying electronic transitions of the pPII conformation. The corresponding absorption spectra are displayed in Figs. in Appendix A for GAG in the pPII and β-conformation, respectively. The CD spectra exhibit a negatively-biased positive couplet with a positive maximum at 202 nm and a negative maximum at 172 nm, which resemble the shape of a canonical pPII signal. As one

would expect, the calculated CD maxima and corresponding absorption bands are at higher energies than the corresponding experimental values, from which they appear blueshifted by 14 and 18 nm. The couplets obtained with implicit and explicit water are very similar. It is slightly redshifted from the position of the corresponding absorption band for implicit water, whereas it coincides with the absorption band for explicit water.

Experimentally, the CD spectrum of the β-strand conformation exhibits a negative couplet with a weak and broad negative maximum at ca. 212 nm, and a positive maximum at ca. 190 nm, similar to the difference spectrum shown in Fig. 5.1 [145, 152, 164]. The calculation for the optimized β-strand structure of GAG in implicit water reproduces the experimental spectrum only partially (Fig. 5.3). The spectrum in the region between 220 and 160 nm exhibits local maxima at 202 and 172 nm, respectively. The latter is redshifted relative to the corresponding absorption band. In contrast to this result for implicit water, the CD spectrum of the β-strand conformation in explicit water meets the consistency criterion. It features a positive maximum at 176 nm (slightly redshifted from the absorption band) and a broad negative maximum at 202 nm. The much closer correspondence between the spectrum calculated with explicit water and the experimental spectrum suggest the necessity of performing calculations on peptide–water clusters for a good reproduction of β-strand CD spectra. This result is generally in line with the findings of other researchers16,21 who found that the inclusion of explicit solvent is essential for calculations aimed at reproducing the CD spectra of folded polypeptides.

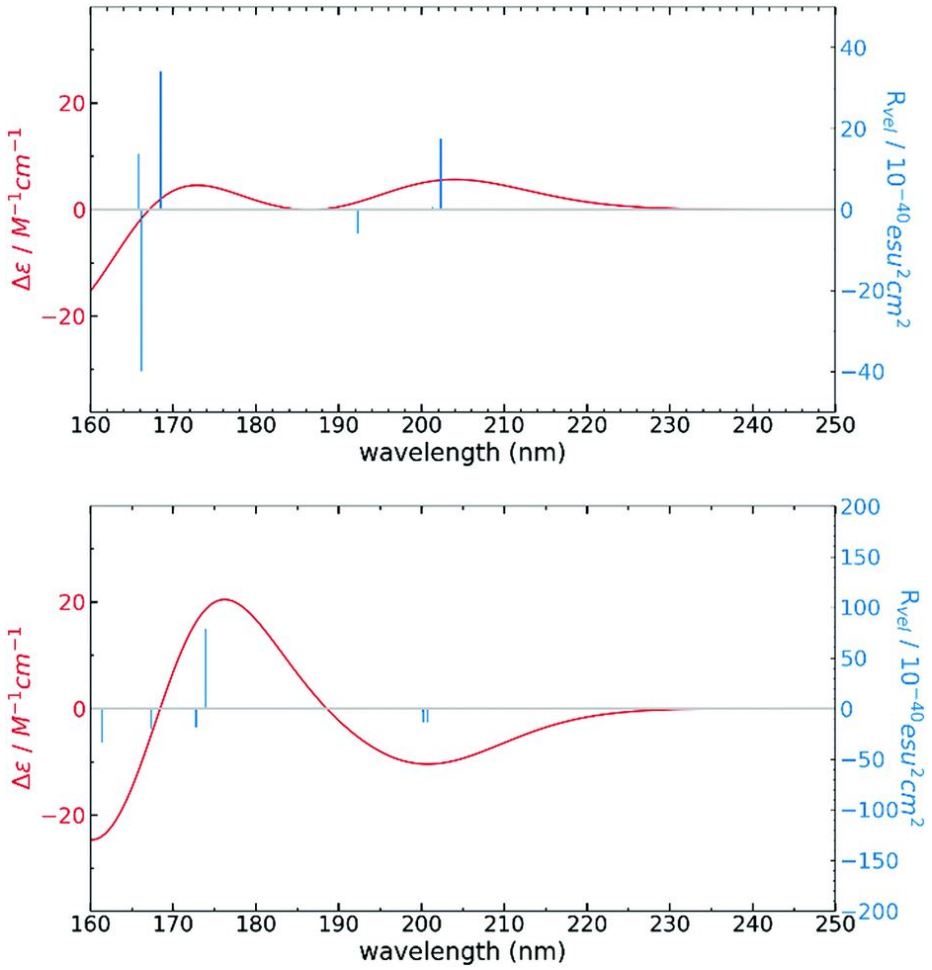


Figure 5.3: ECD spectra of GAG in implicit (upper panel) and explicit water (lower panel) calculated for the optimized β-strand conformation.

It is generally believed that the absorption band at 190 nm results from a $\pi \rightarrow \pi^*$ transition between the HOMO and LUMO, and the corresponding CD spectra reflect excitonic couplings between electronic transitions of different peptide groups [150]. Our calculations reveal a more complex and significantly different picture. The band at 176 nm, which corresponds to the 190 nm band in the experimental spectrum, reflects contributions from multiple electronic transitions, each of which is produced by configurational interactions from electron transfer excitations between occupied and unoccupied molecular

orbitals. The number and oscillator strength of these transitions depend on the peptide conformation as well as on the water model. In the CD spectrum of the explicit water model, the negative maximum of the pPII signal reflects two major and two minor contributions. The corresponding band in the β-strand spectrum is dominated by a single transition. Since the analysis of the excitations involve complex configuration interactions (CI) composition, we employ natural transition orbitals [165] (NTO) to analyse the nature of these excitations. NTOs provide a compact representation of the orbitals and offer a convenient way to visualize excitations via 'hole' and 'particle' descriptors, respectively. Some of the natural transition orbitals involved in these transitions are in part delocalized over both peptide groups and the alanine residue between them. Others are more localized at one of the peptide groups and/or the C-terminal carboxylate group. (Figure in Appendix A). The lowest energy peak is dominated by the second excited state, and visualization of the NTOs implies charge transfer character for this excitation. The hole is localized towards the alanine residue, and the particle is localized towards the carboxylate group. The negative dominant peak involves the fifth and sixth excited states, and NTOs depict the redistribution of electrons away from the carboxylate group. A list of relevant energies with their rotatory strength for the first five excited states are provided in Tables S2 and S3 (ESI) for GAG in the pPII and β conformations, respectively.

The influence of water and geometric conformation on the electronic structure of GAG can be inferred from the Figures in Appendix A which depict the NTOs that contribute most to the relevant excited states. The influence of explicit water and the backbone conformation can be illustrated by a comparison of corresponding NTOs. For

implicit water, the holes of pPII and the β-strand are both mostly localized in the plane of the C-terminal peptide group, while the corresponding particle is localized perpendicular to the plane of the C-terminal peptide group. Hence, the underlying transition clearly has charge-transfer character. For GAG in the pPII conformation with implicit water, the corresponding NTOs contribute around 99% to the transition into the first excited state at around 202 nm. Similarly, the negative maximum at around 170 nm arises from the fourth excited state where the relevant NTOs contribute nearly 93% for this transition. The hole is mostly delocalized across the two peptide bonds while the particle is localized mostly on the N-terminal peptide group. We note that in the negatively-biased positive couplet observed here in the CD spectra, the peaks are dominated by a transition into a single excited state rather than by an admixture of multiple excited-state transitions, as we will see for the case of spectra obtained with explicit water molecules around the GAG in the pPII conformation.

In explicit water, the NTOs of the β-strand conformation are delocalized over both peptide groups and the Cα–Cβ bond of the alanine residue. Both the hole and particle of the pPII conformation are delocalized over the C-terminal peptide and the adjacent carboxylate group for the first and second excited states. For the fifth excitation, which contributes to the negative peak in the CD spectra, the particle is delocalized over both peptide groups while the hole is localized at the alanine residue. In both cases the hole exhibits contributions from water orbitals. The particle, however, involves less water. In the β-strand conformation, water molecules hydrogen-bonded to the peptide groups contribute to the NTO whereas the particle does not exhibit a recognizable contribution from water

molecules. The simulation of the CD spectrum for GAG adopting a β-strand conformation in explicit water yielded a positively-biased negative couplet. The lowest-energy peak arises from transitions into the first and second excited state which are very close in energy. From Fig. in Appendix A, we note that the negative peak at around 202 nm arises from transitions into the first and second excited state that have no significant charge transfer. We do not show particle/hole pairs for the second excited-state transition since they are very similar to those of the first excited-state transition. The positive peak at around 178 nm is dominated by transitions into the fourth excited state. The NTOs show a slight charge-transfer character in which the hole is more localized towards the alanine residue, and the particle is localized at the peptide bond at the C-terminal carboxylate group.

5.3 Conclusion

Taken together, our results show that the electronic structure of accessible excited states are far more complicated than generally envisaged. The degree of NTO delocalization is both structure and solvent dependent. Water is involved in the NTOs of both states, though the distribution depends on the conformation. These results reveal that the classical excitonic coupling models are unsuitable for describing the electronic structure of polypeptides and that the explicit consideration of hydrogen bonded water (not just the effect of hydrogen bonding) is pivotal for understanding the energetics and electronic structure of peptides. Apparently, one does not have to invoke polarization effects due to very high-lying electronic transitions to reproduce the experimental CD spectra of pPII conformations.

The very fact that the UVCD spectra of very short peptides and unfolded polypeptides or proteins look practically indistinguishable strongly suggests that the results obtained for GAG offer some general insight about the electronic structure of unfolded peptides and proteins. UVCD spectra of different GxG peptides exhibit the similar negatively-biased signals, which differ only in terms of the peak intensities of their negative and positive maxima. Some of the peptide spectra do not show the latter. This reflects different pPII/β-strand mixtures [146, 152, 156]. The structure dependence of the delocalization of occupied orbitals invalidates Flory's isolated pair hypothesis, which is in line with experimental results [166, 167]. The huge difference between the corresponding occupied MOs of GAG in implicit and explicit water strongly suggest that solvation energies of different residues are not additive.

One might wonder whether a calculation for only two equilibrium conformations can really account for the observed CD spectra. Since the energy landscape is relatively flat, peptides still sample a significant portion of the Ramachandran plot. For GAG, this is particularly true for the glycine residues, while the conformation space of the alanine residues is rather restricted [147]. However, the similarity of UVCD spectra of GxG peptides with rather different positions of pPII and β-strand-like distributions suggest that the spectra are either pPII or β-like as long as the peptide samples the Ramachandran plot about a ψ-value of 100°. Hence, a consideration of the peptide's dynamics by means of MD/DFT calculations would most likely solely vary the amplitudes of the calculated CD couplet. Any attempt to explicitly account for electronic peptide–water interactions would require that a hydration shell be included in the TDDFT part of the calculations.

Chapter 6

Water-Mediated Electronic Structure of Oligopeptides Probed by Their UV Circular Dichroism, Absorption Spectra, and Time-Dependent DFT Calculations

This work resulted from collaboration with Prof. Reinhard Schweitzer-Stenner's group in the Department of Chemistry at Drexel University. I exclusively performed all the TD-DFT calculations and wrote the manuscript and the supporting information. Prof. Reinhard provided his assistance with the initial drafting of this manuscript. This work is published in *The Journal of Physical Chemistry B*. [2]

6.1 Introduction

Short peptides that contain only a few amino acid residues are generally assumed to be in a so-called random coil state. For individual residues, the generic Ramachandran plot suggests that they sample the entire sterically accessible space, which is considered to be nearly isoenergetic. With the exception of proline and glycine, this space and, thus, the conformational ensemble would be very similar for all amino acid residues [168–175]. In such a case, it makes sense to assume that the random coil conformations can be represented by a single absorption spectrum. However, over the past 15 years, experimental and computational studies have revealed that individual residues differ with regard to the conformational sampling, with the conformational entropy of short peptides and disordered and unfolded proteins being significantly less than assumed [147,148,151,166,176–180]. Recent results from MD and DFT computations have pointed to the same direction [155,181–186]. All residues predominantly sample the upper left quadrant of the Ramachandran plot, and they differ mostly in terms of their population of the polyproline II (pPII) ($\phi > -100^\circ$, $\psi > 100^\circ$) and β-strand region ($\phi \leq -100^\circ$, $\psi > 100^\circ$). The former is stabilized enthalpically while the latter is favored entropically [174].

Besides NMR and vibrational spectroscopies, UV circular dichroism (UV-CD) has emerged as a suitable tool to distinguish between different conformational ensembles of short peptides and unfolded proteins [146,150,187–189]. Owing to the preponderance of the sampling of pPII and β-strand-like conformations [148,190], all spectra exhibit a negative maximum around 190 nm, the intensity of which reflects the pPII content of the distribution. If the pPII content is high, a weak positive maximum also appears in the region at 215 nm.

Generally, the intensities of both maxima decrease with increasing temperature because of the increasing population of β-strand-like conformations [146,189,191]. It is widely assumed that the conformational changes involving a pPII $\leftrightarrow$ β-strand transition lead exclusively to changes in the UV-CD signal due to an excitonic coupling mechanism. However, a recent computational analysis of the electronic transitions underlying the UV-CD spectra of the cationic GAG in explicit water suggests that the electronic structure of pPII and β-strand are different in both the ground and excited states of the peptide [1]. We therefore wondered whether these differences are also reflected by the corresponding far-UV absorption (UV-A) spectra.

While far UV-CD spectroscopy has been the spectroscopic technique of choice for the secondary structure analysis of proteins for over a period of 60 years (if we include its predecessor, optical rotary dispersion spectroscopy), [145,156,192] the corresponding UV-A spectra have been mostly neglected after a period of exploration in the 1950s and 1960s [193–195]. The neglect of this topic has several reasons. The UV-A spectrum of peptide/protein backbones generally displays one broad band with a peak in the region between 190 and 200 nm [196] that overlaps with a broad absorption band of water. Compared with the respective UV-CD spectrum, its structure sensitivity is generally considered as somewhat limited. While UV-A spectra of β-sheets and α-helices differ with regard to their respective molar absorptivities due to the hypochronism of the latter, the spectra of β-sheets and so-called random coil conformations were reported to be similar to each other [196]. As a consequence, they do not appear as suitable tools for the analysis of unfolded peptides and proteins. Amino acid residues with sulfhydryl groups or unsubstituted amide bonds can

contribute to the far UV-A spectra, whereas their contribution to the corresponding UV-CD spectrum does not seem to be significant [195]. Overall, the lack of major fingerprints in the UV-A spectra of peptides makes it a less widely studied tool compared to UV-CD for structure analysis.

To check whether the UV-A spectra of statistical coil-type conformations can still be distinguished, this study compares the temperature dependence of the UV absorption of a series of Gly-x-Gly tripeptides (GxG), where the middle -x- residue is a nonaromatic residue guest in the minimal neighbor environment of glycine residues. To shed some light on the origin of the observed changes, we performed time-dependent density functional theory (TDDFT) calculations of UV-CD and UV-A spectra for a selected number of cationic GxG peptides in explicit and implicit water. Recently, we showed that the explicit consideration of water is necessary to calculate the UV-CD spectrum of GAG in β-strand conformations. Our results revealed rather complicated electronic structures of the investigated peptides, which are inconsistent with the traditional excitonic coupling approach [1]. Herein, we further show that (1) the background (water) corrected UV-A absorption spectra of most of the investigated GxG peptides depend on the temperature below 200 nm, and (2) this temperature dependence depends on the choice of the host residue. Generally, our results shed new light on the charge-transfer character of electronic transitions that give rise to the well-studied UV-CD and UV-A spectra of peptides and proteins. Our results call for a critical assessment of theoretical models that have been used to model respective polypeptide and protein spectra over the past 70 years.

6.2 Methodology

6.2.1 TDDFT Calculations

All quantum mechanical calculations were performed using the Gaussian 09 program package [197]. We optimized the geometry of the peptides after obtaining the coordinates from the Supporting Information reported by Ilawe et al. [198] These calculations were followed by a harmonic frequency analysis to assess whether the stationary points obtained were local minima. After this, we performed excited-state TDDFT calculations to obtain the UV absorption and electronic circular dichroism spectra of the peptides. Previous studies [158–160] have utilized TDDFT calculations to examine the underlying physics of the CD spectra of model peptides and proteins in their folded state. A similar investigation of unfolded peptides was performed in our previous work [1]. This work expands the previous study to include other variants of short peptides. As discussed before, we recently performed DFT calculations on the cationic GAG peptide in implicit and explicit water to obtain the energetics of their pPII and β-strand conformations. The present study considers the optimized geometry of these two conformations of GxG (x = A, I, L, V, and R) peptides for subsequent TDDFT calculations at the ωB97X-D/cc-pVTZ level of theory to calculate the UV-CD and corresponding UV-A spectra of GxG in implicit and explicit water. To add a GxG peptide with a charged side chain to this list, we first optimized the geometry of pPII and β strand conformation of GRG in implicit and explicit water (10 H2O) and subsequently calculated absorption and UV-CD spectra with TDDFT. The water positions in explicit water models were chosen based on a previous study of ours [161], which was based on MD/DFT studies of hydrogen-bonded water molecules [199]. We specifically

chose the ωB97X-D/cc-pVTZ functional for our studies since it contains an asymptotically correct (range-separated) portion of nonlocal exchange in conjunction with dispersion corrections, which are essential for accurately predicting charge-transfer excitations and hydrogen-bonding interactions. It is also important to note that prior work by Neto et al. [163] has shown that the optimal range-separated parameter, ω, required to accurately predict excited states in the explicit solvent is 0.25, which is already close to the default value of $\omega = 0.2$ used in the ωB97X-D/cc-pVTZ functional. We showed in our previous work that (1) changing the range separation value, (2) using different exchange-correlation functionals, (3) utilizing a larger basis set (such as aug-cc-pVTZ), and (4) altering the positions of explicit water molecules did not change the spectra significantly within the relevant wavelength range (see Supporting Information).

To explore how the UV-CD spectra change when more excited-state electronic transitions are included, we also performed additional TDDFT calculations with 50 excited-state transitions. We did not observe any difference between the UV-CD spectra computed with 30 and 50 excited-state transitions in the relevant wavelength range (see Supporting Information). Therefore, we calculated the absorption and UV-CD spectra within the relevant wavelength range at the ωB97X-D/cc-pVTZ level of theory and 30 excited-state transitions for this work. We followed the same method reported in our previous study [1] to obtain UV-CD and absorption spectra. The spectral region covered by our calculation comprises two bands that are generally termed NV_1 and NV_2. Experimentally, NV_1 appears at around 190 nm in peptide and protein spectra and is generally assigned to a HOMO $\rightarrow$ LUMO transition of the peptide π-electron system. NV_2 generally appears at much lower

wavelengths and is thought to be mostly assignable to the C = O bond. As we will show in this paper, the electronic underpinnings of the two bands are more complex. The Supporting Information contains optimized geometries of all GxG peptides, detailed absorption and UV-CD spectra with the underlying transitions, comparisons between spectra of different conformations, lists of NV_1 and NV_2 transition peaks, and tables with a detailed list of the first five excited states.

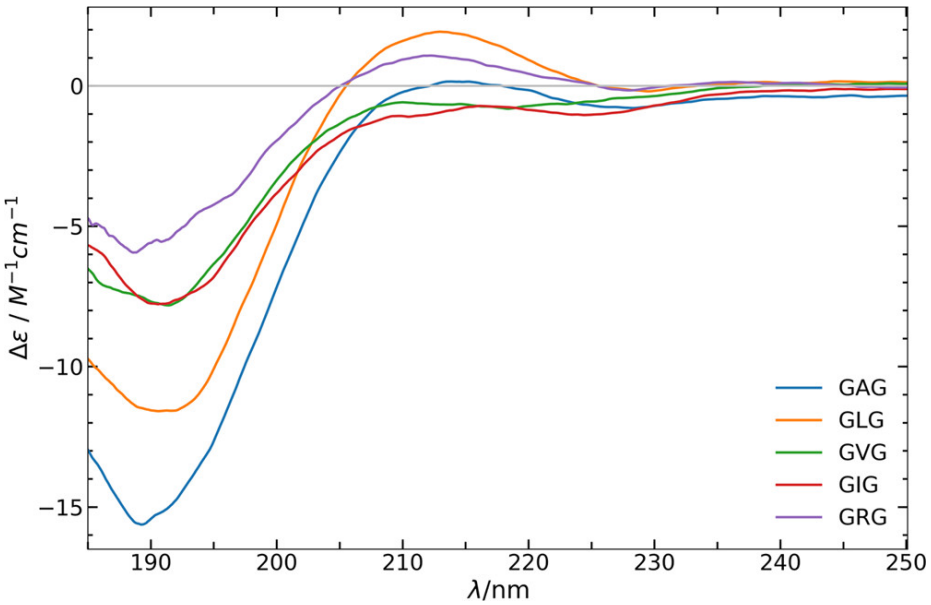


Figure 6.1: UV-CD spectra of cationic GxG (x = A, L, V, I, and R) in water measured at room temperature (20 C). Individual dichroism values of these spectra were reported by Toal et al. [166].

.

6.3 Results and Discussions

6.3.1 TDDFT Calculations of CD and Absorption Spectra

TDDFT calculations were performed for five peptides with side chains, i.e., GAG, GVG, GIG, GLG, and GRG in explicit and implicit water for the respective pPII and β-

strand conformations. The explicit water calculation considered 10 water molecules that are either hydrogen-bonded to the peptide and are inserted in a hydrogen-bonding network of water molecules. The structures for the four aliphatic peptides were recently obtained by geometry optimization [161]. Additionally, we optimized the geometries of cationic GRG for pPII and β-strand conformations in explicit (10 H_2O) and implicit water to explore a situation in which the side-chain transitions can be expected to contribute to the measured absorptivity below 180 nm to a significant extent. We calculated the CD as well as the absorption spectra for these peptides and with both water models. Figs. 6.2 and 6.3 depict the respective spectra obtained with the explicit water model. As usual, the energies of electronic transitions are overestimated by our TDDFT calculations. Hence the wavelength ranges in Figs. 6.2 and 6.3 are all blue-shifted relative to that of the experimental spectra in Fig. 6.1. Experimental limitations rule out measuring any absorption and CD spectra below 185 nm.

The following observations are noteworthy. First, all explicit water calculations yielded the classical feature of pPII, namely negatively biased couplets with a negative maximum at 175 nm and a positive maximum at 210 nm. For β-strand conformations in explicit water, we observed nearly equally intense positive and negative maxima at ca. 185 and 205 nm, respectively. For GLG, the two extrema are very weak. None of the calculations with implicit water reproduce this couplet signal. This underscores the earlier reported notion that explicit water must be considered to account at least quantitatively for the spectra of β-strand conformations [1].

The influence of explicit water and geometric conformation on the electronic structure can be inferred from dissimilar rotational strengths of the NV_1 transition peaks obtained for the β conformations. The implicit water calculation yielded a series of transitions with mostly weak negative rotational strengths. In explicit water, one of these transitions gains a positive rotational strength, giving rise to the positive maximum that is characteristic of the experimental β-strand CD spectra. Differences between the results of the two water models are less pronounced for pPII, where the negative maximum associated with the NV_1 band comprises several transitions with negative rotational strengths.

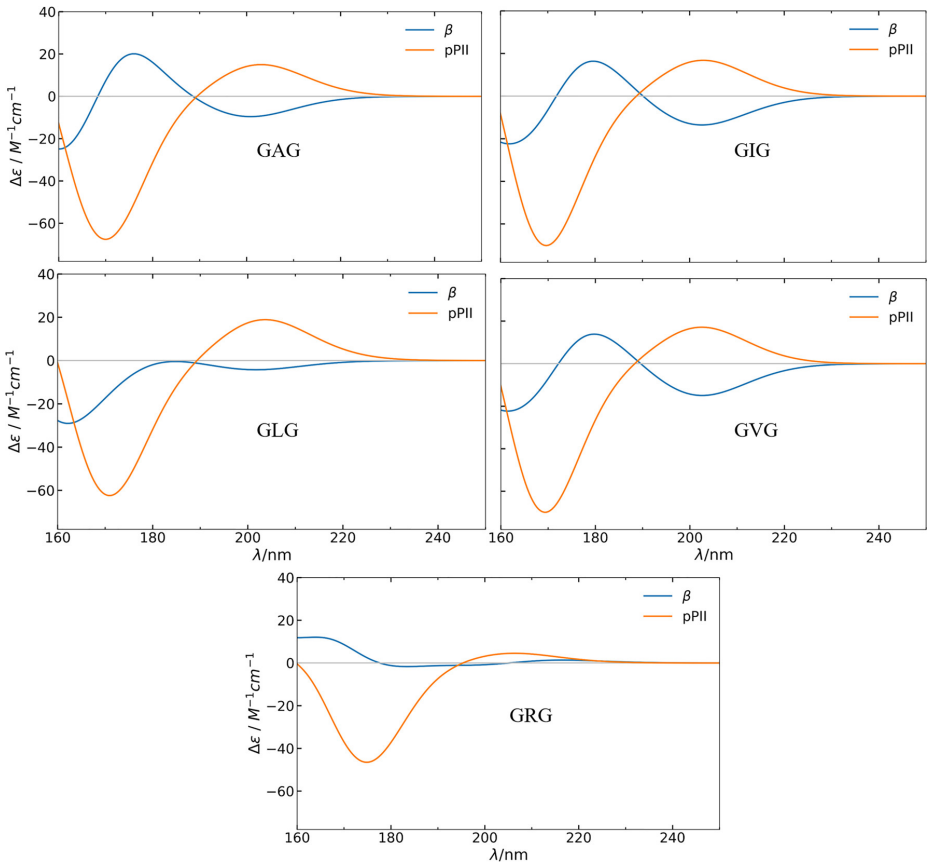


Figure 6.2: Calculated UV-CD spectra of cationic GxG (x = A, I, L, V, and R) peptides in their β and pPII conformation with explicit water.

The calculated absorption spectra (see Fig. 6.3 for explicit water) display two major absorption bands: one around 140 nm and another between 170 and 180 nm. Conventionally, they would be interpreted as indicating the peptide groups' NV_2 and NV_1 transitions from the π-orbital HOMO to the two lowest π^*-orbital LUMOs [194, 195], but as we show below, this assignment would be way too simplistic. However, for the sake of conciseness, we term the corresponding bands NV_1 and NV_2 in the following.

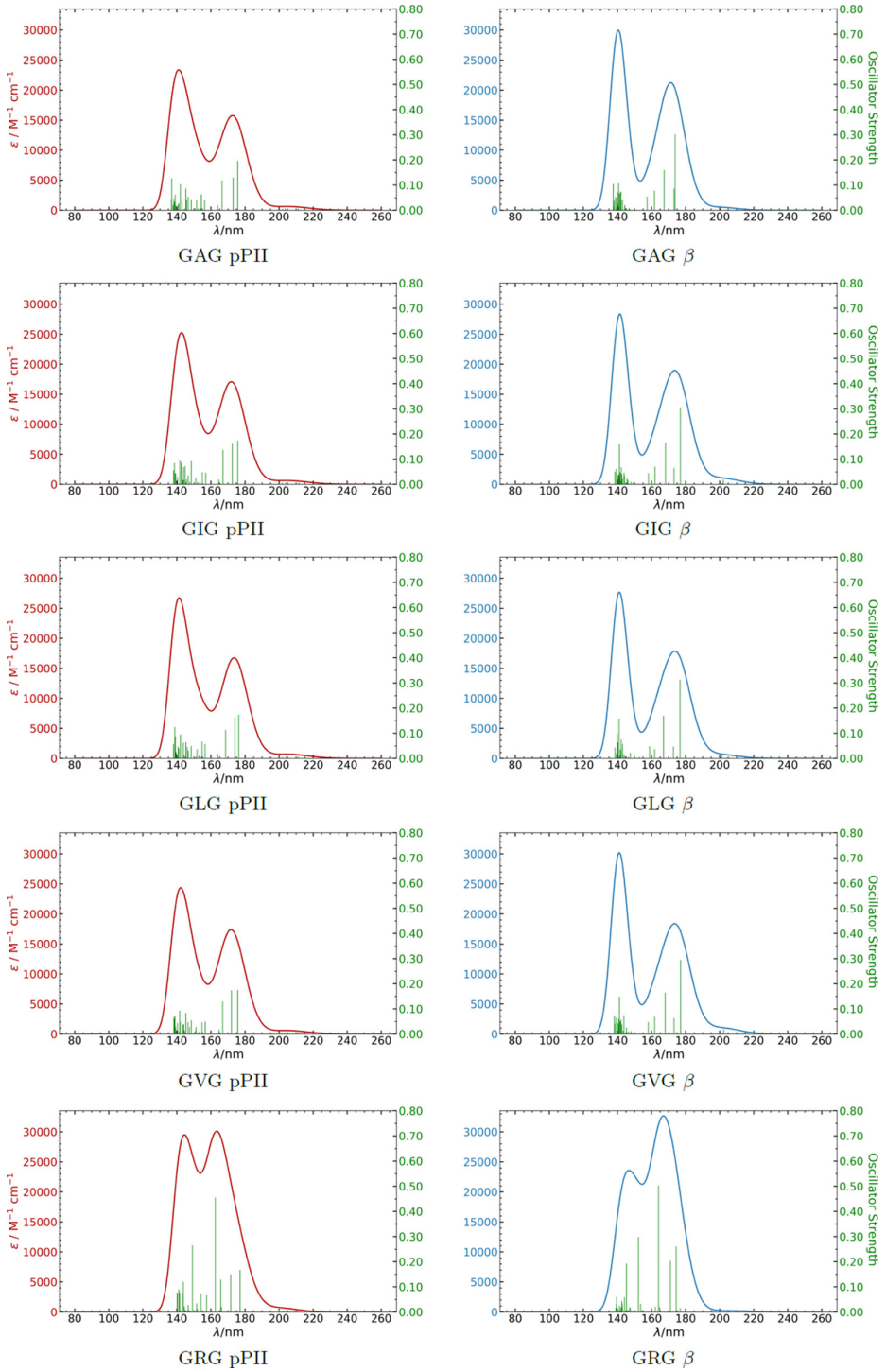


Figure 6.3: Calculated distributions of oscillator strengths associated with bands NV_1 and NV_2 in the pPII and β-strand conformations of the indicated GxG peptides (x = A, I, L, V, and R) obtained with the explicit water model. The sticks represent the underlying electronic transitions.

As can be inferred from Fig. 6.3, the NV_1 bands of β strands exhibit higher oscillator strengths than their respective pPII conformations for the aliphatic peptides GAG, GIG, GLG, and GVG in explicit water. Only for GRG does the pPII intensity of NV_1 exceed that of the corresponding β conformation. It is also noteworthy that the β conformations of GVG, GIG, and GLG exhibit lower transition energies for their NV_1 transition peaks than the corresponding pPII, with the average difference of 1.2 nm.

Figure. 6.3 shows the calculated distributions of oscillator strengths associated with bands NV_1 and NV_2 in the pPII and β-strand conformations of the investigated GxG peptides (x = A, I, L, V, and R) obtained with the explicit water model. The corresponding distributions obtained with the implicit water model are depicted in SI. In the absorption spectra of the β-strand conformations of the aliphatic peptides, the respective transitions into the fourth excited state dominate the NV_1 peak region. For the corresponding pPII conformations, transitions into the fourth and the fifth excited states account for most of the NV_1 absorption peak. The NV_2 band of the aliphatic peptides is just the envelope of a rather large number of electronic transitions (Fig. 6.3). The influence of the water model on the calculation of the NV_2 band is significant. First, the explicit consideration of water molecules leads to an increase of transitions and thus an increase of the overall oscillator strengths. Second, the dispersion of NV_2 transitions is significantly more pronounced in the pPII conformations. As a consequence, the corresponding NV_2 bands appear sharper with higher peak intensities in the β-strand conformations compared with their respective of the pPII conformations.

The calculated absorption spectra of both the β and pPII conformation of GRG in implicit water show a peak at around 160 nm and a shoulder at around 142 nm (Figure S8). The shoulder disappears when the absorption spectra are calculated with the explicit water model, and the absorption spectra show distinct peaks at around 168 nm (162 nm) and 148 nm (144 nm) for the β (pPII) conformation (Fig. 6.3). The NV_1 peaks for both the β and pPII conformation have excited-state transitions with a large oscillator strength (0.50). None of the calculated absorption spectra of the other GxG peptides showed such a large oscillator strength for a transition into any excited states.

6.3.2 Comparison of Experimental and Computational Spectra

In view of the apparent superiority of the explicit water model regarding the calculation of the CD spectra of β-strand conformations, we solely compare experimental and explicit water spectra. The pPII spectra of all aliphatic peptides are very similar, which suggests that the rotational strength of this conformation does not depend very much on the nature of the side chain. This is an important finding for any quantitative analysis of the UV-CD spectra of unfolded peptides. The situation is somewhat different for the β-strand conformation. The corresponding spectra are similar for GAG, GIG, and GVG, while the rotational strength of GLG is much less. These results suggest that the different intensities of the negative maxima in Fig. 6.1 are a good indicator of pPII propensities. Our computational results also explain why the experimental spectra of GAG and GLG are very similar even though the pPII propensity of the latter (0.56) is significantly lower than that of the former (0.79) [147], since a weaker positive maximum of the β-strand leads to a more intense negative maximum of pPII/β mixtures. It should be noted that

the absolute dichroism displayed by the calculated spectra indicate a much larger rotational strength than the experimental ones. However, it has to be noted as well that the respective rotational strengths that emerged from the calculations with the implicit water model are even larger to a significant extent. This observation suggests that the involvement of water molecules decreases the rotational strength, and one could therefore expect that calculations with an even larger number of water molecules would tune the CD signals into the correct order of magnitude.

At first glance, the calculated UV-A spectra do not seem to account for the observed temperature dependence of the NV_1 band. For all four peptides, they indicate a more intense band in the β-strand conformation. The observed temperature dependence suggests the very opposite behavior. However, a closer analysis of the data suggests a different picture. First, for reasons described above, one has to take into account that the NV_2 bands of the β-strand conformation are significantly sharper than those of the corresponding pPII state, which leads to a drop of intensity between the two bands. This is fully consistent with the observed intensity loss at the high-energy site of the NV_1 band. Second, a comparison of the NV_1 oscillator strengths calculated with implicit and explicit water reveals important differences that reflect the pivotal role of hydration water. For GAG, GIG, and GVG, the combined oscillator strength of transitions 4 and 5 of the respective β-strand conformations were calculated to be a factor of 1.4–1.5 larger in the implicit than in the explicit water spectra. The difference between the corresponding pPII spectra is much less pronounced (ca. 1.1). GRG is a special case because of side chain contributions that apparently make the contributions to the β-strand spectrum comparable for both models.

Our results suggest that explicit water reduces the NV_1 oscillator strengths mostly in the β-strand conformation. This effect is very likely to be more pronounced for a peptide with a complete hydration shell. As a consequence, the overall NV_1 oscillator strength of the β-strand would become less than that of the respective pPII conformation, in agreement with our experimental results. Therefore, the temperature dependence of the UV absorption spectra of the investigated peptide reflects the degree by which water affects the electronic transitions in the β-strand. Furthermore, it is safe to infer a concomitant increase in the oscillator strength of NV_2 from our results. At the current level of theory, it is not possible to quantitatively interpret the temperature dependence of the investigated GxG peptides. Only for protonated GDG, available experimental data provide us with an explanation of the very weak temperature dependence of the UV-A (Figure S3). As shown by Rybka et al. [200], the conformational distribution of the central D residue of this peptide at room temperature is dominated by β-strand conformations and various turn-like conformations. Only 20% of the peptides sample pPII. As a consequence, a further increase of the β-strand population should not cause a significant temperature dependence of the UV-A, particularly because the turn-like content of the Ramachandran distribution is nearly temperature independent [184].

6.3.3 NTO Analysis

The contribution of peptide and terminal groups as well as side chains and hydration water of GxG peptides to NV_1 transitions can be obtained by a natural transition orbital (NTO) analysis, which is based on the singular value decomposition of the one-particle transition density matrix. The resulting frontier orbitals can represent any

one-electron property associated with the electronic transition. As such, NTOs provide a compact representation of the orbitals and offer a convenient way to visualize excitations via occupied hole and virtual particles. The NTO analysis was also employed to determine the difference between the electronic structure of pPII and β of investigated GxGs based on the nature of the excitations, which results in the observed absorption and UV-CD spectra. An NTO analysis imparts further impetus in characterizing the dichotomy in the spectra for β and pPII conformations. The influence of water and the backbone conformation on the electronic structure of GxGs can be inferred from Figures 6.4-6.8 for explicit (check the Supporting information for implicit water calculations). These figures depict the NTOs that can be attributed to the dominant transitions. The influence of explicit water and the backbone conformation can be illustrated by a comparison of corresponding NTOs.

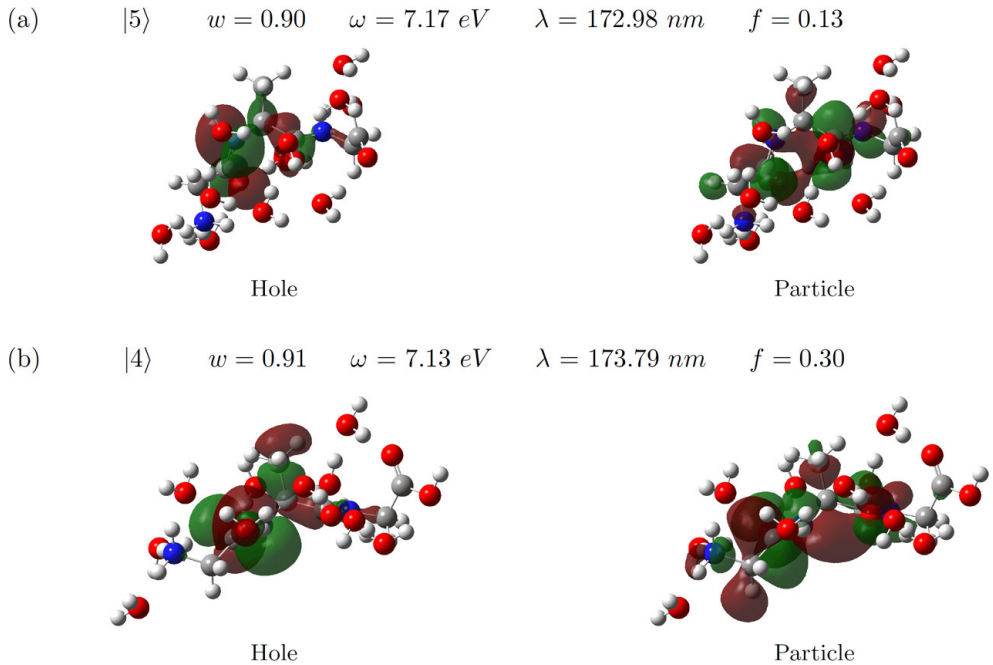


Figure 6.4: Natural transition orbitals (NTOs) of GAG in the (a) pPII and (b) β conformations. The panels give the excited-state number, associated eigenvalue (w), transition energy (ω), wavelength (λ), and oscillator strength (f).

The transition into the fifth excited state gives rise to the NV_1 band obtained for the pPII conformation of GAG (Fig. 6.4). The particle is mostly delocalized across the two peptide bonds, while the hole is delocalized mostly on the N-terminal peptide group. For GAG in the β-strand conformation, the particle associated with the most intense fourth transition is mostly delocalized across the two peptide groups. The corresponding transition of the hole to the particle involves the transfer of some electron density from the alanine residue to the peptide groups. For GIG, the visualized NTOs associated with the NV_1 band of the pPII conformation suggest that particles are mostly localized at the carboxylate group while the hole is predominantly localized at the C-terminal peptide group. In the corresponding β-strand conformation, the particle is delocalized over both peptide groups, whereas the hole is mostly localized at the N-terminal peptide bond (Fig. 6.5). In the case of GLG, transitions into excited states 4 and 5 both contribute to the NV_1 band of the pPII conformation. An analysis of the hole and particle for the fourth excited state shows that the hole is localized at the carboxylate group, while the particle resides at the C-terminated peptide bond. The fifth excitation shows a hole slightly localized at the leucine residue and predominantly at the N-terminated peptide bond, whereas the particle is delocalized across both the peptide bonds. A third transition (number 21) contributes significantly to the NV_1 absorption. Both the hole and particle are delocalized over the peptide backbone. The hole exhibits some minor contributions due to the side chain. For the respective β-strand conformation, the NTO analysis yields a different picture. The particle is delocalized over the ammonium group and the adjacent peptide bond with some admixture from the C_α–C' bond, whereas the hole is localized on the leucine residue. Hence, this transition again has

a charge-transfer character. We wondered whether this was a very peculiar character of the electronic transitions that leads to the very weak CD signal of this conformation. For the pPII conformation of GVG (Fig. 6.7), both the fourth and fifth excitations contribute to the NV_1 band. The hole in the fourth excited state is localized at the carboxylate group, whereas the particle resides at the C-terminal peptide bond group. For the fifth excitation, the hole is mostly localized on the valine residue, and the particle is delocalized across the peptide bonds. Some of the NTOs associated with the NV_1 band of the β-strand conformation of GVG are delocalized over both the peptide group and the valine residue between them. The hole in the fourth excited state is delocalized over both the peptide groups while the particle is localized at the ammonium group.

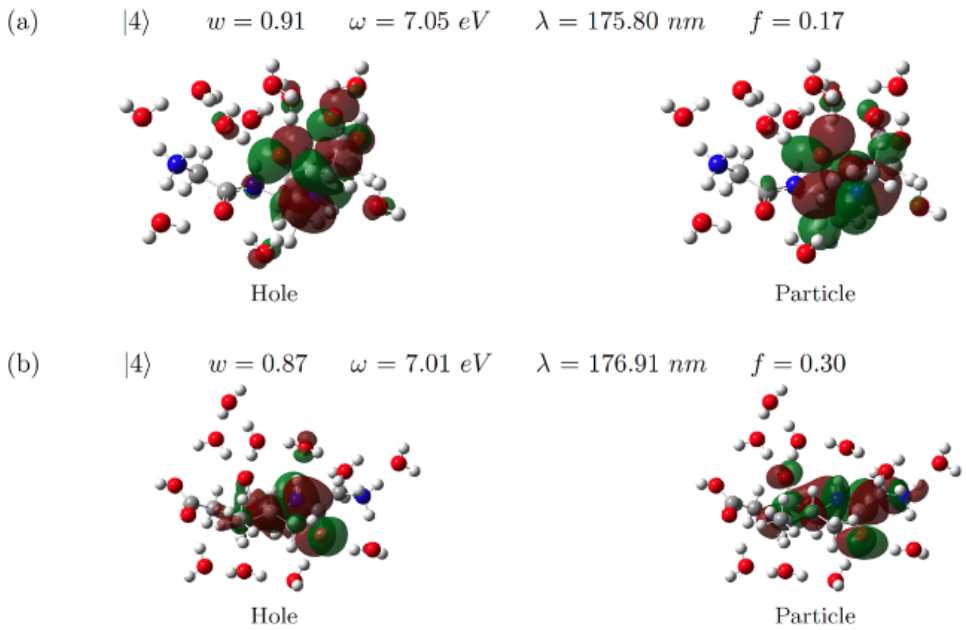


Figure 6.5: Natural transition orbitals (NTOs) of GIG in the (a) pPII and (b) β conformations. The panels give the excited-state number, associated eigenvalue (w), transition energy (ω), wavelength (λ), and oscillator strength (f).

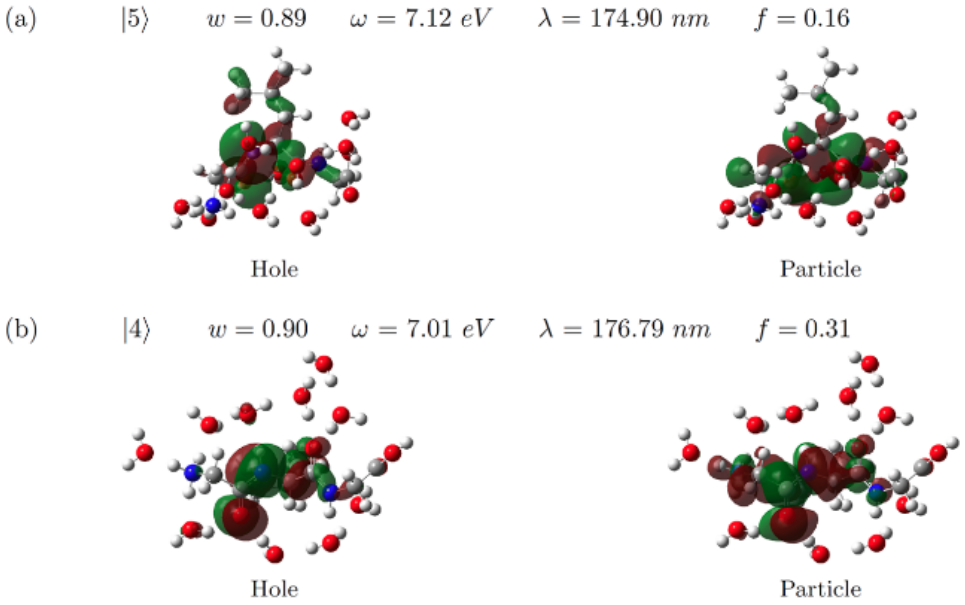


Figure 6.6: Natural transition orbitals (NTOs) of GLG in the (a) pPII and (b) β conformations. The panels give the excited-state number, associated eigenvalue (w), transition energy (ω), wavelength (λ), and oscillator strength (f).

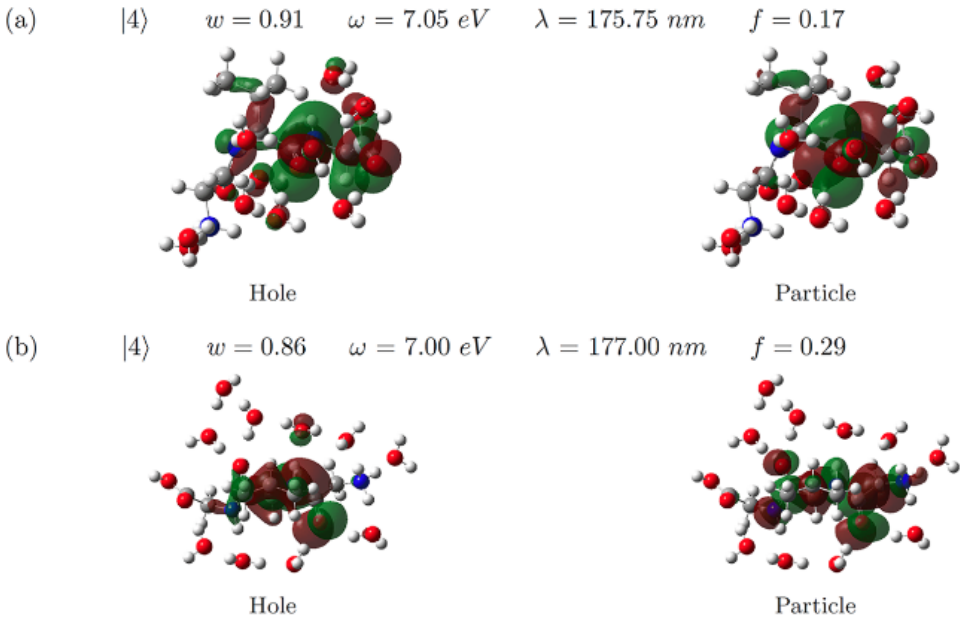


Figure 6.7: Natural transition orbitals (NTOs) of GVG in the (a) pPII and (b) β conformations. The panels give the excited-state number, associated eigenvalue (w), transition energy (ω), wavelength (λ), and oscillator strength (f).

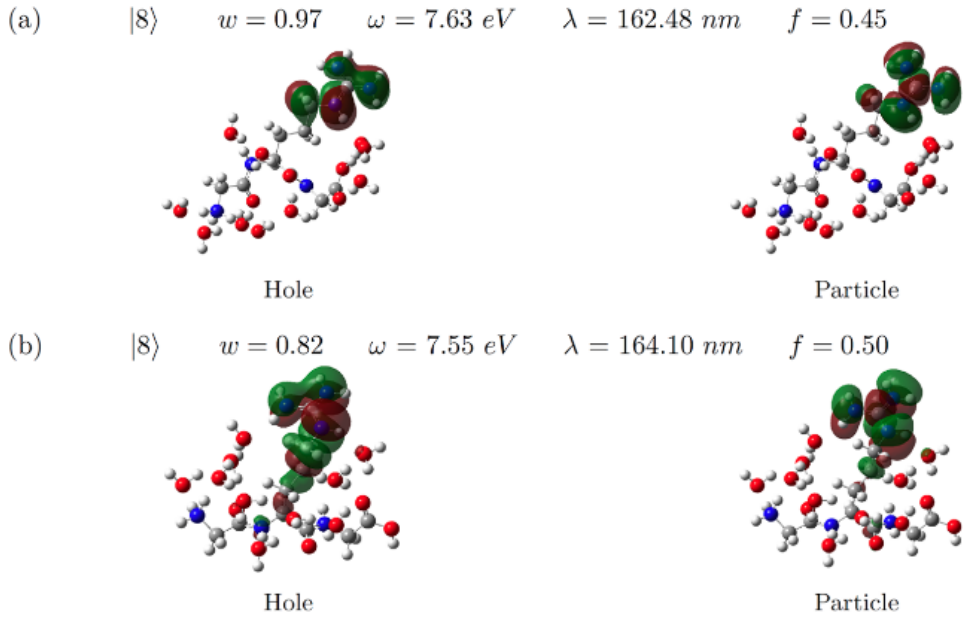


Figure 6.8: Natural transition orbitals (NTOs) of GRG in the (a) pPII and (b) β conformations. The panels give the excited-state number, associated eigenvalue (w), transition energy (ω), wavelength (λ), and oscillator strength (f).

For all the peptides investigated in this study, the NTOs in explicit water exhibit contributions from water molecules. In the pPII conformations of GIG, GLG, and GVG, the particles of the NTOs responsible for NV_1 transition are mostly localized over the carboxylate group. The NTOs of the β conformations are more delocalized over the water molecules than the NTOs of the corresponding pPII conformation. This observation is consistent with the previously discussed influence of explicit water on the oscillator strength of both NV_1 and NV_2.

The NV_1-related NTOs of GRG are peculiar in that the dominant transition of the respective pPII conformation is completely localized on the side chain (both hole and particle). The situation is somewhat different in the β-strand conformation. While transition 8 is still mostly localized on the side chain, transition 19 has some charge transfer

character involving a hole localized over the C-terminal peptide, the C-terminal carboxylate group, and water, while the particle is nearly entirely on the carboxylate group (Fig. 6.8).

In the β conformations of GAG, GIG, GLG, GVG, and GRG, NTOs responsible for the NV_2 transition are mostly localized on the water molecules that are around the peptides. For the corresponding pPII conformations of these peptides, the NTOs are mostly on the peptides' main chain, except for GRG in which the NTOs are mostly on the side chain of the peptide.

The NV_2 transition peaks in the absorption spectra of the pPII conformation for GAG, GIG, GLG, and GVG red-shifts by around 4–5 nm when using an explicit water model. The corresponding NV_2 transition bands of the β conformation red-shifts by around 7–8 nm. This shift in the NV_2 peak suggests again that the absorption spectra of the β conformation of the corresponding peptides—GAG, GIG, GLG, and GVG—are affected more by the explicit water model than their pPII counterparts.

Taken together, the results of our NTO analysis suggest that the electronic structure of the ground and excited state of the investigated peptides changes with its conformation and depends on the character of the side chain. The contribution of water in the hydration shell is substantial for the NV_2 transition and significant only in the more extended β-strand conformation. Quite a few of the dominant electronic transitions have a charge-transfer character; some of them involve the side chains. These findings call into question how the electronic system of polypeptides and proteins have been treated thus far. This issue is briefly discussed below.

6.3.4 Comparison with Literature

Over the past 70 years, a multitude of papers have been published on identifying the underlying physics of the CD and absorption spectra of polypeptides and proteins. In particular, efforts have focused on explaining the spectra of regular secondary structures. Both the oldest and newest attempts are built on the validity of the excitonic coupling model [150, 156, 192, 194, 201, 202], which assumes that the electronic ground state of a polypeptide can be approximately described as a superposition of individual peptide molecular orbitals that are localized on the respective peptide groups. Hence, the respective wave function could be written as a product of wave functions of individual peptide groups. NV_1 excitations, which are generally assigned to a HOMO $\rightarrow$ LUMO transition, cause a delocalization of electronic states. The coupling between transitions has earlier been described by the transition dipole formalism while more recent attempts consider electrostatic interactions between charge distributions [201, 202]. This type of modeling has persisted in the literature because it had a lot of success, particularly in regard to the prediction of the UV-CD spectra of right-handed helical conformations. However, our results suggest that the basic assumptions on which this model is built might be incorrect. First of all, our results suggest that one cannot assign NV_1 to a single transition. Second, we found that the character of the transitions and the resulting excited electronic states are side-chain-dependent. Some transitions might have a charge transfer character where the electron moves from the side chain to the orbitals that are delocalized over peptide groups. The degree of delocalization in the excited state is conformation and peptide dependent. Moreover, the role of water cannot be neglected for extended β-strand conformations.

Such a direct influence of water has not yet been taken into account in the field. Another implication of our results deserves to be emphasized. Currently, most of the force fields used for molecular dynamics simulations are built on the assumption that the torsional force with respect to the dihedral angles ϕ and ψ does not depend on the respective side chain [182, 184, 203–205]. Only recently have attempts been made to consider the specific influence of side chains on the potential function associated with the backbone dihedral angles [206, 207]. The obtained involvement of side-chain MOs in the highest occupied states of peptides and the partially delocalized character of the electronic ground state argues in favor of side-chain-dependent torsional forces. The contribution of water orbitals to the electronic ground state, particularly of the β-strand conformation, suggests the necessity to modify the way how backbone–water interactions are accounted for in molecular dynamic force fields. In this context, the question arises whether force fields with force constants obtained from DFT calculations on model peptides in the gas phase or even in implicit water can sufficiently describe the energetics and dynamics of peptide/protein backbones [183].

It is obvious that a more complete picture than that reported in this paper could be gained by TDDFT calculations with a larger number of hydration water molecules. DFT-based geometry optimizations of tripeptides have recently been achieved with up to 30 water molecules [199, 208]. The corresponding TDDFT calculations would be very computational expansive. Combined MD-TDDFT calculations would only provide an improvement if the hydration shell would become part of the TDDFT part of the calculation.

6.4 Conclusion

The experimental UV-A spectra of some cationic GxG peptides in water indicate that the two dominant conformations, namely polyproline II and β-strand, have different absorption spectra. This notion is supported by TDDFT calculations for a selection of aliphatic peptides with a hydration shell of 10 water molecules. Our calculations also account for the UV-CD spectra of these peptides. The computational results strongly suggest that the two canonical NV_1 and NV_2 bands (the latter is obtainable only with synchrotron radiation) are composed of multiple transitions that result from configurational interactions between numerous transitions. The dispersion of these transitions depends on the backbone conformation and in part on the side chain. Many of the individual transitions have a charge-transfer character. Transitions involving water MOs contribute predominantly to the β-strand conformation. Our results contradict the basic assumption of the excitonic coupling models that are generally used for the theoretical analysis of CD and absorption spectra of polypeptides and even proteins.

Chapter 7

Efficient Predictions of Formation Energies and Convex Hulls from Density Functional Tight Binding Calculations

This chapter's aim is to accelerate the calculation of formation energies and convex hull at different compositions of binary alloys. I exclusively interfaced DFTB with CASM and performed all the CASM+DFT/DFTB calculations. Zulfikhar Ali helped with drafting the manuscript. We contributed equally to the analysis of the convex hull results. This work is published as a full article in the *Journal of Materials Science and Technology.* [4]

7.1 Introduction

Point defects, such as vacancies, play a vital role in the electronic and structural properties of semiconductor materials. These properties ultimately dictate the performance of electronic devices, structural alloys, and functional materials; therefore, a deep understanding of vacancies at the atomistic level can provide a rational path towards their improvement [209–212]. One of the most important quantities to characterize a defect is its formation energy, which can, in principle, be computed with Density Functional Theory (DFT) [213, 214]. Although accurate and widely transferable, DFT can be computationally prohibitive for the routine exploration of the numerous types of defects in material systems. Alternatively, semi-empirical methods, such as Density Functional Tight Binding (DFTB) [215–218], have recently emerged as efficient approaches for addressing these computational bottlenecks. In particular, the DFTB formalism has already been used to calculate a variety of large systems such as metallic nanoparticles [219, 220], explicitly-solvated chromophores [103], massive biological structures [221], molecules/clusters with numerous conformations [222, 223], and immense nanostructures [224].

To enable fast and accurate calculations of formation energies, we combined DFTB with the Clusters Approach to Statistical Mechanics (CASM) [225, 226] software package to predict thermodynamically stable phases of materials for the first time. Specifically, our new capability allows the rapid and accurate calculation of formation energies and the convex hull (if favorable). In short, the convex hull provides a global view of the relative stabilities of structures after the formation energies are calculated. Although computationally demanding, the calculation of formation energies and convex hulls has enabled the discovery

of new materials, including superconducting hydrides [227–231], metal nitrides [232], and metal carbides [233]. Predicting the convex hull for general materials is time-consuming since it requires the classification of a vast number of energy minima on the lattice energy surface. Software programs, such as USPEX [234, 235] and CALYPSO [236, 237] have been used to explore the vast compositional phase space of these materials. Compared to other crystal structure prediction codes, the main advantage of CASM is its ability to evaluate the kinetic and thermodynamic properties of multi-component crystalline solids using group theoretical techniques.

Using this new capability, we calculate the formation energies of various silicon carbide (SiC) and zinc oxide (ZnO) configurations to highlight the efficiency of our DFTB+CASM implementation. SiC is one of the most promising materials for high-temperature, radiation-resistant, power and high-speed electronics [238–241]. ZnO is an affordable, earth-abundant, wide band gap transparent conducting oxide with applications in electronics, optoelectronics, pharmaceuticals, sensors, and catalysis [242]. ZnO crystallizes in many forms, with hexagonal wurtzite (B4), zinc blende (B3), and cubic rocksalt (B1) being the most common. Using SiC and ZnO as representative examples, we compare the accuracy and efficiency of DFTB and DFT for predicting formation energies and convex hulls of these binary compounds.

Our paper is organized as follows: 7.2 gives a brief overview of DFTB, DFT, structure generation algorithm in CASM, formation energies, and convex hulls. 7.3 provides computational details and 7.4 presents our results and discussion. Finally, we conclude with closing remarks and a summary in 7.5.

7.2 Theory and Methodology

7.2.1 DFT and DFTB

Both the DFT and DFTB theory used in the calculations of electronic and structural properties of alloys are covered in chapter 2 and chapter 8, respectively. As such, readers are encouraged to refer to these chapters for a condensed review of these theories. Simulations parameters used for the calculations are present in the Computational Details section 7.3.

7.2.2 Structure Generation with CASM

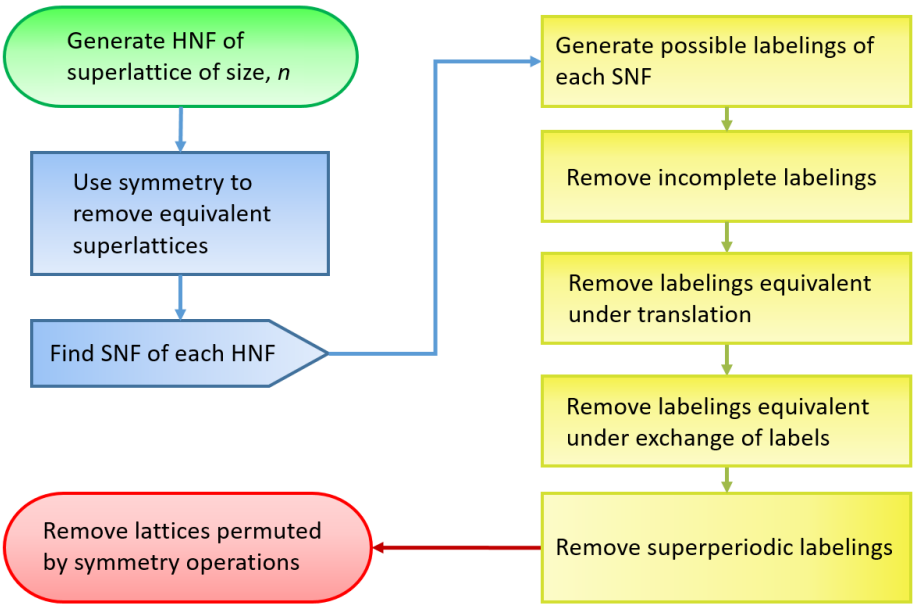


Figure 7.1: Flowchart of the structure generation algorithm used in the CASM software package to enumerate structures.

The algorithms in the CASM software package that enumerate symmetrically distinct configurations utilize an approach based on Hermite Normal Forms of integer matrices [243, 244]. The following is a brief outline of the algorithm [245]:

1. All Hermite normal form (HNF) matrices are generated for each superlattice of size n.

2. Symmetry of the parent lattice is used to remove rotationally equivalent superlattices, thus shrinking the list of HNF matrices.

3. For each index n of the superlattice, the Smith normal form (SNF) is determined for each HNF in the list.

   (a) A list of possible labelings (atomic configurations) is generated for each SNF, which is a list of all k^n numbers in a base k, n-digit system. For the labels, the first k letters of the alphabet, (a, b, $\cdots$) are used.

   (b) Incomplete labelings, where each of the k labels (a, b, $\cdots$) does not appear at least once, are removed.

   (c) Labelings that are equivalent under the translation of the parent lattice vectors are removed. This reduces the list labelings by a factor of $\sim n$.

   (d) Labelings that are equivalent under an exchange of labels, i.e., $a \rightleftharpoons b$, are removed (for example, the labeling *aabbaa* is removed from the list because it is equivalent to *bbaabb*).

   (e) Superperiodic labelings that correspond to a non-primitive superstructure are removed. This can be done without using the geometry of the superlattice.

4. Labelings are removed for each HNF that are permuted by symmetry operations (of the parent lattice) that leave the superlattice fixed.

7.2.3 Formation Energy

The formation energy, e^f, is normalized per primitive unit cell of a particular atomic configuration, σ. For a binary compound AB_x (in this work, atom A is Si/Zn, and atom B is C/O), the formation energy can be calculated with the expression:

$$e^f(\sigma) = e(\sigma) - e^{\text{ref}}(x), \tag{7.1}$$

where $e^f(\sigma)$ is the formation energy of configuration σ, $e(\sigma)$ is the DFT/DFTB total energy (normalized per primitive unit cell) of configuration σ, and $e^{\text{ref}}(x)$ is the DFT/DFTB total energies (normalized per primitive unit cell) of reference state with composition x. The energy of the reference state, $e^{\text{ref}}(x)$, is calculated from the following expression:

$$e^{\text{ref}}(x) = e^{\text{ref}}(x_1) + (x - x_1)\frac{e^{\text{ref}}(x_2) - e^{\text{ref}}(x_1)}{x_2 - x_1}, \tag{7.2}$$

where $e^{\text{ref}}(x_1)$ and $e^{\text{ref}}(x_2)$ are the DFT/DFTB calculated total energies (normalized per primitive unit cell) of the reference states with composition x_1 and x_2, respectively. The composition, x, can be calculated from the expression:

$$x = 1 - (B_n/A_n), \tag{7.3}$$

where B_n and A_n denote the number of C (O) and Si (Zn) atoms in the unit cell, respectively. A value of $x = 1$ implies that the unit cell has only Si (Zn) atoms, while $x = 0$ implies that the unit cell consists of an equal number of Si (Zn) and C (O) atoms. In this study, we chose x_1 and x_2 as 0 and 1, respectively. Setting $x_1 = 0$ and $x_2 = 1$ in Eq. 7.2, the reference state energy, $e^{\text{ref}}(x)$, for a composition x simplifies to:

$$e^{\text{ref}}(x) = e^{\text{ref}}(x_1 = 0) + x\left(e^{\text{ref}}(x_2 = 1) - e^{\text{ref}}(x_1 = 0)\right). \tag{7.4}$$

7.3 Computational Details

The general workflow of our calculations is depicted in Fig. 7.1. We explore thermodynamically stable compositions for binary compounds of the form $A_2B_{2(1-x)}$ using the following four steps: (1) determination of the most stable crystal structure at a fixed composition x using CASM integrated with DFTB, (2) calculation of the formation energy e^f of the compound with respect to its composition at x = 0 and x = 1, (3) repeating the same calculations by changing x, and (4) plotting the formation energy per fixed composition, e^f vs. x. Details of each step are given below.

7.3.1 CASM

To generate the configuration space for the various material compositions, we start with the primitive unit cell of 2H-SiC and B4-ZnO (the 2H prefix denotes a two-layer hexagonal symmetry stacking periodicity). The 2H-SiC unit cell is shown in Fig. 7.3, where the large green atoms represent silicon, and the small gray atoms are carbon. The crystal structure is an AB-type covalent bond crystal, and each Si atom is surrounded by four C atoms. Each 2H-SiC unit cell contains two Si and two C atoms. The space group of 2H-SiC is $P6_3mc$, and the lattice parameters are shown in Table 7.1. B4-ZnO (Wurtzite) has a similar unit cell as 2H-SiC, but is not shown here for brevity. Table 7.1 also lists the lattice parameters of the 2H-SiC and B4-ZnO unit cells used to generate different configurations. Both 2H-SiC and B4-ZnO follow the AB type stacking sequence. We consider a binary ordering between two C/O atoms and a vacancy at the C/O lattice sites in the unit cell. All symmetrically distinct supercells and derivative configurations (up to a supercell vol-

ume that is 4 times the primitive unit cell) were generated using the algorithm developed by Hart et al. [245, 246] in the CASM code. This algorithm enumerates superlattices and atomic configurations in a geometry-independent way using the concept of quotient groups associated with each superlattice to determine all unique atomic configurations. We follow the standard procedure for calculating the formation energy/convex hull [247] using DFT/DFTB with the CASM code.

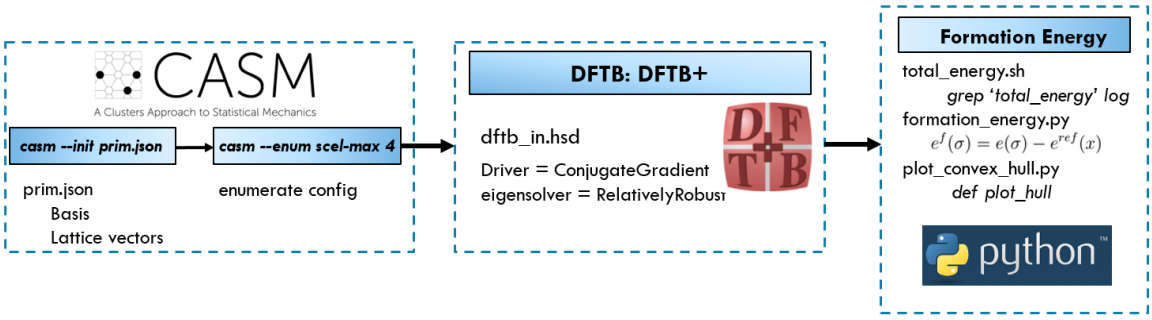


Figure 7.2: General workflow in our DFTB+CASM implementation for calculating formation energies and the convex hull.

After defining the basis and lattice vectors of the system, the symmetrically distinct configurations are generated using the CASM software package. All configurations in symmetrically distinct supercells were generated in 2H and B4 symmetries for SiC and ZnO, respectively. The primitive unit cell, as shown in Fig. 7.3, was used to generate the 401 derivative supercell configurations for each SiC and ZnO system. In the last panel of Fig. 7.2, we calculate the total energies for each configuration via DFT and DFTB.

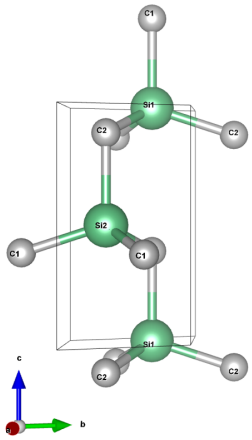


Figure 7.3(a) Lattice Vector $\vec{a}$

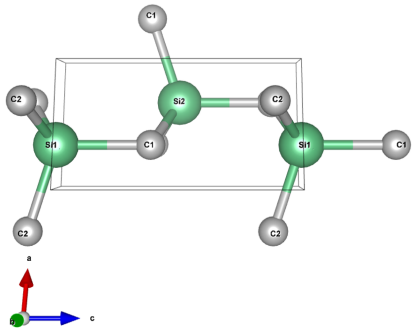


Figure 7.3(b) Lattice Vector $\vec{b}$

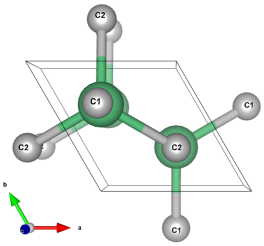


Figure 7.3(c) Lattice Vector $\vec{c}$

Figure 7.3: Primitive unit cells used to generate the various supercells and configurations. Using the CASM software package, we generate all symmetrically distinct configurations with the stoichiometry Si_2C_2. Panels (**a**), (**b**), and (**c**) show the primitive cell in different orientations. Light green and grey atoms represent Si and C, respectively.

7.3.2 DFT Calculations

Density functional theory calculations were carried out with the Vienna Ab Initio Software Package (VASP) code [248, 249]. We used projector augmented wave (PAW) pseudopotentials [250, 251], and the generalized gradient approximation (GGA) exchange-correlation functional, as parameterized by Perdew, Burke, and Ernzerhof (PBE) [252]. We calculated the total energies using a plane-wave energy cutoff of 400 and 520 eV for SiC and ZnO, respectively. All of our calculations used the standard VASP pseudopotentials for all of the atoms. We used a Γ-point centered Monkhorst-Pack k-point mesh (approximately

Table 7.1: Initial lattice parameters of SiC (2H) and ZnO (B4).

Structure	Lattice Vectors (Å)			Lattice Angles		
	a	**b**	**c**	α	β	γ
SiC (2H)	5.54	6.09	6.43	105.50°	89.91°	117.18°
ZnO (B4)	5.56	6.19	6.45	105.21°	89.94°	116.80°

12×12×6) for both SiC and ZnO, and the total energies were found to be suitably converged with this k-point sampling. The k-point sampling differs depending on the unit cell of each configuration, and the CASM software package keeps the mesh density constant for all configurations [253]. In all the optimizations, the geometry was relaxed such that all the forces were less than 0.04 eV/Å. The energy convergence for the electronic degree of freedom was set to 10^{-5} eV. While performing geometry optimizations, the atomic positions, lattice parameters, and angles were allowed to relax for each structure.

7.3.3 DFTB Calculations

As discussed earlier, DFTB is an approximate tight-binding scheme with a low computational cost due to the use of parameterized integrals and a minimal valence basis set [254–260]. In the present study, we used the self-consistent charge formulation of DFTB (SCC-DFTB) in its second-order scheme (DFTB2) [261], which includes the second-order term in the DFT energy expansion around the reference density [261]. Previous studies have shown DFTB to be particularly well suited for describing both SiC and ZnO materials [262–266]. In all of our SiC DFTB calculations, we used a recent DFTB parameterization that accurately reproduces a large dataset of DFT calculations, which includes potential energy surfaces, energies, and forces [267]. We designate the SiC SK files from Ref. [267] as SKfIV

throughout this paper. For our ZnO DFTB calculations, we used standard parameters from the znorg-0-1 SK set [268, 269]. We used a similar k-point mesh as mentioned in the DFT calculations section. In all of our DFTB calculations, the geometry was relaxed with periodic boundary conditions such that all the forces were less than 0.04 eV/Å, and the SCC convergence tolerance was set to 10^{-5} a.u. Both the DFT and DFTB calculations were carried out in a spin-unpolarized formalism. After optimizing the structures, the formation energies and convex hull plots were produced using bash and python scripts.

7.3.4 Formation Energy

As depicted in the last panel of Fig. 7.2, the formation energy for each structure is calculated. The 'total energy.sh' script creates a file that contains the optimized DFT/DFTB ground state energy of each of the configurations. In the next step, the 'formation_energy.py' python script reads the ground-state energies and uses Eqs. 7.2 and 7.1 to compute the formation energy of each of the configurations. Finally, the 'plot_convex_hull.py' python script is used to plot the convex hull from the DFT/DFTB computed formation energies. In the case of SiC, we constructed the convex hull by connecting all the minima of the negative formation energies at various compositions, x, with straight lines.

7.4 Results and Discussion

Using the configuration space described above, we generated distinct compositions up to 4 times the primitive unit cell volume, which produces a total of 401 symmetrically distinct configurations for SiC and ZnO, each. The DFT and DFTB energies of all the enumerated configurations were calculated to obtain the formation energy using Eq. 7.1.

7.4.1 SiC

Convex Hull

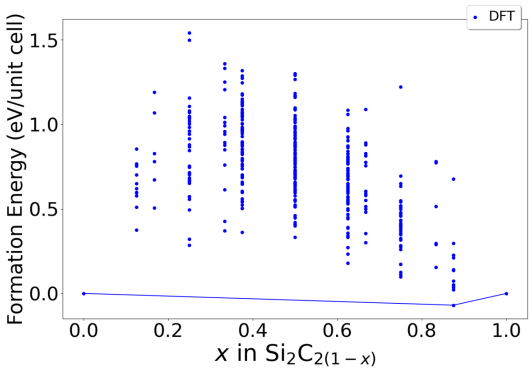


Figure 7.4(a) DFT

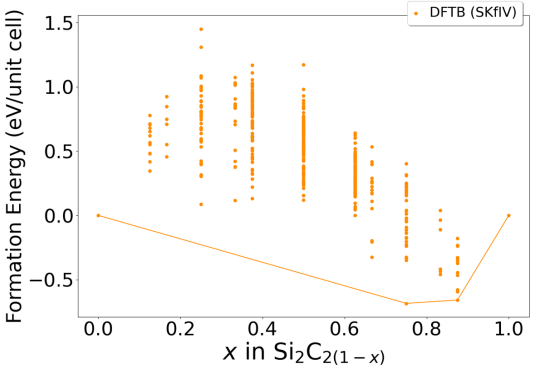


Figure 7.4(b) DFTB (SKfIV)

Figure 7.4: Formation energy convex hull for the Si-C binary system computed with DFT and DFTB. Panel (**a**) shows the DFT formation energy. Panel (**b**) shows the DFTB formation energy calculated using the SKfIV SK files. Each point corresponds to a different crystal structure.

Figs. 7.4(a) and (b) show the formation energies of various SiC configurations calculated with DFT and DFTB, respectively. Although the DFT and DFTB calculations find different minima on the convex hull plot, the results match qualitatively. DFTB predicts a minimum at composition $x = 0.75$ while DFT has a minimum nearby at $x = 0.875$. The two minima structures in the DFTB convex hull plot at $x = 0.75$ and $x = 0.875$

have a formation energy difference of 0.025 eV. This small energy difference is due to the approximations in the SK parameters inherent to the DFTB formalism.

The most probable reason for this small discrepancy in the formation energy is the parameterization of the repulsive DFTB potential between the Si and C atoms in the SKfIV SK files. The repulsive potentials in the SKfIV SK files result in a much stronger repulsion between the Si and C atoms, increasing the bond length between the atoms in the structures, which alters their energetics.

A similar phenomenon was also observed for in a previous study using TiO_2 DFTB SK files [270]. Nevertheless, our results show that the DFT and DFTB formation energies for SiC are similar; if more accurate results are desired, configurations near the convex hull could be first down-selected via DFTB and subsequently refined/re-calculated with DFT to improve their accuracy (which would be more efficient than computing all 401 structures with DFT alone).

Structure Comparison

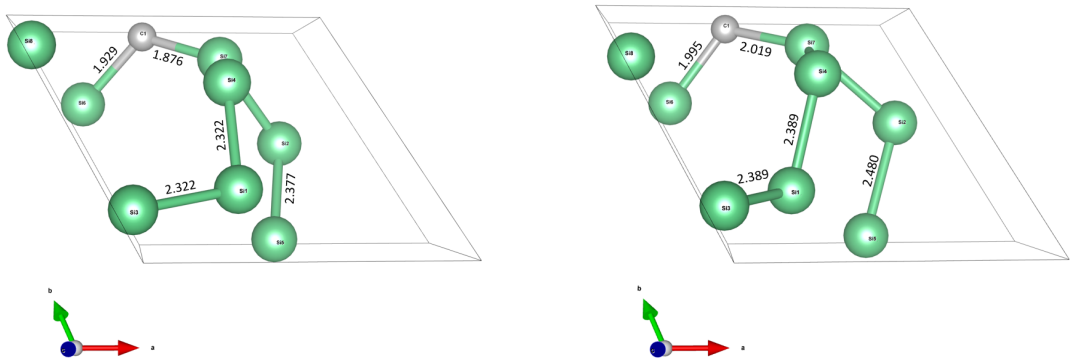


Figure 7.5(a) DFT

Figure 7.5(b) DFTB (SKfIV)

Figure 7.5: Comparison of SiC structural parameters (located on the convex hull) after optimization with (**a**) DFT and (**b**) DFTB (SKfIV), visualized along lattice vector $\vec{c}$. Bond lengths are in Angstroms, and light green and grey atoms represent Si and C, respectively.

The most stable crystal structure obtained via DFT and DFTB is depicted in Fig. 7.5. DFTB predicts a minimum at x = 0.75 while DFT gives a minimum nearby at x = 0.875. The two minima structures in the DFTB convex hull plot at x = 0.75 and x = 0.875 have a formation energy difference of 0.025 eV. As discussed in the previous section, this discrepancy in the minima for the DFTB formation energy results from approximations in the repulsive potential between the Si and C atoms in the SKfIV SK files. Moreover, the crystal structures at x = 0.875 obtained via DFT and DFTB show the same P1 symmetry, which consists of eight Si atoms and one C atom in the unit cell (see Fig. 7.5). Since both the DFT and DFTB calculations correctly predict the same crystal structure and relative ratio of Si/C at x = 0.875, our results show that DFTB can be employed as an efficient computational approach tool for calculating and pre-screening formation energies.

Fig. 7.5 compares the DFT and DFTB optimized structural parameters of the most stable configurations at the convex hull minimum located at x = 0.875. Two types

Table 7.2: Comparison of optimized lattice parameters of the minima structure calculated with DFT and DFTB for SiC.

Method	**Lattice Vectors (Å)**			**Lattice Angles**		
	a	**b**	**c**	α	β	γ
DFT	6.15	6.15	6.16	121.39°	97.24°	110.62°
DFTB (SKfIV)	6.27	6.27	6.40	119.38°	99.60°	109.24°

of Si atoms exist in the unit cell: one that is bonded to only Si atoms and another which is bonded with one C atom. Each Si and C atom has a coordination number of 4. As can be seen in Fig. 7.5(a), the Si-C bond length ranges from 1.93 to 2.38 Å in the DFT-optimized structure. Fig. 7.5(b) shows the bond lengths between various Si and C atoms of the DFTB-optimized structure. DFTB predicts slightly longer bond lengths for almost all the Si-C bonds, which on average are longer by 0.06 Å compared to DFT calculations. Table 7.2 compares the optimized lattice parameters of the minimum structure calculated with DFT and DFTB. DFTB overestimates the optimized lattice parameters and predicts slightly longer lengths for $\vec{a}$, $\vec{b}$, and $\vec{c}$.

As stated earlier, the discrepancy in the structural parameters is due to the parameterization of the repulsive DFTB potential between the Si and C atoms in the SKfIV SK files. The repulsive potentials used in the SKfIV SK files result in a much stronger repulsion between the Si and C atoms which increases the bond length between the atoms in the structures. Previous work has also shown that DFTB predicts longer lattice parameters for systems containing C atoms [271].

7.4.2 ZnO

Formation Energy

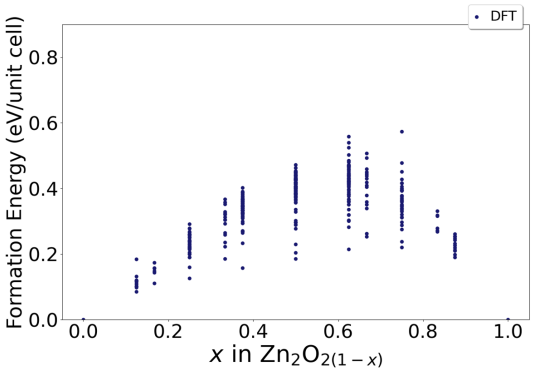


Figure 7.6(a) DFT

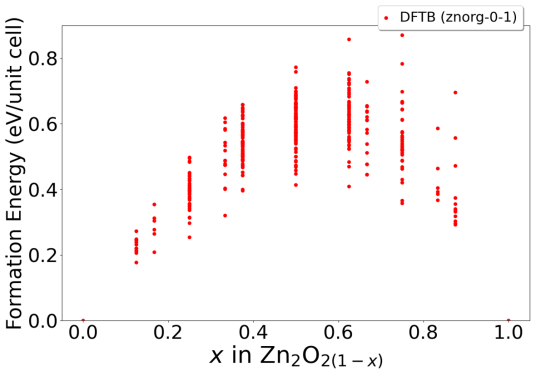


Figure 7.6(b) DFTB (znorg-0-1)

Figure 7.6: Formation energy of the ZnO binary system obtained from DFT and DFTB. Panel (**a**) shows the DFT formation energy, and panel (**b**) shows the DFTB formation energy calculated using the znorg-0-1 SK files. Each point corresponds to a different crystal structure.

We now proceed to ZnO, which is an even more complex material but shows more accurate results between DFT/DFTB. The wurtzite (B4) structure is the most stable form of ZnO at ambient conditions in nature [272]. The zinc blende (B3) structure of ZnO has a less stable cohesive energy than the B4 structure, and is, therefore, energetically unfavorable at zero temperature and pressure [273]. As such, we enumerated various ZnO configurations starting with the B4-ZnO unit cell. We obtain a significant performance improvement for ZnO structures optimized via DFTB (more details on efficiency are discussed in the next section). It is interesting to note that both the DFT and DFTB calculations do not find a convex hull in the formation energy plots of the $Zn_2O_{2(1-x)}$ binary compounds. As shown in Fig. 7.6, the calculated formation energies are all positive, showing no stable minima structure predicted at any composition x. A previous study reported similar findings of

positive formation energies when oxygen vacancies were introduced in the ZnO lattice [274]. Specifically, these previous studies showed that defects often induce occupied states in the bandgap and increase the formation energy [275–278].

7.4.3 Efficiency Analyses for DFT and DFTB

In this section, we give a detailed analysis of the computational timings and efficiency of DFT and DFTB. Fig. 7.7 compares the wall time per SCF iteration step (WT-SCF) as a function of the number of electrons in various ZnO configurations. As the number of electrons increases, the WT-SCF increases rapidly for both DFT and DFTB. For example, the DFT WT-SCF for a 30-electron ZnO configuration is around 16 seconds, whereas a 120-electron configuration is roughly 134 secs. Fitting the DFT WT-SCF data to a cubic polynomial gives a high R^2 correlation coefficient of 0.92, indicating an $O(N^3)$ scaling, where N is the number of electrons. This scaling can be attributed to matrix diagonalization in KS DFT, which is an $O(N^3)$ process, where N is the size of the matrix. Similarly, fitting the DFTB WT-SCF data gives an $O(N)$ linear scaling with an R^2 of 0.91. For all configurations, the DFTB WT-SCF is less than 6 secs, which is significantly faster than DFT. As can be seen from Fig. 7.7, the DFTB WT-SCF for a ZnO configuration with 138 electrons is an order of magnitude faster than DFT. It is worth mentioning that although the DFTB WT-SCF is more efficient than DFT, DFTB geometry optimizations may take more SCF cycles compared to DFT, which may result in a small loss in efficiency.

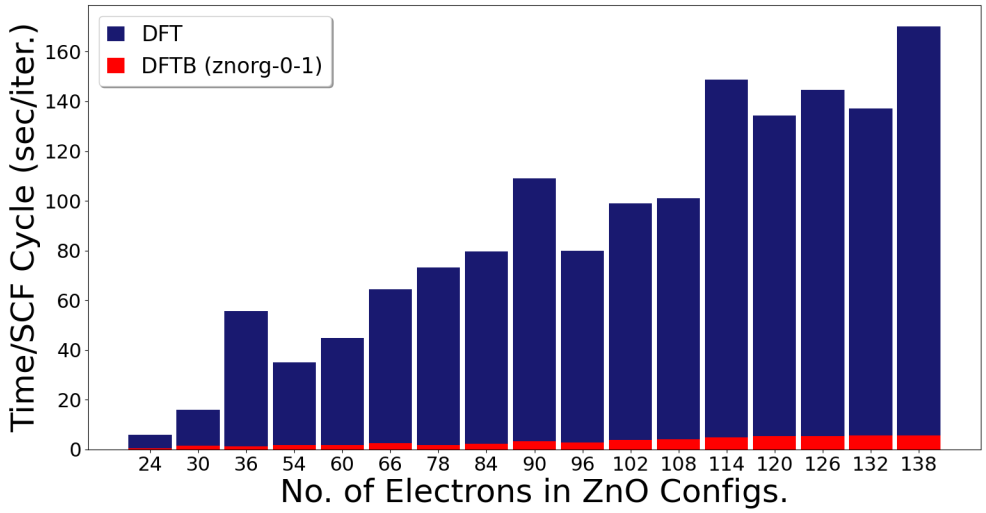


Figure 7.7: Comparison of DFT and DFTB wall times per SCF iteration step as a function of the number of electrons in various ZnO configurations. Each vertical bar value was calculated by averaging multiple configurations having the same number of electrons.

Computational Timings for SiC

Figs. 7.8(a) and (b) compare computational timings for geometry optimizations of various $Si_2C_{2(1-x)}$ structures using DFT (VASP) and DFTB (DFTB+). In general, the DFTB calculations take significantly less time compared to DFT, and Fig. 7.8(b) shows that DFTB can be an order of magnitude faster than DFT in some cases. For most configurations, the DFTB calculations required more SCF cycles for geometry optimization compared to DFT; however, the total compute time for DFTB is still significantly smaller.

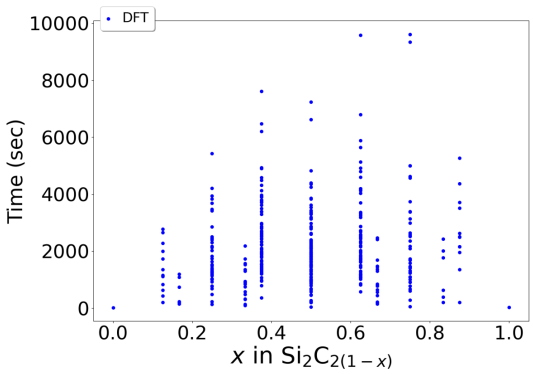


Figure 7.8(a) DFT

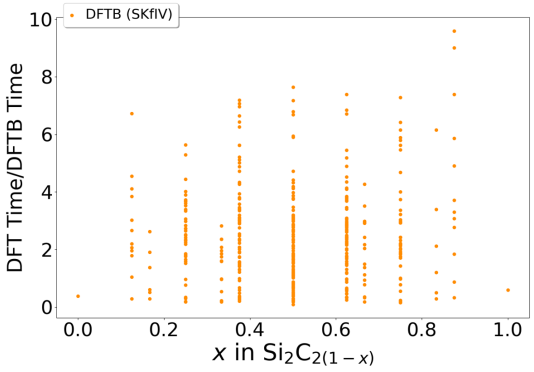


Figure 7.8(b) DFTB (SKfIV)

Figure 7.8: Comparison of wall times for geometry optimization of various $Si_2C_{2(1-x)}$ compositions calculated via DFT and DFTB. Panel (**a**) shows the wall time for optimizing each configuration using DFT. Panel (**b**) shows the ratio between the DFT and DFTB wall times for a geometry optimization of each configuration. Each point corresponds to a different crystal structure.

Computational Timings for ZnO

Fig. 7.9 compares wall times for geometry optimization for various $Zn_2O_{2(1-x)}$ binary compounds calculated via DFT and DFTB. As in the case of SiC, Fig. 7.9(b) shows that DFTB is an order of magnitude faster than DFT in most cases. It is important to note that DFTB is almost 40 times faster than DFT for $Zn_2O_{2(1-x)}$ binary compounds (see Fig. 7.9(b)), whereas the maximum performance enhancement of DFTB is only 10 times for $Si_2C_{2(1-x)}$. In general, as the number of electrons in the system increases, we show that DFTB exhibits more performance gains than conventional KS DFT calculations.

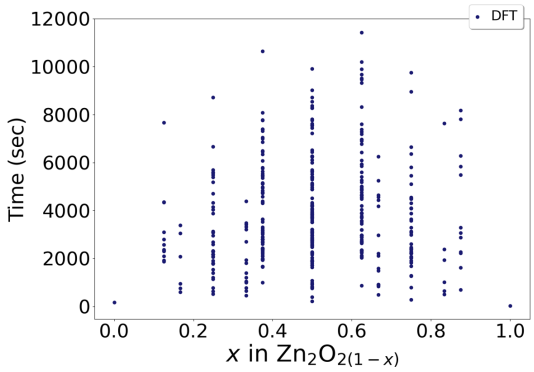


Figure 7.9(a) DFT

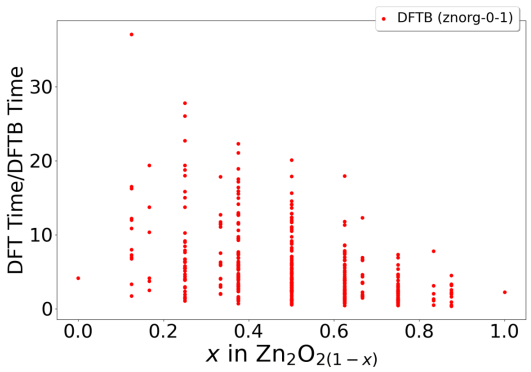


Figure 7.9(b) DFTB (znorg-0-1)

Figure 7.9: Comparison of wall times for geometry optimization of various $Zn_2O_{2(1-x)}$ compositions calculated via DFT and DFTB. Panel (**a**) shows the wall time for optimizing each configuration using DFT. Panel (**b**) shows the ratio between the DFT and DFTB wall times for a geometry optimization of each configuration. Each point corresponds to a different crystal structure.

Finally, Fig. 7.10 compares the total time (sum of individual wall times) for the geometry optimization of $Si_2C_{2(1-x)}$ and $Zn_2O_{2(1-x)}$ binary compounds. For $Si_2C_{2(1-x)}$, the total time for geometry optimization of all 401 configurations is around five days for DFTB, whereas the DFT calculations require nearly ten days. The difference in performance between DFTB and DFT for $Zn_2O_{2(1-x)}$ is even more significant - DFTB geometry optimizations take approximately five days, whereas the DFT calculations take nearly 18 days (more than 3 times longer than DFTB). As mentioned previously, larger systems such as ZnO (which contain more electrons), scale more favorably with DFTB, resulting in speedups that can be orders of magnitude faster than conventional DFT methods.

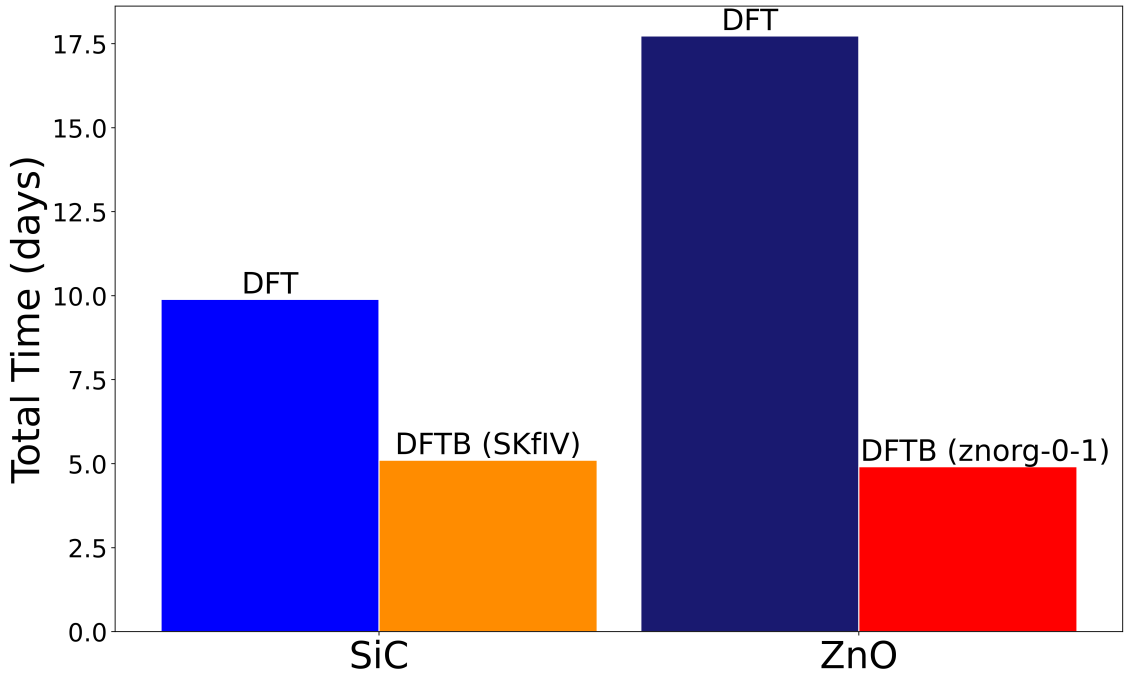


Figure 7.10: Comparison of DFT and DFTB total wall times for geometry optimization of all SiC and ZnO configurations. The DFTB calculations used the SKfIV and znorg-0-1 SK files for SiC and ZnO, respectively.

7.5 Conclusion

In closing, we have interfaced DFTB with the CASM software package for the first time to enable routine and efficient calculations of formation energies and convex hulls. Our extensive calculations show that DFTB can be used as an efficient screening tool to compute the numerous formation energies (and convex hull if it exists) of complex materials. To highlight the efficiency and accuracy of our approach, we calculated and compared the formation energies of SiC and ZnO with both DFT and DFTB. We find that the DFTB approach enables extremely efficient calculations of formation energies in a completely unbiased manner to predict low-lying metastable phases over the entire composition space. By

comparing the convex hull/formation energy from both approaches, we found that DFTB gives similar trends as the DFT calculations. Finally, we performed an extensive benchmark of the computational timings for both DFT and DFTB and found that the DFTB calculations can be an order of magnitude faster (larger systems give even higher computational efficiency). Moreover, we show that DFTB gives accurate results and can be a computationally cheaper alternative to DFT. In summary, our DFTB+CASM implementation allows for an efficient exploration (up to an order of magnitude faster than DFT) of formation energies and convex hulls, which researchers can use to routinely probe other complex systems.

Chapter 8

GPU-Enhanced DFTB Metadynamics for Efficiently Predicting Free Energies of Biochemical Systems

This chapter describes the free energy surfaces (FES) of alanine dipeptide and remdesivir calculated using DFTB and classical force field. We perform large-scale metadynamics calculations on the Azure cloud to evaluate FES. To overcome the limitation associated with the classical approach in modeling biochemical systems, we use the quantum-based DFTB approach. Employing GPU acceleration, we showed that the GPU-DFTB can be used as a highly efficient and effective (more accurate than Amber) method in capturing the quantum interactions at the atomic level. We collaborated with Dr. Giulia Palermo's

group at the University of California, Riverside, to perform the classical Amber calculations of remdesivir. Prof. Giulia's group wrote the Amber calculations section (8.3.1) and performed the classical calculations for remdesivir. This work is published as an article in the journal *Molecules*, where the complete research is provided. [5]

8.1 Introduction

Molecular dynamics (MD) simulations are used to study a wide range of dynamic atomistic effects, including free energetics of chemical processes [279–281], protein folding [282,283], self-assembly [284], nucleation [285–287], glass formation [288,289], and chemical dynamics in solutions at interfaces [290]. The relevant physical processes in these studies are often rare events where a property of interest occurs on a time scale not accessible via simulation (within a reasonable amount of time) due to the presence of a large energy barrier separating local minima along the free-energy landscape. This well-recognized limitation of MD has led to the development of metadynamics approaches [291, 292] to enhance the sampling of free-energy states and the rare events that allow the crossing of very high free-energy barriers [293]. Metadynamics is often applied in conjunction with classical molecular dynamics, where the atomistic interactions are approximated by classical force fields that are predetermined functions of the atomic coordinates.

When coupled with metadynamics, classical force field simulations of large systems can be used to estimate the structure and thermodynamics of relatively complex chemicals and materials. However, classical force fields can be inaccurate [281] and fail to capture the quantum interactions at the electronic level. For example, chemical reactions in which bonds

are broken/formed cannot be directly simulated using the most common force fields [294, 295]. Additionally, force fields are fitted to experimental data under specific conditions, which makes their transferability to other situations challenging [296,297]. To remedy these issues, *ab initio* metadynamics generated from density functional theory (DFT) calculations can be used to accurately capture bond breaking and formation in various chemical dynamics processes. In DFT-based metadynamics, interatomic forces are computed on the fly [298, 299], leading to more computationally demanding calculations than classical metadynamics simulations. Moreover, the enormous computational cost associated with the DFT sampling of free energies restricts its applicability to relatively small chemical systems (less than 20 atoms) [300]. A promising alternative is the use of semiempirical methods such as density functional tight binding (DFTB), which can serve as a bridge between (efficient but inaccurate) MD and (costly but accurate) DFT calculations. In previous studies, our group and others have used DFTB calculations to gain computational speedups of up to 2–3 orders of magnitude compared with those of conventional DFT calculations [103, 219–221, 224, 260, 301–303].

In a previous work [221], we developed a massively parallelized heterogeneous CPU+GPU approach for carrying out large-scale DFTB MD simulations (2 ps) of an entire explicitly solvated protein (HIV protease) for the first time. Building on our experience with GPU-enhanced DFTB simulations of large biochemical systems, we now apply these techniques to long-term metadynamics simulations (10 ns). Because the computational bottleneck in metadynamics simulations is the diagonalization of the Hamiltonian matrix (which is performed several times during a single molecular dynamics trajectory) [221],

many of our GPU-acceleration techniques can be harnessed for these calculations. To further accelerate our metadynamics calculations, we used massively parallelized cloud computing, which has recently emerged as a new computational platform for running large, complex electronic structure calculations. We first validate our approach by calculating the free-energy surface of alanine dipeptide (ADP), a chemical system typically used as a reference standard in the scientific literature for benchmarking metadynamics algorithms. Our GPU-based DFTB calculations are compared against the results obtained from classical force fields and hybrid DFT (PBE0) methods (the latter is the most accurate benchmark of ADP to date). Finally, to further extend our GPU-enhanced DFTB approach, we also carried out a 10 ns metadynamics simulation of remdesivir, which is prohibitively out of reach for routine DFT-based metadynamics calculations. Based on our benchmark tests, analyses, and extensions to large biochemical systems, we highlight the use of our GPU-based DFTB approach for accurately and efficiently predicting the free-energy surfaces/thermodynamics of large biochemical systems.

8.2 Theory and Methodology

8.2.1 DFTB Formalism

We briefly discuss the DFTB formalism in this section since it is used extensively to calculate free-energy surfaces/thermodynamics of biochemical systems in this study. Specifically, we used the third-order expansion of the Kohn–Sham (KS) DFT energy around a reference density, which is commonly referred to as DFTB3. To derive the DFTB3 total energy, the DFT total energy expression [254] is chosen as the starting point, which is given by [260]

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + E_{\text{ext}} + E_{\text{H}} + E_{\text{nn}} + E_{\text{xc}}[\rho(\mathbf{r})], \tag{8.1}$$

where T is the kinetic energy of the electrons, E_{ext} is the electron–nuclei interaction energy, E_{H} is the mean-field (Hartree) energy, E_{nn} is the interaction energy of the nuclei, and E_{xc} is the exchange-correlation (XC) energy.

To obtain the expression of the DFTB3 total energy, a Taylor series expansion of Equation (8.1) around the reference density, $\rho^0(\mathbf{r})$, is carried out up to third order in the density fluctuations, $\delta\rho(\mathbf{r})$. The reference density is constructed as a superposition of atomic densities $\rho^0_{\text{A}}(\mathbf{r})$ on neutral atom A, i.e., $\rho^0(\mathbf{r}) = \sum_{\text{A}} \rho^0_{\text{A}}(\mathbf{r})$. Substituting $\rho(\mathbf{r}) = \rho^0(\mathbf{r}) + \delta\rho(\mathbf{r})$ into Equation (8.1) and invoking a minimal basis set with a monopole expansion (among other approximations) [301], we obtain the DFTB3 total energy given as

$$\begin{aligned} E_{\text{DFTB3}} &= \sum_{i}^{\text{occ}} \langle \psi_i | \, \hat{H}^0 \, | \psi_i \rangle + \frac{1}{2} \sum_{\text{AB}}^{\text{M}} \Delta q_{\text{A}} \Delta q_{\text{B}} \gamma^h_{\text{AB}} + \frac{1}{3} \sum_{\text{AB}}^{\text{M}} \Delta q^2_{\text{A}} \Delta q_{\text{B}} \Gamma_{\text{AB}} + \frac{1}{2} \sum_{\text{AB}}^{\text{M}} V^{\text{AB}}_{\text{rep}} \\ &= E_{\text{BS}} + E_{\gamma} + E_{\Gamma} + E_{\text{rep}}. \end{aligned} \tag{8.2}$$

The second, third, and fourth summations in Equation (8.2) run over the number of atoms,

M, in the system. The first term, E_{BS}, in Equation (8.2) is a sum over occupied orbital energies and corresponds to the band-structure energy. It can be obtained from the diagonalization of the non-self-consistent DFTB Hamiltonian $\hat{H}^0$, whose matrix elements are given by [304]:

$$H^0_{\mu\nu} = \begin{cases} \epsilon_\mu^{\text{free atom}}, & \text{if } \mu = \nu \\ \langle \phi_\mu | \, \hat{T} + \nu_{\text{eff}}[\rho_0^{\mathrm{A}} + \rho_0^{\mathrm{B}}] \, | \phi_\nu \rangle, & \text{if } \mu \in \mathrm{A},\ \nu \in \mathrm{B},\ A \neq B \\ 0, & \text{if } A = B,\ \mu \neq \nu \end{cases} \tag{8.3}$$

where ϕ_μ and ϕ_ν form a minimal Slater-type atomic basis, with μ and ν representing the indices of the valence atomic basis function associated with atoms A and B, respectively. In Equation (8.3), $\hat{T}$ is the kinetic energy operator, ρ_0^{I} is the reference density of neutral atom I, and ν_{eff} is an effective Kohn–Sham potential. To obtain $H^0_{\mu\nu}$, we first calculate ϕ_μ, ϕ_ν, and $\epsilon_\mu^{\text{free atom}}$ by solving a modified Kohn–Sham equation given by [305]:

$$\left[-\frac{1}{2} \nabla^2 + V^{\text{eff}} \right] \phi_{\mu/\nu}(\mathbf{r}) = \epsilon_{\mu/\nu} \phi_{\mu/\nu}(\mathbf{r}), \tag{8.4}$$

where V^{eff} is the pseudoatomic potential, which includes the confinement potential [305].

Based on the form of Equation (8.3), only two-center elements are treated within the DFTB framework, which are explicitly calculated using analytical functions. Specifically, the Hamiltonian and overlap matrix elements are stored in Slater–Koster (SK) files for all pairs of chemical elements as a function of the distance between atomic pairs. As such, no explicit integral evaluation occurs during the simulation, which significantly improves the computational efficiency of the DFTB method [221, 306]. The second term in Equation (8.2), E_γ, accounts for the charge fluctuation contributions to the energy, where

γ^{h}_{AB} describes the effective on-site electron–electron interaction [305]. The third term, E_{Γ}, captures the changes in chemical hardness with respect to atomic charge, which improves the description of localized charges [261, 307]. The last term, E_{rep}, is a sum of pairwise repulsive functions, which are obtained by fitting to the DFT calculations of reference structures/molecules [256]. Similar to the Hamiltonian and overlap matrix elements, E_{rep} is pre-tabulated and stored in SK parameter files. By applying the variational principle, we obtain the Kohn–Sham equations [301]:

$$\sum_{\nu}^{\mathrm{M}} c_{\nu i} \left(H_{\mu\nu} - \epsilon_i S_{\mu\nu} \right) = 0, \quad \nu \in B \text{ and } \forall A,\ \mu \in A,\ i \tag{8.5}$$

$$S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle, \quad \forall \mu \in A,\ \nu \in B. \tag{8.6}$$

The DFTB Hamiltonian, $H_{\mu\nu}$, in Equation (8.5) is given by:

$$\begin{aligned} H_{\mu\nu} &= \langle \phi_\mu | \, \hat{H}_0 \, | \phi_\nu \rangle \\ &+ S_{\mu\nu} \sum_{\xi}^{\mathrm{M}} \Delta q_\xi \left(\frac{1}{2} \left(\gamma_{A\xi} + \gamma_{B\xi} \right) + \frac{1}{3} \left(\Delta q_{\mathrm{A}} \Gamma_{\mathrm{A}\xi} + \Delta q_{\mathrm{B}} \Gamma_{\mathrm{B}\xi} \right) + \frac{\Delta q_\xi}{6} \left(\Gamma_{\xi\mathrm{A}} + \Gamma_{\xi\mathrm{B}} \right) \right), \end{aligned} \tag{8.7}$$

where $\mu \in$ A, $\nu \in$ B, and $S_{\mu\nu}$ are the overlap matrix of the atomic orbitals; $\Delta q_{\mathrm{A/B}} = q_{\mathrm{A/B}} - q^{0}_{\mathrm{A/B}}$ is the net charge of atom A/B. The summation in the second term of Equation (8.7) is performed over the number of atoms, M, in the system, and γ_{AB} is an analytical function of the interatomic distance. Because the atomic charges depend on the one-particle wave functions, ϕ_i, Equation (8.5) must be iteratively solved by repeatedly diagonalizing the updated Hamiltonian until self-consistency is reached. This particular step is typically performed numerous times during a DFTB-MD simulation and accounts for 90–95% of the total wall time [308]. To overcome this computational bottleneck, we previously implemented a GPU-enabled eigensolver [221] to efficiently diagonalize the Hamiltonian in Equation (8.5).

8.2.2 Hamiltonian Diagonalization

As discussed in the previous section, the primary bottleneck in DFTB-based MD simulations is the diagonalization of the Hamiltonian matrix in Equation (8.7), which is typically performed numerous times along an MD trajectory [221]. The Hamiltonian diagonalization can be classified as a generalized symmetric definite eigenvalue problem of the form:

$$\mathbf{A} \cdot \mathbf{x} = \lambda \mathbf{B} \cdot \mathbf{x}, \tag{8.8}$$

where $\mathbf{A}$ and $\mathbf{B}$ are real and symmetric matrices, respectively; $\mathbf{B}$ is positive definitive; λ is the eigenvalue; and $\mathbf{x}$ is the eigenvector. Applying a Cholesky factorization on matrix $\mathbf{B}$ ($\mathbf{B} = \mathbf{L} \cdot \mathbf{L}^{\mathrm{T}}$, where $\mathbf{L}$ is a lower triangular matrix), Equation (8.8) can easily be reduced to a standard symmetric eigenvalue problem ($\mathbf{C} \cdot \mathbf{y} = \lambda \mathbf{y}$, where $\mathbf{C} = \mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-\mathrm{T}}$and $\mathbf{y} = \mathbf{L}^{\mathrm{T}}\mathbf{x}$), which facilitates Hamiltonian diagonalization. Standard diagonalization routines can then be employed to solve the standard symmetric eigenvalue problem to obtain the eigenvalues and eigenvectors. In our previous study, we implemented GPU enhancements for the QR, Divide-And-Conquer, and RelativelyRoubust diagonalization routines [309] in an older version of the DFTB+ code [221]. In the DFTB v19.1 code [308], only the Divide-And-Conquer eigensolver routine is enhanced with GPU parallelization via the MAGMA library [310]. Since this particular routine is extensively used during the metadynamics simulations in our study (via Hamiltonian diagonalization, which occurs numerous times in each MD trajectory), we briefly review this routine in the following section.

8.2.3 Divide-And-Conquer

The Divide-And-Conquer eigensolver is based on recursively breaking down a problem into two or more sub-problems, which are subsequently solved to obtain a solution to the original problem [221]. This algorithm takes advantage of deflation [311], which occurs when an eigenpair of a submatrix of a tridiagonal matrix is an eigenpair of a larger matrix. After Equation (8.8) is reduced to a standard symmetric eigenvalue problem of the form $\mathbf{C} \cdot \mathbf{y} = \lambda \mathbf{y}$, the matrix $\mathbf{C}$ is reduced to a block-tridiagonal matrix, $\mathbf{T}$:

$$\mathbf{T} = \begin{pmatrix} & & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \mathbf{T_1} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & \beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \beta & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \mathbf{T_2} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \end{pmatrix}. \tag{8.9}$$

The Divide-And-Conquer approach uses the fact that a block-tridiagonal matrix is very close to a block-diagonal matrix [311], $\tilde{\mathbf{T}}$, having the following form:

$$\tilde{\mathbf{T}} = \begin{pmatrix} & & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \tilde{\mathbf{T}}_\mathbf{1} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \tilde{\mathbf{T}}_\mathbf{2} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \end{pmatrix}. \tag{8.10}$$

Because of its block-diagonal form, the eigenvalues and eigenvectors of the full matrix $\tilde{\mathbf{T}}$ can be obtained from diagonalizing $\tilde{\mathbf{T}}_\mathbf{1}$ and $\tilde{\mathbf{T}}_\mathbf{2}$; as such, solving these two smaller problems is almost always faster than solving the original problem. First, the block-tridiagonal matrix, $\mathbf{T}$, is written as a block diagonal matrix, $\tilde{\mathbf{T}}$, plus a correction, $\mathbf{C}$, i.e.,

$$\mathbf{T} = \tilde{\mathbf{T}} + \mathbf{C} \tag{8.11}$$

$$\mathbf{T} = \begin{pmatrix} & & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \tilde{\mathbf{T}}_\mathbf{1} & & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & & & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & \tilde{\mathbf{T}}_\mathbf{2} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \beta & \beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \beta & \beta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{8.12}$$

The eigenvalues and eigenvectors of $\tilde{\mathbf{T}}_\mathbf{1}$ and $\tilde{\mathbf{T}}_\mathbf{2}$ are subsequently calculated by recursively calling the Divide-And-Conquer algorithm. In the last step, the eigenvalues and eigenvectors of the original matrix $\mathbf{T}$ are built.

The Divide-And-Conquer algorithm scales as $\mathcal{O}(n^3)$ [312], where n is the matrix dimension. The steps used in the Divide-And-Conquer eigensolver are summarized in the following algorithm flowchart. [311]

Algorithm 1: The tridiagonal Divide-And-Conquer algorithm.

For a real symmetric tridiagonal matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$, this algorithm computes the spectral decomposition of $\mathbf{T} = \mathrm{Q}\Lambda\mathrm{Q}^T$, where Λ is a diagonal matrix of eigenvalues and Q is an orthogonal matrix.

if $\mathbf{T}$ is 1×1 **then**
  return $\Lambda = \mathbf{T}$ and $\mathrm{Q} = 1$

else
  Partition $\mathbf{T} = \left[\begin{array}{c|c} \mathrm{T}_1 & 0 \\ \hline 0 & \mathrm{T}_2 \end{array}\right] + \rho\mathbf{u}\mathbf{u}^T$
  Use T_1 to obtain Q_1 and Λ_1 as output
  Use T_2 to obtain Q_2 and Λ_2 as output
  Construct $\mathrm{D} + \rho\mathbf{v}\mathbf{v}^T$ from Λ_1, Λ_2, Q_1, and Q_2
  Find the eigenvalues, Λ, and eigenvectors, Q', of $\mathrm{D} + \rho\mathbf{v}\mathbf{v}^T$
  Construct $\mathrm{Q} = \begin{bmatrix} \mathrm{Q}_1 & 0 \\ 0 & \mathrm{Q}_2 \end{bmatrix} \cdot \mathrm{Q}'$, which are the eigenvectors of T
  return Λ and Q

8.2.4 Metadynamics

Metadynamics is an accelerated sampling method that can be used to explore the free-energy landscape of a system as a function of collective variables (CVs) [291]. Within this formalism, a history-dependent bias potential composed of Gaussian functions is added to the Hamiltonian of the system. These external potentials "fill" the underlying free-energy basins, thus enabling an efficient exploration of the free-energy landscape. In well-tempered metadynamics (WT-MetaD) simulations, the Gaussian height is decreased during

the simulation, which avoids overfilling the free-energy basins and ensures convergence of the final bias potential to the actual free energy (within a constant) [292, 313]. As such, WT-MetaD simulations address the convergence problems associated with conventional metadynamics and allow the exploration of physically relevant regions of conformational space [314]. The WT-MetaD bias potential $V_{\mathrm{B}}(\mathbf{s}, t)$ at time t is constructed from the sum of Gaussian "hills" [292]:

$$V_{\mathrm{B}}(\mathbf{s}, t) = \sum_{t'=\tau,2\tau,\ldots}^{t'<t} W \exp\left[\frac{-\beta V_{\mathrm{B}}(\mathbf{s}, t')}{\gamma}\right] \exp\left[-\sum_i \frac{(s_i - s_i(t'))^2}{2\sigma_i^2}\right], \tag{8.13}$$

where W is the initial height of the bias potential, τ is the time between deposited Gaussians, $\beta = (k_B T)^{-1}$ (where k_B and T are the Boltzmann constant and temperature, respectively), γ is the bias factor, and σ_i is the width of the Gaussians for the ith CV in the set, $\mathbf{s}$, of collective variables. The first exponential term in Equation (8.13) decreases the height of the deposited Gaussians where previous bias potentials had been added. This reduction in the Gaussian height reduces the error and avoids exploring high free-energy states that are thermodynamically irrelevant [315]. The bias factor γ determines the rate at which the magnitude of the newly added potential decreases; a lower bias factor leads to a faster decrease in the bias potential. The last exponential is a product of Gaussians in the direction of the i^{th} CV with width σ_i and centered at the CV value at time t'. Using this approach, the system's dynamics are enhanced, and different conformations are explored by adding an extra force (potential) to the system. The additional bias force for the i^{th} atom is given by [315]:

$$\mathbf{F}_i^{\mathrm{B}}(t) = \left.\frac{\partial V_{\mathrm{B}}(\mathbf{s}, t)}{\partial \mathbf{s}}\right|_{\mathbf{s}=\mathbf{s}(t)} \left.\frac{\partial \mathbf{s}(\mathbf{r})}{\partial \mathbf{r}_i}\right|_{\mathbf{r}=\mathbf{r}(t)}. \tag{8.14}$$

In Equation (8.14), V_{B} is the bias potential, $\mathbf{s}$ is a set of collective variables, $\mathbf{r}$ contains the position vector of all atoms, and $\mathbf{r}_i$ is the position vector of the i^{th} atom. If the conformational space is sampled for a sufficiently long simulation time, the free-energy landscape over CV($F(\mathbf{s})$) is obtained from the bias potential using the following expression [292]:

$$\lim_{t\to\infty} V_{\mathrm{B}}(\mathbf{s},t) = -\frac{(\gamma-1)}{\gamma}F(\mathbf{s}). \tag{8.15}$$

8.3 Computational Details

The free energy, potential energy, and entropy landscapes of ADP and remdesivir were calculated using the GROMACS [316] and DFTB+ [308] software programs for classical and quantum metadynamics simulations, respectively. Both the classical and DFTB calculations were performed on a single remdesivir molecule without any explicit or implicit solvent. The following sections provide the detailed settings and parameters used in our study for each of these approaches.

8.3.1 Amber Calculations

All-atom molecular dynamics simulations for the remdesivir molecule were performed with the Amber ff19SB force field [183] and generalized Amber force field (GAFF) [317] parameters via AntechAmber to collect an overall ensemble with a 2 µs sampling. It is important to note that although classical force fields are parameterized and validated under explicit-solvent conditions, they are routinely used in calculations performed in vacuo [318–320]. Charge parameters for remdesivir were assigned using a restrained electrostatic potential (RESP) [321] charge in vacuo. The structure of remdesivir was obtained

from the Protein Data Bank (PDB ID: 7BV2). The remdesivir structure was first optimized at the DFT/B3LYP/6-31G(d,p) level of theory using Gaussian 09 [135], and the RESP charges were calculated. All bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Temperature control (300 K) was performed via Langevin dynamics [322] with a collision frequency of $\gamma = 1$ ps. The system was then subjected to energy minimization. The system was further heated from 0 to 100 K in a canonical ensemble (NVT) by running two simulations of 5 ps each and imposing position restraints of 100 kJ mol^{-1} $Å^{-2}$. The temperature was further increased to 200 K in ≈100 ps of MD simulations in the NVT ensemble while reducing the restraint to 25 kJ mol^{-1} $Å^{-2}$. Subsequently, all restraints were released, and the temperature of the system was raised to 300 K in a single NVT simulation of 500 ps. After ≈1.1 ns of equilibration, ≈10 ns of NVT runs were carried out. All classical MD simulations were performed with the GPU-enhanced version of AMBER 20 [323]. The well-equilibrated system was used as starting point for the subsequent well-tempered (WT) metadynamics [292] simulations.

A structural assessment of remdesivir was performed using metadynamics simulations, which determined the conformational preferences of the dihedral angles in the main scaffold. Gaussian hills with an initial height of 1.2 kJ mol^{-1} and a hill width of 0.35 kJ mol^{-1} were applied to the system. In this WT scheme, Gaussian functions were rescaled with a bias factor of 10. The temperature was kept constant by a V-rescale thermostat (NVT step) with a coupling constant of $\tau = 0.1$ ps. The Lincs [324, 325] method was applied to constrain covalent bond lengths, allowing an integration step of 2 fs. The GROMACS 2019.6 [316] software package interfaced with the PLUMED plugin package

2.6.4 [326] was employed, and the "sum hills" tool from the PLUMED package was used to compute the free-energy surfaces.

8.3.2 DFTB calculations

All DFTB calculations in this study utilized high-performance computing hardware (40 Intel Xeon Platinum 8168 CPUs and 8 NVIDIA Volta V100 GPUs) executed on virtual machines (VMs) from Microsoft Azure cloud computing resources. Using the high-performance computing container maker (HPCCM) [327], an open-source tool for deploying the HPC components into container images, we created a docker image on the Azure cloud for DFTB v19.1 with the required libraries and dependencies (Intel MKL, Open MPI, Cuda, PLUMED v2.6 [326], and MAGMA v2.5.3 [310]). As such, this study also demonstrates the viability and readiness of cloud computing for high-performance computing workloads for first-principles computational approaches [328, 329]. In the present study, we used the self-consistent-charge formulation of DFTB (SCC-DFTB) in its third-order scheme (DFTB3), which includes the third-order term in the DFT energy expansion around the reference density [261]. We used the 3ob-3-1 Slater–Koster parameter set and its corresponding Hubbard derivative parameters, which have been previously shown to work well for biochemical systems [330–332]. We included DFT-D3 dispersion effects [333, 334] to accurately describe the London dispersion interactions in these biochemical systems. All the initial geometries for our metadynamics calculations were relaxed with nonperiodic boundary conditions (i.e., a cluster geometry), such that all the forces were less than 0.04 eV $Å^{-1}$. All the DFTB calculations were performed without any implicit or explicit solvent. All subsequent metadynamics calculations were performed after running an NVT equilibration for 2 ps. For all

the metadynamics runs, the temperature in the NVT ensemble (T = 300 k) was controlled using a Nose–Hoover thermostat [335,336]. Metadynamics calculations were performed with the PLUMED code [326] patched with DFTB+ [308]. All of our DFTB-based MD simulations used a time step of 1.0 fs, and all metadynamics calculations were carried out until the free energy converged with respect to each CV. In our metadynamics calculations, the height and width of the Gaussian hills were set to 1.2 and 0.35 kJ mol^{-1}, respectively. The deposition rate of the Gaussian hills was 500 MD steps, and a bias factor of 10 was used. Finally, we used the "sum hills" tool in the PLUMED package to compute the free-energy surfaces. To obtain the potential energy/entropy surfaces, we calculated the local average of the internal energy computed on a (ϕ, ψ) grid using our in-house pandas-based [337] python scripts. To smoothen the noise in our energy/entropy surfaces for visualization purposes, we used a Gaussian filter.

8.4 Results and Discussion

8.4.1 Timing Benchmarks

To evaluate the computational speedup gained from our heterogeneous CPU+GPU DFTB metadynamics calculations, we benchmarked the timings for carrying out eight SCC iterations of the first MD step on protease (PDB ID: 6LU7), which consists of 5,029 atoms. Table 8.1 compares the performance of various combinations of CPUs (Intel Xeon Platinum 8168) and GPUs (NVIDIA Volta V100) for performing eight SCC steps using the Divide-And-Conquer eigensolver in DFTB+. As shown in Table 8.1, the introduction of GPUs provides a significant speedup of nearly 140%. While this speedup is only for eight SCC

steps in a single MD step, this improvement scales exponentially for full MD calculations because multiple SCC steps are performed during each MD step. It is also worth mentioning that increasing the number of GPU cores from two to four did not enhance performance. One possible reason is the steep communication overhead associated with data transfer from the CPU to GPU, which adversely affected computational performance. Moreover, the matrix dimension is "only" 12,642 × 12,642, which is too small for effective multi-GPU scaling. It is also interesting to note that increasing the number of CPUs from four to eight did not increase performance. One of the reasons for this is Amdahl's law [338], which limits the scaling based on the dimension of the matrices being solved (which, in turn, depends on the size of the molecular system studied). Based on our benchmark timings (Table 8.1), a hardware configuration of two CPUs and two GPUs gave the best timings for calculations on protease 6LU7. We also performed similar benchmarks with both ADP and remdesivir and found that two CPUs and two GPUs also provided the most optimal hardware configuration for efficient metadynamics simulations.

Table 8.1: Comparison of timings for various hardware configurations for carrying out 8 SCC iterations on protease 6LU7.

Hardware Configurations		**Wall Clock (min)**
Number of CPUs	**Number of GPUs**	
40	4	23.43
20	4	7.74
10	4	7.98
8	4	7.89
4	4	5.59
8	2	5.87
4	2	5.17
2	**2**	**3.89**
8	1	6.05
4	1	3.95
2	1	3.93
1	1	32.45
8	0	14.74
1	0	59.09

8.4.2 Metadynamics Benchmarks on Alanine Dipeptide

To assess the efficiency and accuracy of our DFTB-based metadynamics calculations, we first performed a benchmark analysis with alanine dipeptide (ADP). In the scientific literature, ADP is frequently used as the archetypal system to evaluate the performance of various enhanced sampling methods, including, for example, the extended harmonic superposition approach [339], replica exchange solute tempering [340], string methods [341], as well as numerous metadynamics approaches [292, 342, 343]. The most accurate first-principles calculation of the free-energy surface of ADP to date is an ab initio molecular dynamics (AIMD) simulation at the PBE0 level of theory by de Pablo et al. using the adaptive biasing force method [300]. Two dihedral angles, ϕ and ψ, shown in Figure 8.1, are chosen as the collective variables (CVs) and are used to describe the thermodynamics of ADP. Using these two dihedral angles as CVs, we were able to identify three different minima in a Ramachandran plot, which describes the peptide's secondary structure. The first minimum, denoted as β, is located at $(\phi, \psi) = (-2.5, -2.5)$ radians and shows an almost-planar geometry. The second and third minima (located at $(\phi, \psi) = (-1.5, 1.2)$ and $(1.0, -1.2)$ radians, respectively) are stabilized by an intramolecular hydrogen bond and are denoted as $C_{7\mathrm{eq}}$ and C_{7ax}, respectively.

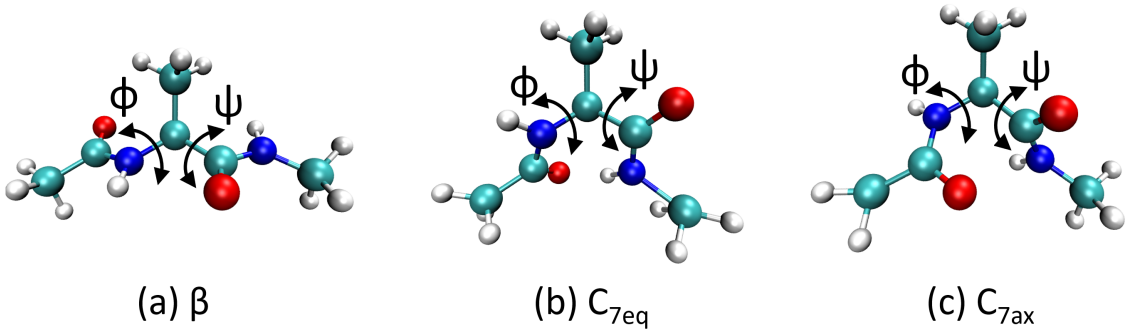


Figure 8.1: Molecular structures of three metastable minima: (**a**) β, (**b**) C_{7eq}, and (**c**) C_{7ax} of alanine dipeptide, which is composed of 22 atoms. Each panel depicts the two dihedral angles (ϕ, ψ) used to bias and analyze our calculations. The H, C, N, and O atoms are shown in white, cyan, blue, and red, respectively.

Figure 8.2 depicts the FES of the alanine dipeptide projected onto the ϕ and ψ dihedral angles obtained from well-tempered metadynamics simulations at 300 K. Figure 8.2 compares the FES of ADP obtained with the Amber99sb classical force field (Figure 8.2a), DFT at the PBE0 level of theory (Figure 8.2b), and DFTB3 (Figure 8.2c). The data used to plot Figure 8.2a,b were taken from Ref. [[300]]. As described in Ref. [[300]], there are clear differences between the DFT-PBE0 and classical force fields. These differences in the FES are distinctly visible near the maximum located at (ϕ, ψ) = (2.3, 1.2) radians in the Ramachandran plot, which is more pronounced in the AIMD calculations. In addition, the Amber99sb force field predicts a significantly larger barrier that spans the entirety of ψ at ϕ = 2.2 radians, which restricts conformational transitions across the dihedral angle.

In contrast with the classical force field results, the FESs obtained with DFT-PBE0 and DFTB3 (Figure 8.2) have the same general morphology and show some similar trends. In particular, both DFT-PBE0 and DFTB3 predict the same local minima regions on the Ramachandran plots (the β, C_{7eq}, and C_{7ax} local minima/metastable states are indicated by the ●, ■, and ▼ markers, respectively, in Figure 8.2). Moreover, DFT-PBE0 and DFTB3

predict the same maxima regions on the FES, which are located at approximately ϕ = 0 radians. Although morphologically similar, the DFT-PBE0 and DFTB3 FESs do exhibit some differences. DFTB3 predicts a much smaller barrier that spans the entirety of ψ at ϕ = 2 radians, likely allowing conformational transitions across the dihedral angle. The most noticeable discrepancies between the DFTB3 and DFT-PBE0 FESs appear at the center of the plots at (ϕ = 0, ψ = 0) radians. The DFTB3 plot shows a maximum near (0, 0) radians that is surrounded by two valleys constituting natural pathways between the C_{7eq} and C_{7ax} minima. The FES region near (0, 0) radians predicted by DFT-PBE0 shows no distinct local maxima; nevertheless, both DFTB3 and DFT provide similar geometries for the two C_{7ax} and C_{7eq} conformations. Moreover, our energies and geometries are consistent with those in Ref. [[344]], where the authors also used DFTB to calculate the FES of ADP. Taken together, our results demonstrate that DFTB3 reproduces the main features of the DFT-PBE0 FES (despite slight underestimation of barrier heights); most importantly, finite-temperature configurational properties/energetics predicted by DFTB3 are typically more accurate than those predicted by classical force fields.

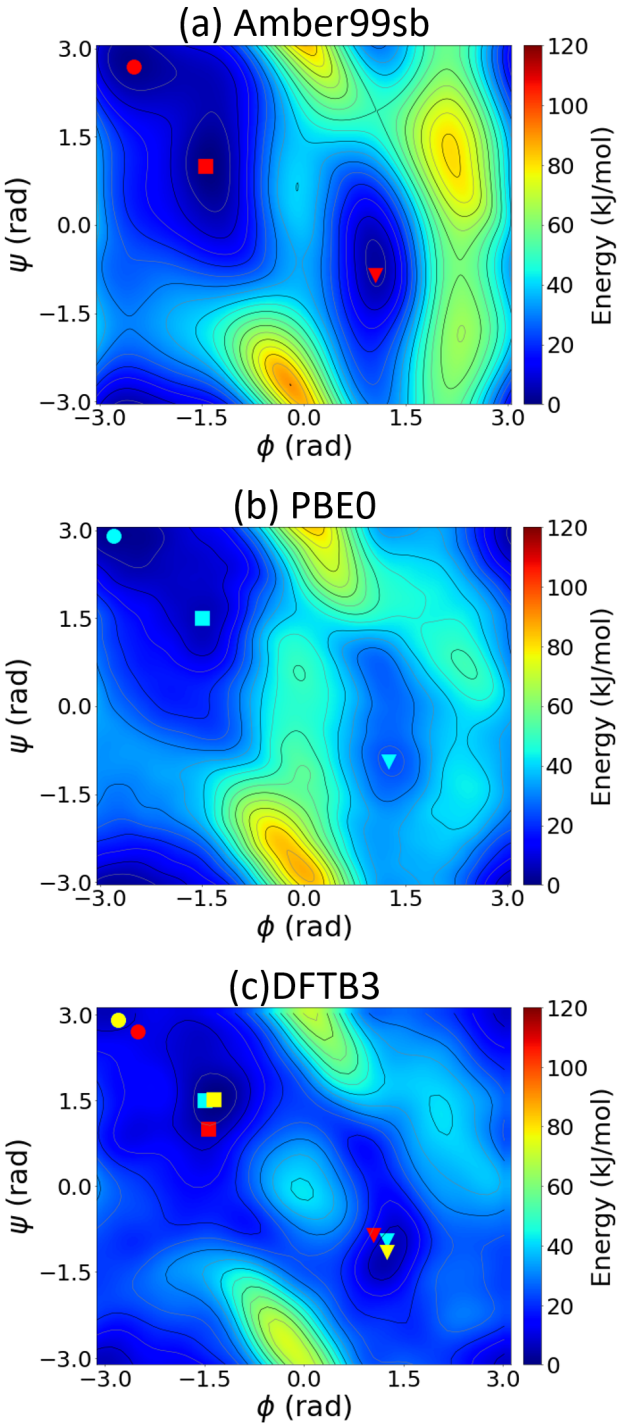


Figure 8.2: Two-dimensional free-energy surface of alanine dipeptide as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (**a**) classical MD with the Amber99sb force field, (**b**) DFT-PBE0 calculations, and (**c**) SCC-DFTB3 calculations. The red, cyan, and yellow points in panels (**a**–**c**) represent the local minima obtained using the Amber99sb force field, PBE0, and SCC-DFTB3, respectively. ●, ■, and ▼ denote the β, C_{7eq}, and C_{7ax} minima/metastable structures, respectively.

To understand the origin of the FES differences, we further investigated the contribution of internal energy and entropy to the free energy in the MD simulations. The change in free energy is given by

$$\Delta A(\phi, \psi) = \Delta U(\phi, \psi) - T\Delta S(\phi, \psi), \tag{8.16}$$

where A is the free energy, U is the internal energy, T is the temperature, and S is the entropy. The internal energy contribution to the FES was calculated using the local average of the internal energy computed during the MD simulations on a (ϕ, ψ) grid. The entropic term, $T\Delta S$, was calculated from the difference between the internal and free energy (i.e., $\Delta A(\phi, \psi) - \Delta U(\phi, \psi)$). Figure 8.3 compares the potential energy surfaces obtained from the various methods. The classical force field predicts a higher internal energy than the DFT-PBE0 and DFTB3 methods in the region corresponding to $\phi = 2$ radians, which is also reflected in the FES in Figure 8.2. The differences between the DFT-PBE0 and DFTB3 FESs are mirrored here, as the barriers predicted by the PBE0 functional are higher than those calculated by DFTB3 in the region corresponding to $\phi = 2$ radians. The PES obtained from the classical MD, DFT-PBE0, and DFTB3 approaches differ most in the low-probability states (i.e., states with high energy values): DFTB3 and DFT-PBE0 predict similar locations of the local minima, whereas the classical Amber force field gives quite different results.

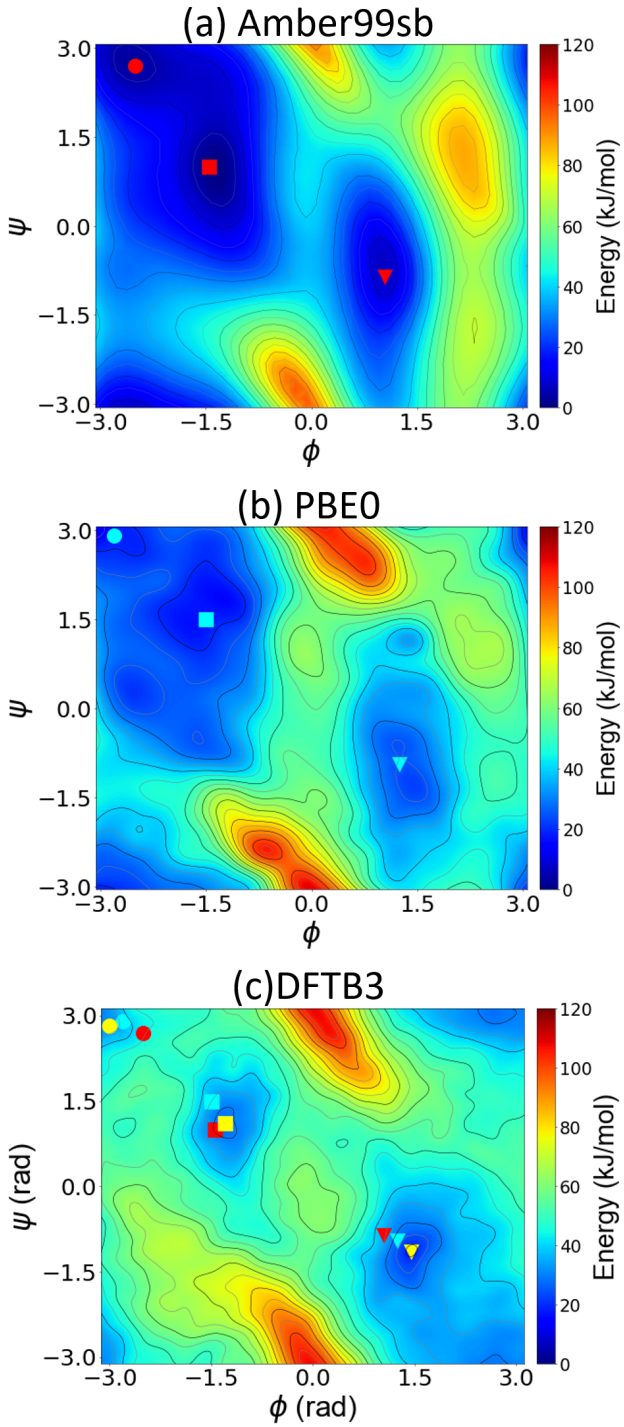


Figure 8.3: Two-dimensional potential energy surface of alanine dipeptide as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (**a**) classical MD from the Amber99sb force field, (**b**) DFT-PBE0 calculations, and (**c**) SCC-DFTB3 calculations. The red, cyan, and yellow points in panels (**a**–**c**) represent the local minima obtained using the Amber99sb force field, PBE0, and SCC-DFTB3, respectively. ●, ■, and ▼ denote the positions of the local minima.

The entropic contributions to the free energy also exhibit significant differences (Figure S1). The classical force field predicts a low entropy compared with the DFT-PBE0 and DFTB3 approaches. The entropy maxima for all three cases are located near the same locations as their corresponding FES and PES maxima. The DFTB3 plot in Figure S1c shows that two of the minima correspond to C_{7eq} and C_{7ax}, while the third one located at $(\phi, \psi) = (-0.5, -3)$ radians does not correspond to a minimum in the FES or a well-defined structure. The classical force field severely underestimates the entropic contribution to the free energy because entropy is not explicitly included in the fitting of the force field. As such, our GPU-enhanced DFTB calculations support the recent claim [300] that the entropic contribution is *essential* for obtaining an accurate description of large peptides, especially for folding and unfolding processes. In particular, our DFTB3 calculations for the FES/PES qualitatively agree with computationally intensive DFT-PBE0 benchmarks, whereas classical force fields give significant errors.

Most importantly, the computational effort/time for our DFTB calculations is significantly less than that of full DFT (while still being more accurate than classical MD). As reported in a previous study by de Pablo et al. [300], the DFT-PBE0 calculations took 4.5 weeks to carry out a 1.5 ns metadynamics simulation. However, our DFTB metadynamics simulation on ADP (22 atoms) took only ≈18 hours to obtain a 5 ns converged FES, indicating that our GPU-DFTB approach is nearly two orders of magnitude faster than DFT-PBE0. As noted previously, because the diagonalization algorithm scales as $\mathcal{O}(n^3)$ (where n is the matrix dimension), increasing the system size twice would incur an eight-fold increase in computational cost. Therefore, a system size of ≈80 atoms is well within the

capabilities of our GPU-enhanced DFTB approach (i.e., a metadynamics simulation of 10 ns would take ≈21 days), which cannot be easily obtained with DFT-based metadynamics.

8.4.3 Large-Scale GPU-DFTB Metadynamics Simulations of Remdesivir

With our GPU-enhanced DFTB approach validated against the high-level DFT-PBE0 ADP benchmarks, we then proceeded with metadynamics calculations of remdesivir as a proof-of-concept example of a system that is impractical to calculate with full DFT. Remdesivir has garnered recent attention in treating various ailments [345–347] and is a structurally complex molecule consisting of three key fragments: an adenine analogue base, a pentose sugar unit, and a phosphoramidate side chain. Broadly, predicting the FES landscape of promising drug candidates can guide the calculation of binding affinities and/or transition pathways to accelerate the drug design process. As such, the use of accurate computational approaches to efficiently predict the FES (such as the GPU-enhanced DFTB approach used here) can provide essential thermodynamic information for directed structure-based drug design.

As mentioned previously, converged metadynamics calculations with full DFT on large chemical systems such as remdesivir are impractical. More specifically, previous DFT-PBE0 metadynamics calculations on the 22-atom ADP molecule required 4.5 continuous weeks of computing time [300], and simulations on the 77-atom remdesivir molecule at that same level of theory would take several months. As such, the remdesivir metadynamics calculations in this study are excellent extensions of our GPU-enhanced DFTB capability to highlight and test its computational limits. To compare our DFTB3 calculations against those of conventional MD approaches, we also carried out classical Amber force field metady-

namics calculations. Figure 8.4 shows the structure of remdesivir with the dihedral angles, ϕ, and ψ, used to bias the metadynamics calculations.

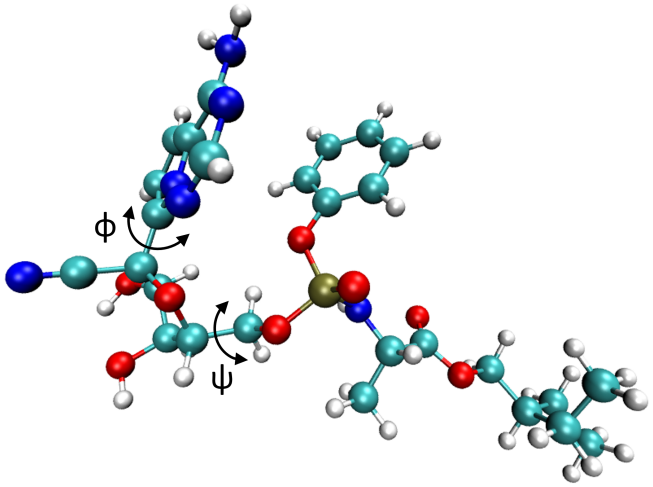


Figure 8.4: Molecular structure of remdesivir, which is composed of 77 atoms. The two dihedral angles, ϕ, and ψ, are used to bias and analyze our calculations. The H, C, N, O, and P atoms are shown in white, cyan, blue, red, and yellow, respectively.

Before calculating the FES, we examined the convergence of our DFTB WT-MetaD simulations by calculating the free energy as a function of time. In general, when a metadynamics simulation is converged, the resulting FES profiles are similar within a constant offset. Figure S2, in the Supporting Information, depicts the FES calculated every 0.5 ns up to a total time of t = 10.0 ns. We found that the FES did not change appreciably from t = 9.0 to 10.0 ns (other than a constant offset), which indicates that our simulations fully converged. Figure S3 shows the one-dimensional free energies (extracted from the full-dimensional metadynamics calculations) along the dihedral angles ϕ and ψ as a function of simulation time. We found that the free-energy differences of -6.92 and -3.80 kJ/mol associated with the basins near $\phi = -2$ and 2 radians and $\psi = -2$ and 1 radians, respectively, did not change appreciably, which indicated the FES calculations converged.

Figure 8.5 depicts the FES of remdesivir projected onto the ϕ and ψ dihedral angles calculated from well-tempered metadynamics simulations at 300 K via the Amber-ff19SB force field (Figure 8.5a) and DFTB3 (Figure 8.5b) approaches. The energy barriers and transition pathways for each of the plots were computed using the nudged elastic band (NEB) method as implemented in the Metadynminer package [348]. Using these two CVs for remdesivir, we were able to identify two dominant minima in a Ramachandran plot (see points A/B and C/D for the Amber force field and DFTB3 approaches, respectively), which describe the metastable structures of the molecule.

There are clear differences between the DFTB3 and classical force field predictions for the FES. The most discernible difference is near the maximum at $(\phi, \psi) = (2.8, 2.8)$ radians in the Ramachandran plot, which is less pronounced in the DFTB3 calculations. In addition, the Amber-ff19SB force field predicts a much larger barrier that spans the entirety of ψ at $\phi = 1.4$ radians, likely restricting conformational transitions across the dihedral angle. In addition, the DFTB3 calculations predict a much smaller barrier that spans the entirety of ψ at $\phi = 2$ radians, likely allowing conformational transitions across the dihedral angle. The local energy maximum predicted by DFTB3 near $(\phi, \psi) = (1.0, 1.5)$ radians is surrounded by a valley that constitutes natural pathways between the two minima at $(\phi, \psi) = (-0.50, 0.72)$ and $(1.80, 0.72)$ radians. The region near $\psi = 0.72$ radians predicted by DFTB3 is significantly different as there are no prominent maxima. Both DFTB3 and Amber provide similar geometries for the dominant minima near $(\phi, \psi) = (-0.50, 0.72)$ radians (points B and D in Figure 8.5). To assess the accuracy of the DFTB3 and Amber calculations, we carried out single-point hybrid-DFT calculations on remdesivir geometries

extracted from points A, B, C, and D to understand which of the two approaches are consistent with the more-accurate DFT calculations. Table S1 in the Supporting Information shows that the hybrid DFT calculations predict the molecular geometry at point C to lie lower in energy than any of the geometries extracted from points A, B, or D. As such, this test indicated that DFTB3, which predicts point C to be the global minimum in Figure 8.5b, is consistent with full DFT (in contrast, the Amber calculations shown in Figure 8.5a incorrectly predict the global minimum to lie at point A).

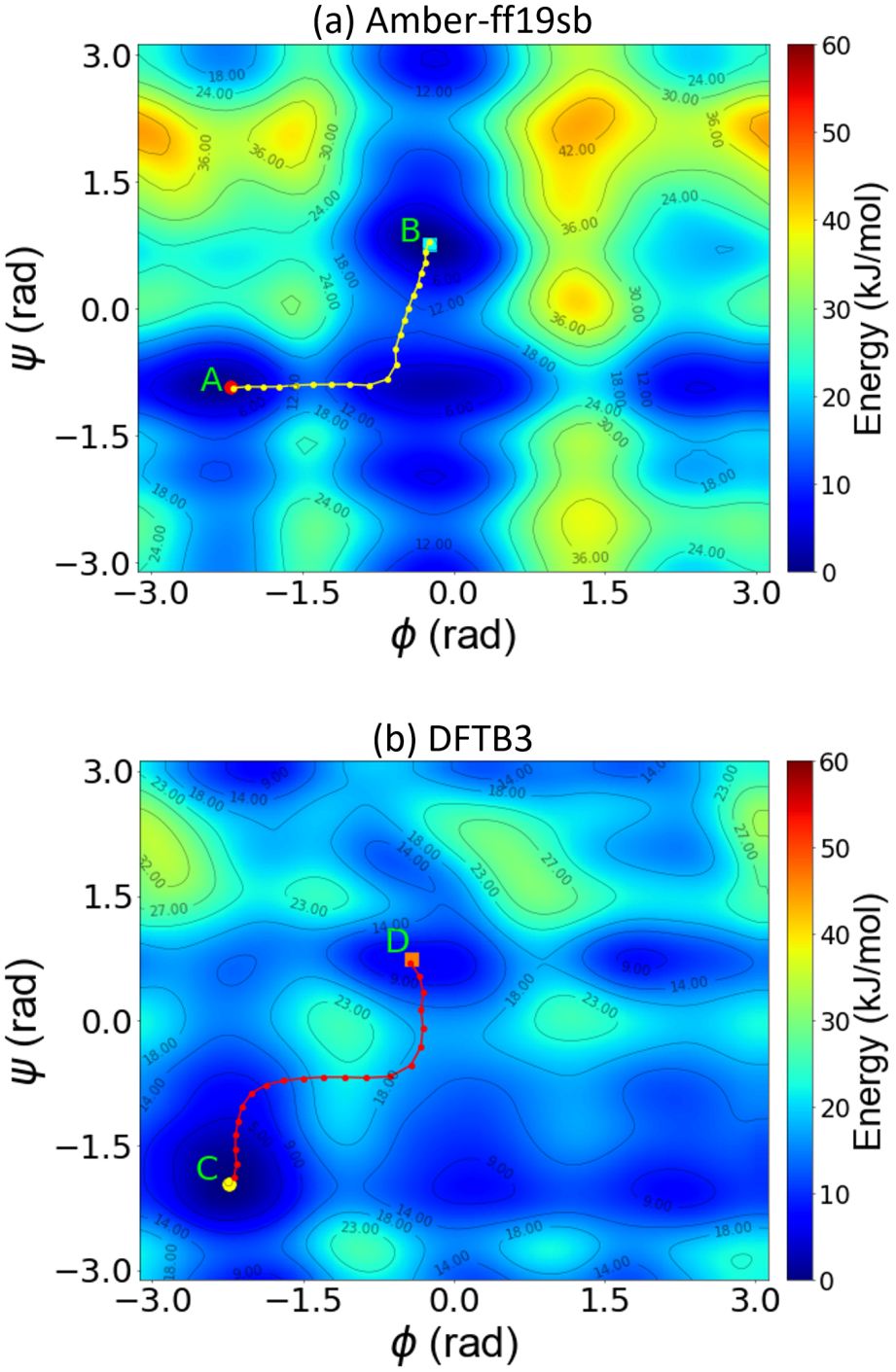


Figure 8.5: Two-dimensional free-energy surface of remdesivir as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (**a**) classical MD from the Amber-ff19SB force field and (**b**) SCC-DFTB3 calculations. Points A/B and C/D represent the dominant minima along the transition pathway (shown as a dotted line) obtained from the classical Amber force field and DFTB3 calculations, respectively.

Similar to our analysis on ADP, we investigated the origin of the FES differences by computing the contribution of internal energy and entropy for remdesivir. Figure 8.6 compares the potential energy surfaces obtained from the classical and DFTB3 approaches. The classical force field predicts higher internal energy than the DFTB3 methods in the entire region near ψ = 2 radians, which is also reflected in the FES in Figure 8.5. The differences between the Amber and DFTB3 FES are mirrored here, as the barriers predicted by the Amber calculations are higher than those predicted using DFTB3 in the region near ψ = 2 radians. These differences are also observed for the global minimum, which the Amber force field predicts to be less stable. Similar to ADP, the remdesivir PES predicted by Amber and DFTB3 differ most in the low-probability states. Figure S4 shows the entropic energy surfaces obtained using various methods. The classical force field predicts a higher entropy than DFTB3 in the region corresponding to ϕ = 2 radians, which is also reflected in the FES in Figure 8.5. The differences between the DFT-PBE0 and DFTB3 FES are mirrored here, as the barriers predicted by the PBE0 functional are higher than those calculated at the DFTB3 in the region near ϕ = 2 radians.

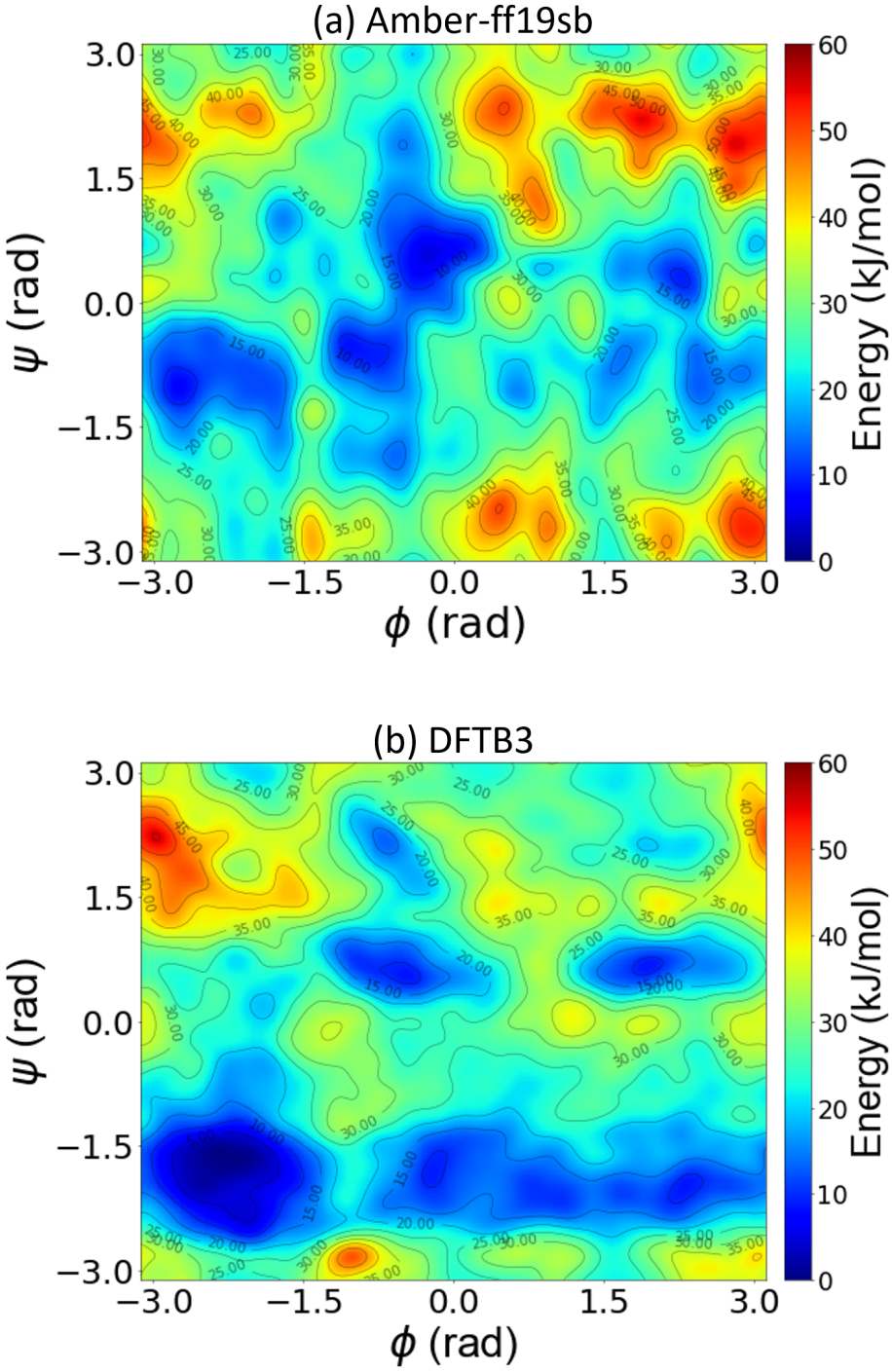


Figure 8.6: Two-dimensional potential energy surface of remdesivir as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (**a**) classical MD from the Amber-ff19SB force field and (**b**) SCC-DFTB3 calculations.

In summary, our results emphasize the importance of including quantum effects for accurately probing the metadynamics of remdesivir. In particular, our DFTB3 and Amber classical field calculations give qualitatively different predictions for the remdesivir FES. To estimate the accuracy of these two computational approaches, we carried out benchmark tests showing that the DFTB3 results are much more consistent with full DFT than the Amber classical force field. The resulting errors in the Amber classical force field manifest themselves in the FES by overestimating the internal energy contribution, particularly in the high free-energy remdesivir configurations. Taken together, our results show that our GPU-enhanced DFTB approach is a promising approach for accurately calculating the long-term metadynamics of remdesivir, which cannot be easily carried out with full DFT calculations.

8.5 Conclusions

In conclusion, we have extended our GPU-enhanced DFTB approach to enable efficient simulations of long-term metadynamics calculations of complex biochemical systems. Carrying out metadynamics calculations on these large biochemical systems is a natural extension of our GPU-enhanced DFTB approach because the diagonalization of the Hamiltonian matrix is performed several times during a single MD trajectory, which can be accelerated with massively parallelized GPUs. To enable these large simulations, we also carried out these calculations on Microsoft's Azure cloud platform to demonstrate the viability of cloud computing resources for quantum simulations.

After testing the performance of our GPU-DFTB approach on various hardware configurations for optimal performance, we verified the accuracy of our computational ap-

proach by calculating the free-energy surfaces of alanine dipeptide, which is a standard reference system for evaluating the performance/accuracy of enhanced sampling methods. In contrast to classical force fields, which give qualitatively different results than DFT-PBE0 benchmarks, we found that our GPU-enhanced DFTB calculations are in good agreement (with a much lower computational cost) with the computationally intensive DFT-PBE0 benchmarks. To further extend our GPU-enhanced DFTB approach, we also carried out a 10 ns metadynamics simulation of remdesivir, which is prohibitively out of reach for routine DFT-based metadynamics calculations. To the best of our knowledge, this is the first time that a quantum-based FES has been calculated for remdesivir for a relatively long sampling time of 10 ns. We found the free-energy surfaces obtained from classical and DFTB3 calculations differ significantly. Compared to DFTB3, the classical force field overestimates the internal energy contribution of high free-energy states in remdesivir, which produces dissimilar transition pathways that connect different minima on the free-energy surface. Taken together, our calculations and benchmark studies indicate that GPU-enhanced DFTB metadynamics is a promising technique for sampling the long-term thermodynamics of biochemical systems that require more accuracy than classical force fields but cannot be easily calculated with full DFT methods.

Chapter 9

Harnessing Deep Neural Networks to Solve Inverse Problems in Quantum Dynamics: Machine-Learned Predictions of Time-Dependent Optimal Control Fields

This chapter's work resulted from collaborating with a former lab colleague, Xian Wang. We contributed equally to designing the machine learning model to obtain the

amplitude and frequency of the optimal electric field required for enabling desired electronic transitions in quantum systems. Xian Wang independently designed the neural network for *directly* predicting the electric field (section 9.4). All parts of this work are included for completeness. This work is published in *Physical Chemistry Chemical Physics* journal [3].

9.1 Introduction

Inverse problems arise in many domains of quantum dynamics, with quantum optimal control being one of the most well-known examples. In the context of molecular systems, the field of quantum optimal control [349] seeks to steer a chemical system from a known initial state to a desired target state via an external field, $E(t)$, typically a tailored electromagnetic pulse. Predicting the explicit time-dependence of $E(t)$ is central to providing critical initial conditions for experiments across multiple chemical physics domains including light-harvesting complexes [198, 350–353], quantum information processing [354–356], laser cooling [357, 358], and ultracold physics [359, 360]. As such, the capability to fully harness these optically-driven systems has tremendous potential to grow as we understand how to *control* the excited-state quantum dynamical processes that govern these systems.

Although several approaches and algorithms have been proposed on optimizing quantum control fields (each with their own purposes and advantages [361–364]), all of these prior approaches are iterative in nature and require complex numerical methods to solve for these optimal control fields. Due to the nonlinear nature of these dynamical optimization problems, the number of iterations and floating point operations required by these algorithms can be extremely large, leading to extremely slow convergence (even

for relatively simple one-dimensional problems [363, 365]). Furthermore, when an optimal control field for a *new* quantum mechanical system is desired, the entire iteration process has to be re-started *de novo* since the algorithm has no prior "memory" of previously converged cases. Because of these computational bottlenecks, we wondered whether machine learning, particularly deep neural networks (DNNs), could offer a promising approach for obtaining solutions to this complex, inverse problem in quantum dynamics.

In recent years, machine learning has emerged as a powerful tool in the physical sciences for finding patterns (particularly those that evade human intuition) in high-dimensional data. While the majority of machine learning efforts in the chemical sciences have focused on equilibrium properties such as thermodynamic, [366–368], structural, [369–372] and ground-state properties [373–375] (to name just a select few), considerably less attention has focused on non-equilibrium *dynamical processes*, such as the explicitly time-dependent optimal fields discussed previously. As such, the use of machine learning in this largely unexplored application of quantum dynamics is a first step towards the design of machine-learned, time-dependent fields for efficiently controlling directed electron/energy transfer in these complex systems.

To this end, we present the first machine learning effort for solving time-dependent quantum control problems in reduced-dimensional chemical systems. These dynamical time-dependent systems pose a unique challenge for conventional machine learning techniques, and we investigate a variety of approaches for predicting optimal control fields, $E(t)$, in these systems. The present paper is organized as follows: Section II briefly outlines the basic concepts of quantum control and the requisite datasets used by the machine learning

approaches in our work. Section III describes a neural network approach for predicting the frequency and amplitude content of the power spectrum in the frequency domain (i.e., the Fourier transform of $E(t)$), whereas Section IV provides a cross-correlation neural network approach for directly predicting $E(t)$ in the time domain. Finally, Section V concludes with a brief discussion and perspective look at potential future applications of our machine learning approach.

9.2 Theory and Computational Methodology

9.2.1 Brief Overview of Quantum Control

Since the main purpose of this work is to harness machine learning techniques for controlling dynamic chemical systems, we only give a brief overview of quantum optimal control and point the interested reader to several topical reviews in this area. [376–379] For chemical systems, the quantum optimal control formalism commences with the time-dependent Schrödinger equation for describing the temporal dynamics of nuclei, which, in atomic units is given by

$$i\frac{\partial}{\partial t}\psi(x,t) = \left[-\frac{1}{2m}\frac{\partial}{\partial x^2} + V(x) - \mu(x)E(t)\right]\psi(x,t). \tag{9.1}$$

In the equation above, x denotes the reduced coordinate along a chosen reaction path, [380–383] m is the effective mass associated with the molecular motion along the reaction path, [384,385] $V(x)$ is the Born-Oppenheimer electronic energy of the molecule, $\mu(x)$ is the dipole moment function, $E(t)$ is the time-dependent external electric field, and $\psi(x,t)$ represents the probability amplitude for the motion of the nuclei along the reduced coordinate path.

Both $V(x)$ and $\mu(x)$ can be obtained from a standard quantum chemistry calculation by carrying out a relaxed potential energy scan. [386, 387]

With x and $V(x)$ properly chosen/computed, Eq. (9.1) allows us to mathematically answer the question: "*Given an electric field* $E(t)$, how does an initial state, $\psi_0(x, t = 0)$, evolve after some final time T has elapsed?" However, as mentioned in the Introduction, the field of quantum optimal control is an inverse problem and instead seeks the answer to the "inverse" question: "If we want to reach a desired final state $\psi_{N-1}(x, t = T)$ at time T (after $N - 1$ propagation steps), *what does the functional form of* $E(t)$ *look like?*" To be more mathematically precise, quantum control seeks the functional form of an external electric field, $E(t)$, that maximizes the functional $J[\psi_{N-1}, E]$ given by

$$J[\psi_{N-1}, E] = \left| \int_{-\infty}^{\infty} \psi_f^*(x)\psi_{N-1}(x)dx \right|^2 - \alpha \int_0^T E(t)^2 dt, \tag{9.2}$$

where ψ_f is a known desired final target wavefunction (given by the user), and ψ_{N-1} is obtained after applying $N-1$ successive propagation steps of the time-dependent Schrödinger equation (i.e., Eq. 9.1). It should be noted that the first term in Eq. (9.2) is essentially a measure of the similarity of the final target and the propagated wavefunction. The second term in Eq. (9.2) is a fluence and acts as a penalty to prevent unphysically large values of the electric field, where α is a positive constant (set to 0.001 in this work) to be chosen by the user. Providing accurate and efficient answers to this inverse question is the ultimate goal of the machine-learning approaches described in this work.

9.2.2 Generation of Datasets Used for Machine Learning

To generate the data required for our machine learning approaches, we utilized the NIC-CAGE (Novel Implementation of Constrained Calculations for Automated Generation of Excitations) program developed in our previous work. [388] Given a potential, $V(x)$, this program iteratively calculates a numerical representation of $E(t)$ that enables a $\approx$100% transition probability between two desired electronic transitions (which, in this work, are the ground and first-excited state, schematically shown in Figs. 9.1a and 9.1b. In simple terms, our NIC-CAGE program can be seen as a black box that accepts potential functions, $V(x)$, as input and subsequently outputs optimal electric fields, $E(t)$, corresponding to the inputted potentials. It is important to note that the optimal electric field, $E(t)$, can also be represented in the frequency domain as a power spectrum, $\sigma(\omega)$, by applying a fast Fourier transform (FFT) to $E(t)$ (cf. Fig. 9.1c). In this work, we seamlessly switch between the time and frequency domains to provide different machine learning approaches for predicting optimal control fields in these dynamic systems.

While the NIC-CAGE program [388] can obtain transition probabilities with notable accuracy (typically over 97%), it can take hundreds of iterations (or longer) to converge to the final electric field for each potential. Moreover, as mentioned in the Introduction, when a new potential is inputted, the iteration process has to be re-started anew since the program has no prior memory of previously converged cases. For these reasons, the prediction of optimal electric fields for a *general* potential energy function is a natural application for a data-driven solution. To generate a large dataset for our machine learning approaches, a vast number of potentials were generated as input to the NIC-CAGE program to produce

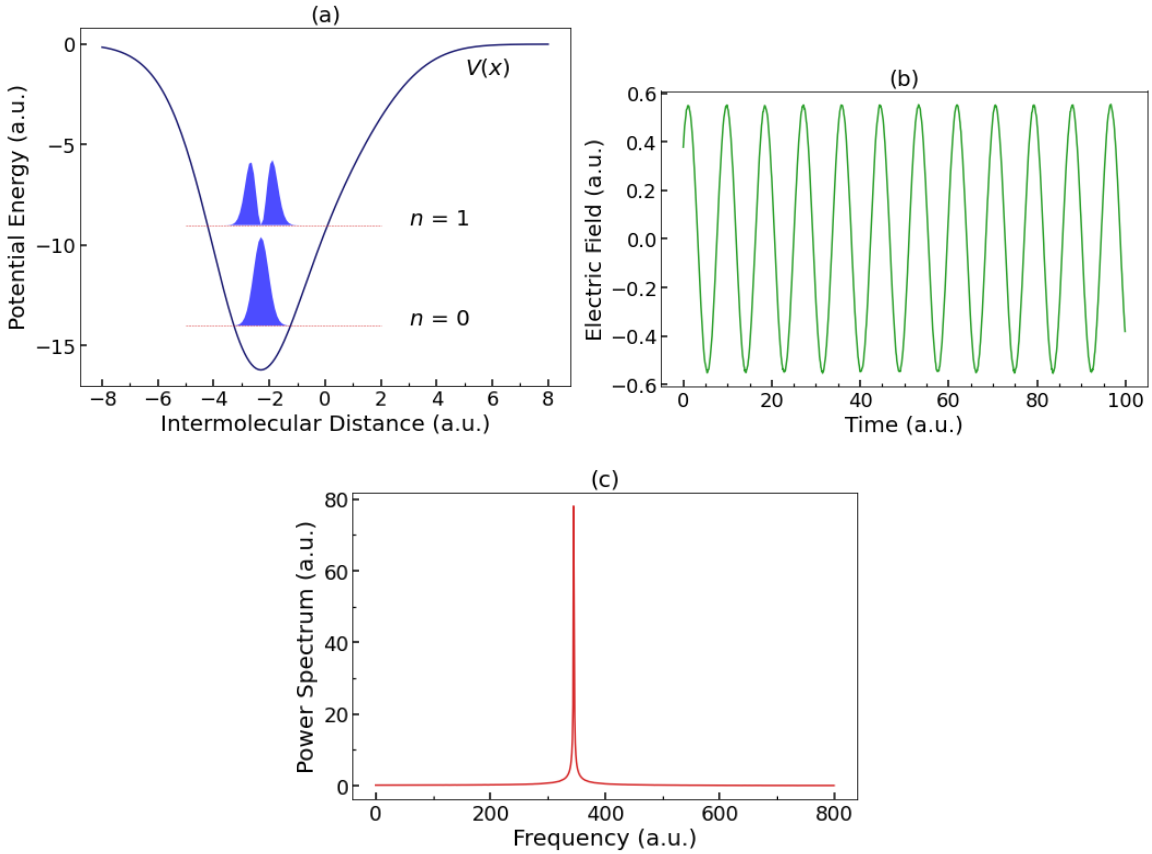


Figure 9.1: Schematic example of (a) a potential well, $V(x)$, as a function of intermolecular distance x. The horizontal dashed lines denote the energy levels of the ground and first excited state, and their respective probability wavefunctions, $|\psi(x)|^2$, are depicted as blue curves above the energy levels; (b) the optimal electric field $E(t)$ required to excite the transition between the ground and the first excited state; (c) the corresponding power spectrum $\sigma(\omega)$ as a function of frequency ω, obtained from the fast Fourier transform of $E(t)$.

corresponding optimized electric fields. These potential-field pairs served as the training, validation, and test sets for our DNNs.

Our complete dataset consisted of 36,118 randomly generated potential functions, $V(x)$, each of which was evaluated across 192 points in one dimension. For all of these potential functions, the effective mass, m, and dipole moment, $\mu(x)$, were set to 1 and x, respectively. To enable statistical flexibility in this dataset, each potential was constructed

by the summation of three Gaussian functions with varying amplitudes, widths, and centers, according to the following equation: [389]

$$V(x) = -\sum_{i=1}^{3} A_i \exp\left[-\frac{(x-\mu_i)^2}{2\Delta_i^2}\right] \tag{9.3}$$

Specifically, our dataset was created by randomly sampling each of the parameters with the following ranges: amplitude $A \in [1, 10]$, center $\mu \in [-3, 3]$, and width $\Delta \in [0.5, 2]$. As such, each potential function can be fully described by nine randomly generated parameters. In addition, we also visualized this parameter space and found that all parameters were evenly distributed within the selected range, indicating that the randomly generated potential functions sufficiently span this phase space (cf. Fig. 9.2). Each of the 36,118 potential functions was inputted into the NIC-CAGE code, which resulted in an optimized electric field evaluated across 30,000 points in the time domain.

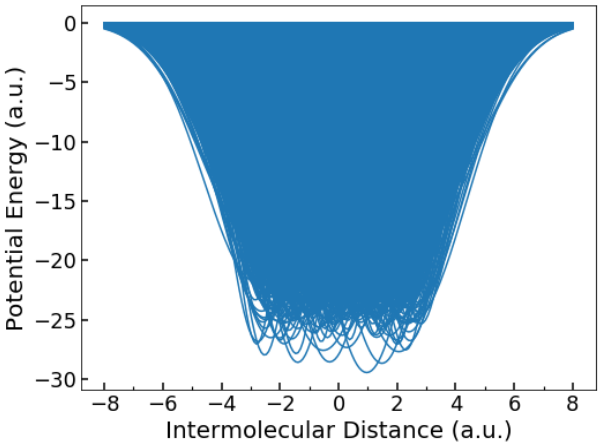


Figure 9.2: Plot of all 36,118 potentials sampled in this work. The center region of the $V(x)$ space is densely packed and fully sampled, indicating that the full set of these potentials sufficiently explores this phase space. The side regions of the figure are not filled by the potential energy curves since the range of the Gaussian centers, μ, were intentionally kept small to prevent the wavefunctions from spreading outside the $x \in$ [-8.0, 8.0] range.

Of the 36,118 potentials examined in this work 26,000 were used for the training set, 5,000 were utilized for the validation set, and the remaining 5,118 potentials were designated for the test set. We ensured that the number of potentials used in the training, validation, and test sets were exactly the same for each training instance to ensure that the results could be compared.

9.2.3 General Neural Network Architectures

We employed feedforward neural networks (FNNs) for this work due to their simplicity as well as their ability to learn complicated mappings between input and target spaces. The FNN architectures used here can be classified as a deep network architecture, with the depth in each network arising from the stacking of multiple hidden layers. Each hidden layer accepts output from the previous layer as input, and returns a non-linear activation as the output. It is worth nothing that the predictive accuracy of the FNNs can be sensitive to several key hyperparameters and training methods, such as the number of hidden layers, the number of nodes in each layer, the learning rate, and the regularization method. As such, multiple models and parameters were tested in this work (i.e., we also tested convolutional neural networks but found that the best results were obtained with FNNs), and we only present FNN architectures and parameters in Sections III and IV with the best performance.

9.3 Neural Networks for Predicting the Resonance Frequency and Amplitude, $\sigma(\omega)$

In this section, we describe our first machine learning approach, which utilizes FNNs to predict the frequency and amplitude of the power spectrum, $\sigma(\omega)$, in the frequency domain. As briefly mentioned in Section II. B., the power spectrum is obtained by a standard numerical procedure in which a fast Fourier transform of a properly converged $E(t)$ is first computed, followed by taking its absolute value. It is worth mentioning that because of the last absolute value operation, the phase of the original electric field is inherently lost and, therefore, only the amplitude and frequency were predicted with our FNNs in this section. To this end, we utilized two independent FNNs to separately learn the frequency and amplitude, and a schematic of the FNN architecture used for both of these predictions is shown in Fig. 9.3.

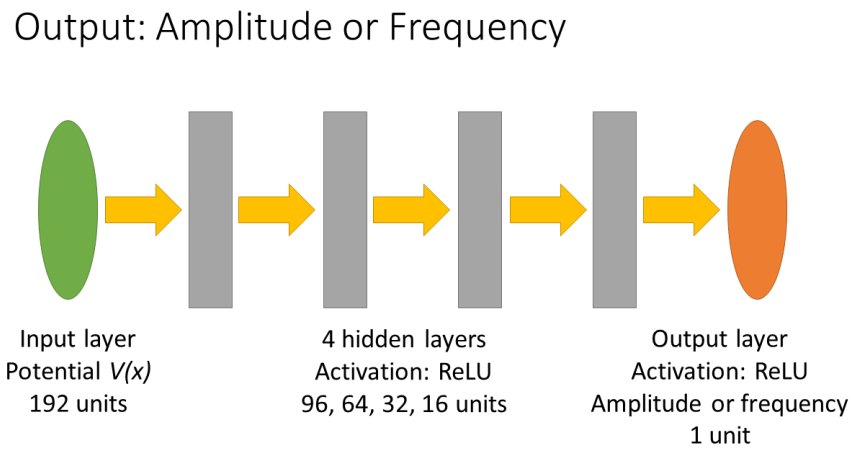


Figure 9.3: Architecture of the FNN used to predict the amplitude and resonance frequency of the power spectrum, $\sigma(\omega)$. The FNN starts with an input layer composed of 192 units (which correspond to the potential, $V(x)$, evaluated across 192 points), followed by four hidden layers of various sizes. The output layer is composed of 1 unit to predict either the amplitude or resonance frequency of $\sigma(\omega)$.

Table 9.1: Hyperparameters and settings of the FNNs used for predicting the amplitude and frequency of the optimized $E(t)$.

Output purpose	Amplitude	Frequency
Neural network structure	Feedforward	Feedforward
Activation	ReLU	ReLU
Learning rate	0.0001	0.0005
Loss function	MSE	MSE
Regularization	L2	L2
Regularization coefficient	0.0001	0.0005
Mini-batch size	1024	1024
Number of hidden layers	4	4
Number of units in hidden layers	96, 64, 32, 16	96, 64, 32, 16

Upon closer inspection of the original test set used in this work, we noticed that 66 of the optimal $E(t)$ fields had extremely large amplitudes (i.e., these specific electric fields were characterized by amplitudes that were an order of magnitude larger than the average $E(t)$ in the test set). Since electric fields with these large amplitudes are difficult to construct in a realistic experiment, we eliminated these 66 data points (they account for only 1.29% of the 5,118 data points), and we designated this dataset as our pruned test set. The input for each of our independent FNNs was the potential $V(x)$ (consisting of 192 data points), whereas the output was the single value of the frequency or amplitude, as depicted in the last step of Fig. 9.3. Both of these two FNNs were constructed and trained using a Tensorflow [390] backend with GPU acceleration powered by NVIDIA CUDA libraries [391]. In each FNN model, all of the weight matrices were initialized with random values satisfying a normal distribution, while all the biases were initialized to 0.001. We chose our loss function based on the definition of the mean square error (MSE), given by the following equation:

$$\text{loss} = \frac{\sum_{i=1}^{N} (y_{\text{true}} - y_{\text{pred}})^2}{N} \tag{9.4}$$

where N is the mini-batch size, y_{true} is the true frequency/amplitude of $\sigma(\omega)$ obtained from the NIC-CAGE program, and y_{pred} is the frequency/amplitude predicted by the machine learning algorithm. An L2 regularization of the weights was applied to prevent overfitting, and the built-in Adam optimizer was utilized. The training, validation, and test sets were kept the same size, and after several tests, we found that the optimal learning rates and regularization coefficients were different for these two FNNs, while all other optimal hyperparameters had the same values. Table 9.1 summarizes the optimal hyperparameters used

in each of these FNNs.

Table 9.2: FNN metrics for predicting the amplitude and frequency, respectively.

Output	Amplitude	Frequency
Number of epochs for best performance	~1000000	~1000000
Loss on original test set	509.1925	286.1925
R^2 for pruned test set	0.6036	0.9814

Fig. 9.4 depicts the results of our machine-learned amplitudes and frequencies. The diagonal line in each plot represents a perfect match between the machine-learned predictions and true values (obtained with 1,000,000 epochs). To further quantify this performance, we computed a coefficient of determination (R^2) for measuring the similarity between y_{pred} and y_{true}:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{\text{pred}} - y_{\text{true}})^2}{\sum_{i=1}^{N}(y_{\text{pred}} - \hat{y}_{\text{pred}})^2} \tag{9.5}$$

where N is the batch size, and $\hat{y}_{\text{pred}}$ is the average of all the y_{pred} values in the batch. A perfect agreement between y_{pred} and y_{true} yields an R^2 value of 1. As visually shown in Fig. 9.4 and from the R^2 values listed in Table 9.2, our machine learning approaches were more accurate in predicting the resonance frequency compared to the amplitude. This difference in performance suggests that the machine-learned mapping from the potential to the amplitude is much more complicated than the mapping from the same potential to the resonance frequency. More concretely, the frequency has a more clear/intuitive physical meaning, which is equal to the energy difference between the ground- and first-excited state. However, the amplitude is much more sensitive to the underlying shape of the potential, $V(x)$, and this sensitivity contributes to the error in predicting the amplitude with our

neural network. This difference in predictive performance can also be seen by comparing the figures of the R^2 values vs. the epoch number on the validation set. In particular, the R^2 values for predicting the frequency show a smooth progression, while that for the amplitude fluctuates significantly as shown in Figs. 9.5c and 9.5d. We also investigated the sensitivity of our results to the size of our training set and found that the accuracy of the machine-learned predictions decreased with the training set size. Specifically, when the training set was reduced to only 10,000 potentials, the R^2 values for predicting the resonance frequency and amplitude in the same validation set decreased to 0.93 and 0.41, respectively. As such, these statistics showed that a sufficiently large training set was necessary to enable accurate machine-learned predictions for these optimal control fields.

We also explored the option of predicting the entire power spectrum instead of just the primary resonance frequency and amplitude. Several attempts were made along those lines, including reducing the size of the output to 800 rather than 15,000 (since the resonance peaks typically had small frequencies), choosing a cross-entropy loss function instead of the MSE, fixing the lineshape of the output to be a Gaussian or symmetric Lorentzian to reduce the number of units (i.e., to 3) required in the output layer to predict the power spectrum, etc. Unfortunately, all of these attempts failed in predicting the correct amplitude of the power spectrum, although some of them were quite successful in predicting the resonance frequency. We attribute these failures to the sharpness of the resonance peak in the power spectrum. Due to the limited resolution inherent to the discrete $\sigma(\omega)$ data, each peak only consisted of a few data points and, therefore, the linewidth was not well-resolved. In other words, since the linewidth of the resonance peak in $\sigma(\omega)$ was inherently imprecise, the FNN

was unable to converge to a proper mapping of the power spectrum. In addition, we also tested one-dimensional convolutional neural networks (CNNs) for predicting the frequency and the amplitude as well as the entire power spectrum. Unfortunately, the results obtained with CNNs were less accurate than those obtained with the FNN approaches used here. Because of these limitations, we investigated other FNN architectures to learn mappings between $V(x)$ and $E(t)$ in the time domain. This is motivated by the fact that if $E(t)$ can be accurately predicted using FNNs in the time domain, $\sigma(\omega)$ could also be accurately resolved (since $\sigma(\omega)$ is merely the Fourier transform of $E(t)$), and we discuss these strategies in the next section.

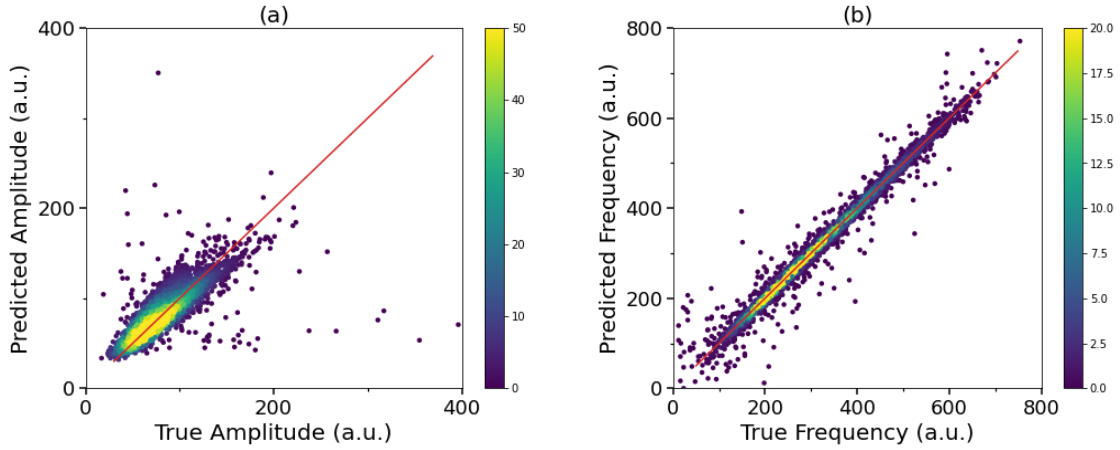


Figure 9.4: Density plots of the machine-learned predicted vs. true (a) amplitudes and (b) frequencies. The diagonal line in each plot represents a perfect match between the machine-learned predictions and true values. Both plots were obtained with 1,000,000 epochs.

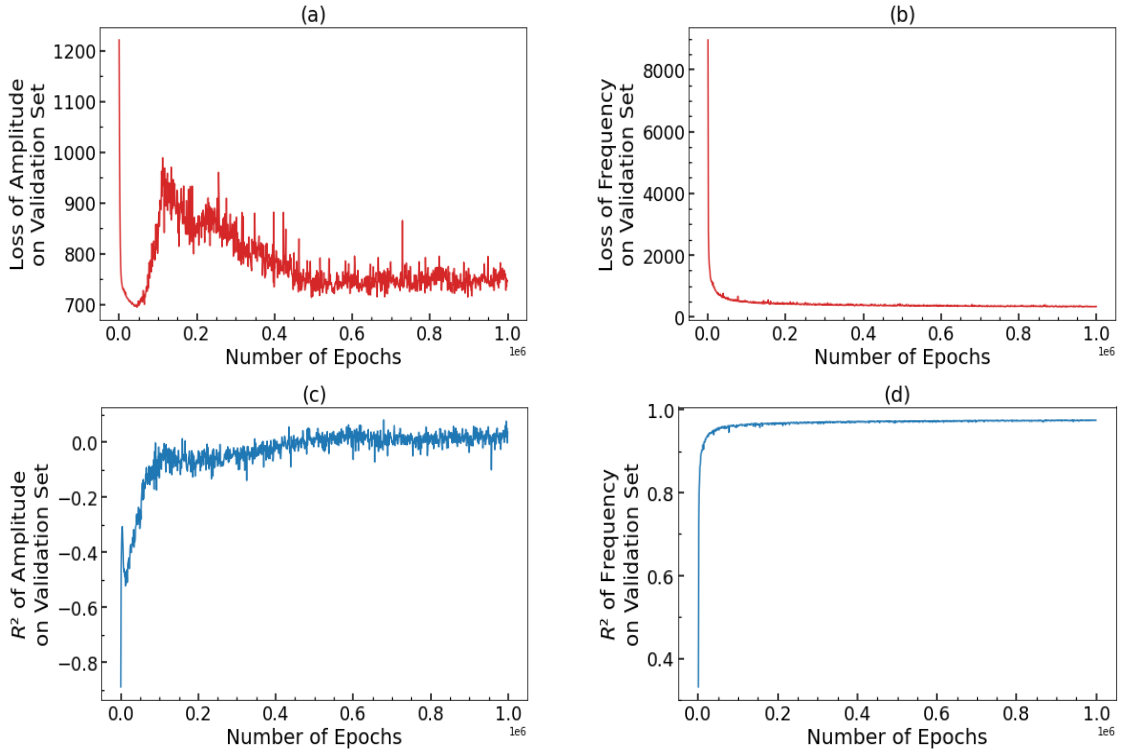


Figure 9.5: Plot of loss vs. number of epochs for FNN predictions of (a) the amplitude and (b) resonance frequency. R^2 values for the FNN-predicted (c) amplitude and (d) resonance frequency. All plots were generated from the validation dataset.

9.4 Neural Networks for Directly Predicting the Electric Field, $E(t)$

While Section III focused on predicting the power spectrum, $\sigma(\omega)$, in the frequency domain, we now investigate whether the electric field in the time domain, $E(t)$, can be predicted with a machine learning approach. Predicting these dynamic fields as an explicit function of time presents unique challenges for machine learning approaches. In particular, while $\sigma(\omega)$ in the frequency domain contains no phase information, $E(t)$ in the time domain *does* contain an explicit phase dependence (cf. Fig. 9.1b) that requires additional care, which we discuss in further detail below.

To predict $E(t)$ as an explicit function of time, we constructed an FNN with three hidden layers, which was trained with the same GPU-accelerated Tensorflow [390] backend and NVIDIA CUDA libraries [391] used in Section III. Our FNN, depicted in Fig. 9.6, was designed such that the number of units increases as data flows towards the output layer. Specifically, the input layer was composed of 192 units (which correspond to the potential, $V(x)$, evaluated across 192 points), followed by three hidden layers having 300, 500, and 750 units, respectively. The output layer, which outputs the electric field as a function of time, was composed of 1,000 (or fewer) units. Similar to the FNN used in Section III, the activation for both the input and hidden layers was chosen to be a ReLU function without any leaky or bounded modification. Since the output array is expected to be sinusoidal with a zero base, the activation of the output layer was chosen to be a tanh function to enable the output of negative values. All of the weight matrices were initialized with random values satisfying a normal distribution, while all the biases were initialized to 0.001. We chose the same loss function (cf. Eq. (4)), L2 regularization, and Adam optimizer described previously in Section III for our FNN. Based on several tests of our data, we found that a regularization coefficient of 0.001 was optimal for balancing regression speed and overfitting.

For the specific case of excitations from the ground to the first-excited state, we noticed that the optimal electric field, $E(t)$, could be closely approximated with a sinusoidal function (with a single frequency and amplitude) regardless of the potential function used. Because of this periodicity, the time-dependent trends in these electric fields could be accurately captured by only considering a smaller portion of the entire periodic signal. To this end, we only extracted 1,000 (or fewer) representative data points within the entire

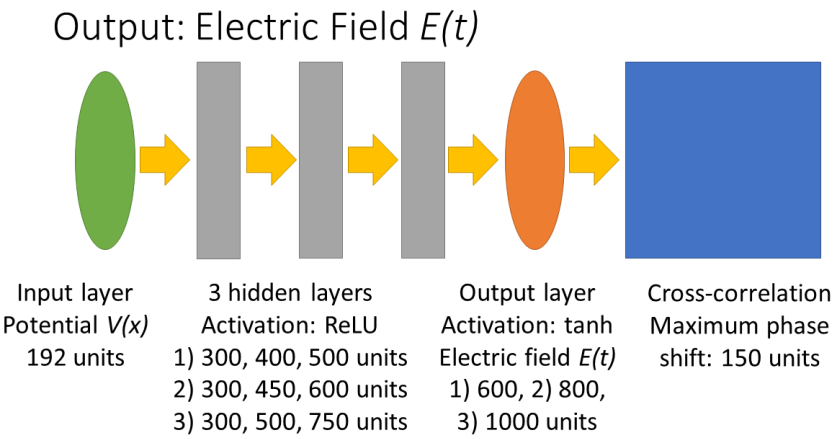


Figure 9.6: Architecture of the FNN used to predict the electric field, $E(t)$. The FNN starts with an input layer composed of 192 units (which correspond to the potential, $V(x)$, evaluated across 192 points), followed by three hidden layers of various sizes. The output layer is composed of 1,000 (or fewer) units and is directly interfaced with a cross-correlation algorithm to predict the final electric field, $E(t)$.

30,000-point electric field for our output set. This simplification allowed us to train our machine learning models more easily due to constraints in holding this large amount of data in RAM, the immense computing time, and associated GPU resources.

In the same spirit of reducing the number of physically relevant parameters needed for our machine-learning efforts, we also explored whether the transition probability was sensitive to the specific phase factors or amplitudes directly obtained from the NIC-CAGE code. To test the first assumption, we inputted several electric fields with different phase shifts, φ (but having the same optimized frequency and amplitude that gives the desired transition), as an initial guess into the NIC-CAGE code (cf. Fig. 9.7a). All of these phase-shifted electric fields gave a transition probability close to unity (with the NIC-CAGE code exiting immediately without further iterations), indicating that the transition probability was not dependent on the phase. However, when we tested the second assumption by in-

putting electric fields with different amplitudes as an initial guess into the NIC-CAGE code (cf. Fig. 9.7b), we observed a completely different phenomenon. Specifically, all of these initial conditions resulted in several subsequent iterations that eventually reverted/converged to the *same* optimal $E(t)$ form (cf. Fig. 9.7c). Taken together, both of these benchmark tests indicate that the optimal $E(t)$ is insensitive to the phase but highly dependent on the amplitude. As such, these tests allow us to construct a streamlined FNN using a cross-correlation technique for predicting $E(t)$ in the time domain (without having to directly predict the phase factor since it has no physical effect on the dynamics), which we describe in further detail below.

For the ground to first-excited state transitions examined in this work, each of the optimal control fields, $E(t)$, can be nearly characterized by a single amplitude, frequency, and phase, φ. Since we showed previously that the transition probability is insensitive to φ, a conventional neural network may be unable to learn any patterns that map between $V(x)$ and φ, since the phase is arbitrary and has no physical meaning. To sidestep this difficulty, we used a cross-correlation approach to shift the predicted $E(t)$ by a series of different phase values. In essence, this generates multiple $E(t)$ functions with exactly the same frequency and amplitude but with a variety of different phases.

To this end, 150 shift-matrices were constructed by shifting the identity matrix along rows with a "roll" function. To more concretely illustrate how we automated these phase shift operations in our machine learning approach, we denote $E(t)$ as a row vector given by,

$$E(t) = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{999} \\ E_{1000} \end{pmatrix}^{\mathrm{T}} \tag{9.6}$$

Therefore, $E(t)$ can be trivially written as

$$\begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{999} \\ E_{1000} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{999} \\ E_{1000} \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{9.7}$$

By shifting the diagonal entry of the identity matrix, the phase of $E(t)$ can be "rolled" or shifted as follows:

$$\begin{pmatrix} E_{1000} \\ E_1 \\ E_2 \\ \vdots \\ E_{999} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{999} \\ E_{1000} \end{pmatrix}^{\mathrm{T}} \cdot \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}. \tag{9.8}$$

Using this approach, the predicted $E(t)$ can be shifted along the time axis by 0 to 150 increments when multiplied by the matrix in Eq. 9.8. As such, each output array was spanned to a set of 150 arrays with exactly the same frequency and amplitude, but with different phases, φ. We also tested the accuracy of this approach by using a smaller number of shift matrices but found that at least 100 of these arrays were needed to sufficiently sample the entire phase space of $\varphi \in [0, 2\pi]$ (i.e., each new array shifts the phase, φ, by at least $2\pi/100$, and 100 or more arrays were necessary to ensure that the phase within the interval $[0, 2\pi]$ was sufficiently represented to give accurate results). With these 150 shift matrices in hand, the MSE loss was computed for each prediction, and when the phase of the prediction matched that of the true $E(t)$, the MSE loss was minimized. The weights and biases of the neural network were then updated using a back-propagation algorithm based on the minimum loss value. It is worth noting that our cross-correlation approach was only used to train the neural network, and after the neural network was successfully trained, the cross-correlation procedure was no longer needed to process/predict new data.

Table 9.3: Hyperparameters and settings of the FNN used for predicting $E(t)$ in the time domain.

Neural network structure	Feedforward
Activation	ReLU (hidden layers) tanh (output layer)
Learning rate	0.0001
Loss function	MSE
Regularization	L2
Regularization coefficient	0.001
Mini-batch size	1024
Multiplicative pre-factor of $E(t)$	80
Maximum phase-shift in cross-correlation	150 increments

Table 9.4: FNN metrics for predicting $E(t)$ in the time domain with the 600-, 800-, 1000-output-layer-unit models.

Number of output layer units	600	800	1000
Number of epochs for best performance	~30,000	~40,000	~50,000
Loss on original test set	25.0653	44.5876	68.2800
R^2 for amplitude on pruned test set	0.3702	0.2594	0.1485
R^2 for frequency on original test set	0.9550	0.9381	0.9370

We optimized some of the hyperparameters used by our cross-correlation neural network approach for the training set, and the optimal learning rate was chosen to be 0.0001. A mini-batch of 1,024 input arrays was chosen from the training set for each training epoch, and the training set was fully shuffled after each epoch. Since the electric fields outputted by the NIC-CAGE program had amplitudes on the order of ~ 0.01, all of the electric fields were multiplied by 80 to avoid numeric underflows and allow the weights and biases to converge faster in our machine learning algorithms. We chose a scaling factor of 80 to ensure that the processed electric field would not exceed 1, since the tanh function used in our output layer has a range of [-1, 1]. Table 9.3 summarizes our selection of hyperparameters and settings used to predict $E(t)$ in the time domain.

To reduce the large RAM requirements and computational effort for our machine learning algorithms, we reduced the number of units for predicting $E(t)$ to 600 and 800 from our original 1000-output-layer-unit model. The number of hidden layer units were also reduced to 300, 400, 500, and 300, 450, and 600, while the size of the input layer remained the same. This reduction in the number of points had a negligible effect on predicting the frequency of $E(t)$, as shown in Figs. 9.8 a-c. In these plots, the predicted $E(t)$ is shifted with the proper phase to allow a more straightforward comparison. Both the frequency and amplitude agree well, and these results show that our cross-correlation approach is able to address the previous issues associated with the random phase of $E(t)$. Similar to the tests carried out in Section III, we also investigated the sensitivity of our results to the size of our training set and found that the accuracy of the machine-learned predictions decreased with the training set size. Specifically, when the training set was reduced to only 10,000 potentials, the R^2 values for predicting the resonance frequency and amplitude in the same validation set decreased to 0.89 and -0.02, respectively. Similar to our findings in Section III, these statistics showed that a sufficiently large training set was necessary to enable accurate machine-learned predictions for these optimal control fields, even in the time domain. Nevertheless, it is still worth noting that when the cross-correlated FNN approach was applied to $E(t)$ fields with large amplitudes (which were originally pruned from the test set as discussed in Section III), the machine learning algorithm was able to still accurately predict the resonance frequency, as shown in Fig. 9.8(d), which indicates the robustness of this approach.

To quantitatively demonstrate that the machine-learned and true $E(t)$ are in excellent agreement, a fast Fourier transform was applied to both of these data sets. The amplitude and frequency of $\sigma(\omega)$ were then compared for each data point in the validation and test set. As before, we computed R^2 values (cf. Eq. 9.5) for each of our 600-, 800-, and 1000-output-layer-unit models, and all of these configurations showed similar R^2 statistics, which are summarized in Table 9.4. The loss and R^2 of the training and validation sets were recorded every 1,000 epochs. The figures in the Supplementary Information show that the 1000-output-layer-unit DNN was sufficiently trained at ~50,000 epochs, and further training introduces overfitting (~30,000 and ~40,000 epochs were required for the 600- and 800-output-layer-unit DNN to reach a minimal loss). It is also worth mentioning that batch normalization and dropout approaches (among others) are machine-learning techniques that could also be used to prevent overfitting of the data; however, since we did not observe any severe overfitting of our training set, we did not employ these techniques in our work. Nevertheless, the R^2 for predicting the frequency on the validation set converged to an impressive ~0.95 value for all three models (cf. Supplementary Information), and both Figs. 9.8 and D.2 show that reducing the number of units in the layers of our cross-correlation neural network approach did not adversely affect its predictive performance.

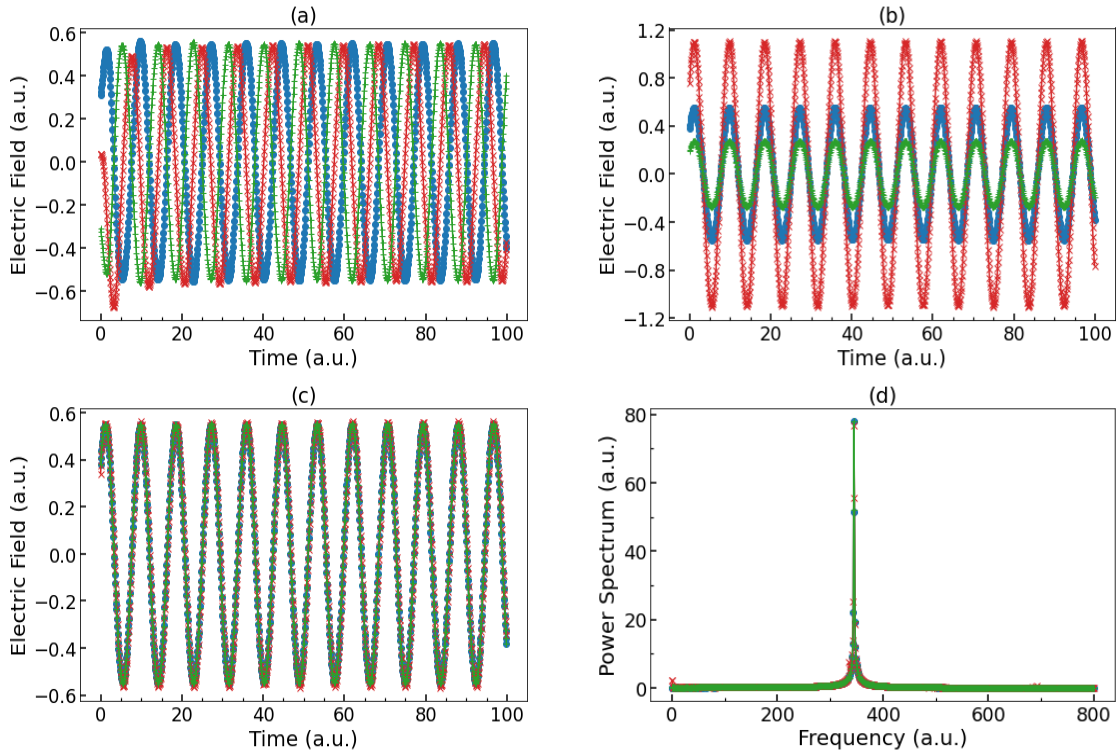


Figure 9.7: (a) Optimized electric field, $E(t)$, with various phase shifts, φ. The blue data points denote the optimized $E(t)$ obtained directly from the NIC-CAGE code. The red curve is the same $E(t)$ with a phase shift of $\pi/2$, and the green curve is $E(t)$ with a phase shift of π. When each of these electric fields are used as initial guesses for propagating the time-dependent Schrödinger equation, all of them gave a transition probability close to unity (which shows that the transition probability is insensitive to the phase, φ). (b) Optimized electric field, $E(t)$, with various amplitudes. The blue data points denote the optimized $E(t)$ obtained from the NIC-CAGE code, and the red and green curves denote the same $E(t)$ with amplitudes multiplied by 2 and 0.5, respectively. When each of these electric fields were used as initial guesses for time propagation, all of them reverted/converged back to the $E(t)$ with the original amplitude shown in panel (c), which indicates that the transition probability depends critically on the electric field amplitude. (d) Power spectra, $\sigma(\omega)$, of the various $E(t)$ fields depicted in (a), showing that they coincide with each other, as expected.

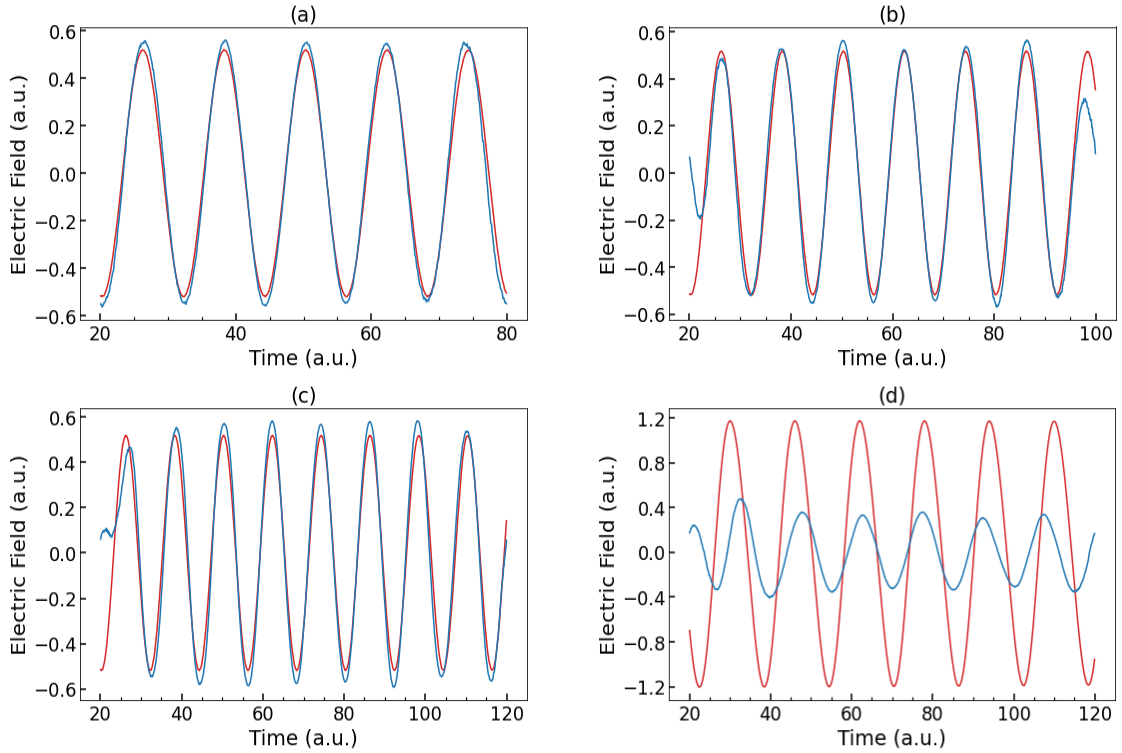


Figure 9.8: Comparisons of true (red) and machine-learned predicted (blue) $E(t)$ fields. The electric fields correspond to the same potential, but with (a) 600, (b) 800, and (c) 1,000 units. (d) True (red) and machine-learned (blue) $E(t)$ for a different potential characterized by a large amplitude. When the true $E(t)$ has a much larger amplitude, the machine learning algorithm is still able to accurately predict the resonance frequency but underestimates the amplitude.

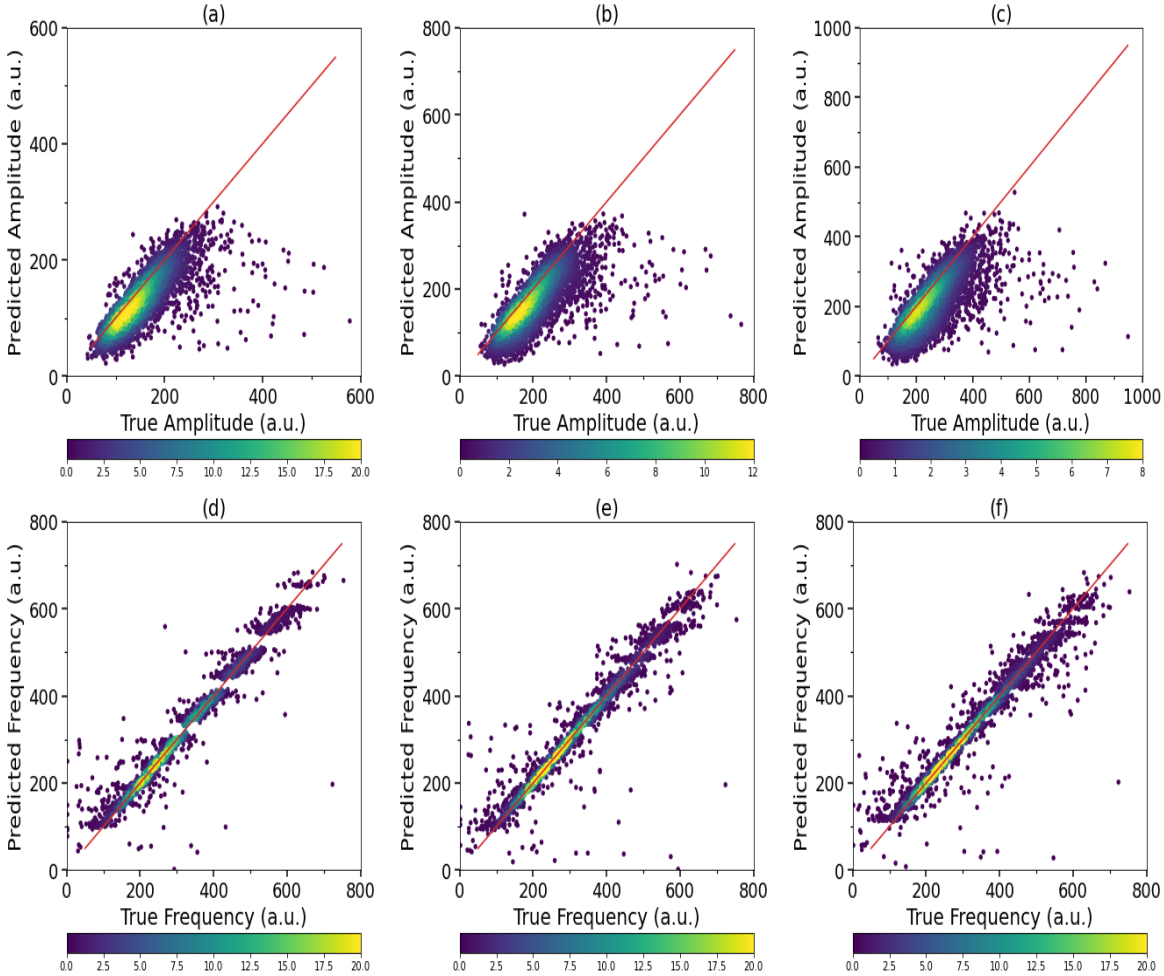


Figure 9.9: Density plots of the predicted and true amplitude for the (a) 600-, (b) 800-, and (c) 1000-output-layer-unit model, respectively. Density plots of the predicted and true resonance frequency for the (d) 600-, (e) 800-, and (f) 1000-output-layer-unit model, respectively. The diagonal line in each plot represents a perfect match between the machine-learned predictions and true values.

In addition, we also investigated the effect of using only 2 hidden layers to predict $E(t)$ with 1,000 units. As shown in the Supplementary Information, the density plot obtained with a 2-hidden-layer FNN was more sparse and spread out. Furthermore, the R^2 values for predicting the frequency on the validation set never exceeded 0.87, showing that the 2-hidden-layer neural network underfitted the data (cf. Supplementary Information).

On the other hand, we also recognized that increasing the number of hidden layers beyond 3 would possibly improve the accuracy of our neural network; however, this modification would also incur an immense computational cost. Specifically, training our 3-hidden-layer FNN to predict $E(t)$ required ~256 GB of RAM and 20 hours on high-performance GPUs. Further training with additional layers would require even more memory and GPU time, which we felt was impractical since we already obtained impressive R^2 values greater than 0.95 with our 3-hidden-layer FNN. As such, these benchmark configuration tests indicated that the use of 3 hidden layers in our neural network was sufficient and practical for accurate predictions. Most importantly, the density plots in Fig. D.2 show that both the resonance frequencies and amplitudes predicted by our cross-correlation neural network approach demonstrate an impressive agreement with the brute-force (and computationally expensive) quantum control results obtained with the NIC-CAGE program.

9.5 Conclusion

In conclusion, we have presented the first machine-learning effort for solving explicit time-dependent quantum control problems in reduced-dimensional chemical systems. Using a variety of deep neural networks, we have shown that the prediction of optimal control fields is an inverse problem that naturally lends itself to a machine-learning approach. Regarding efficiency, we have shown that our machine learning approach only requires knowledge of the potential, $V(x)$, to yield a reliable prediction of an optimal control field, $E(t)$. In other words, a user can simply input a variety of potentials into our neural network model to obtain optimal control fields without having to do a computa-

tionally expensive time-dependent quantum control calculation. In terms of accuracy, we have shown that deep neural networks can predict these optimal control fields within 96% accuracy by directly learning the underlying patterns between $V(x)$ and $E(t)$.

While this work focused on reduced-dimensional quantum systems, we anticipate that the machine-learning techniques explored in this work could be applied to other applications of increasing complexity. For example, we envision that some of the machine learning tactics used here could serve as a first step towards solving more complex quantum dynamics problems in higher dimensions. The use of reduced-dimensional techniques to address full 3D quantum dynamics problems is similar in spirit to ongoing efforts that use machine-learned, ground-state, 1D exchange-correlation functionals [389, 392] for full three-dimensional chemical problems [393]. Finally, we also anticipate that the machine learning techniques used here could be harnessed to predict optimal electric fields for other higher-lying transitions, which are known to exhibit more complex patterns in the time and frequency domains [388]. In particular, cross-correlation neural network approaches, which were used to overcome problems associated with the random phase of $E(t)$, could be useful in (1) predicting optimal electric fields for other higher-energy excitations in the time domain or (2) enabling the prediction of the *full* absorption/emission spectra of molecules since the absorption spectra are merely the Fourier transform of $E(t)$. Taken together, these machine learning techniques show a promising path towards cost-effective statistical approaches for designing control fields that enable desired transitions in quantum dynamical systems.

Chapter 10

Conclusions

We have performed large-scale quantum calculations via DFT and DFTB to evaluate the electronic, optical, and thermodynamic properties of various complex chemical, biochemical, and material systems. Specifically, by integrating DFTB in CASM, we showed that DFTB could be used as an alternative approach to DFT for calculating and prescreening formation energies and convex hulls. Finally, in the end, we showed that by using hardware accelerators such as GPUs with DFTB, large-scale metadynamics calculations on complex chemical systems could be performed in a relatively short timeframe.

In particular, Chapter 3 demonstrates the role of DFT calculations in characterizing the transport properties in doped carbon nanotubes (CNTs). Specifically, using DFT and theoretical calculations, we showed that doping CNTs with Cr atoms results in a high density of states (DOS) near the Fermi level, which facilitates opening new transmission pathways for electrons' conduction resulting in enhanced conductivity. This study has important implications for the field of anisotropic conduction, which has garnered significant

scientific and technological interests.

In chapter 4, we performed an extensive analysis of various range-separated DFT methods along with the wavefunctions-based MP2, CCSD(T), and CCDSD(T)-F12 approaches for predicting the linear polarizability (α) and second hyperpolarizability (γ) of π-conjugated streptocyanine oligomers. We found that the electronic ground states for several streptocynanines are not closed-shell singlets (contrary to the previous studies on these conjugated systems). Our new benchmark found that broken-symmetry LC-ωPBE $_{\omega=0.33}$ functional gives the most accurate results among various DFT methods for predicting both α and γ.

In chapters 5 and 6, using time-dependent DFT calculations on peptides, we found the electronic transitions that resulted in the experimentally observed UVCD spectra of polyproline II and β-strand conformations. Our calculations revealed multitudes of electronic transitions governed by configurational interactions between multiple molecular orbitals transitions of comparable energy. We also showed that reproducing the CD spectra of polyproline II and β-strand conformations requires the explicit consideration of water molecules.

In chapter 7, we introduced DFTB to overcome the computational costs associated with DFT calculations. Specifically, we interfaced DFTB with the CASM software package to calculate the formation energies and convex hulls efficiently. Using SiC and ZnO as representative examples, we show that DFTB gives accurate results and can be used as an efficient computational approach for calculating and screening electronic properties.

In chapter 8, we used the GPU-enhanced DFTB approach for the efficient prediction of free energies of complex biochemical systems. Specifically, we showed that converged free energy surfaces of complex systems could be accurately predicted/described at the quantum level using the DFTB approach. Finally, in chapter 9, as an example of a machine-learning application in accelerating quantum computation, we design the neural network to obtain the optimal electric field required for electron transitions in quantum systems.

In conclusion, this thesis demonstrates the importance of performing quantum calculations to characterize complex chemical and material systems accurately. We emphasize the importance of DFTB in performing these quantum calculations relatively quickly (compared to DFT) without compromising the accuracy of the results. DFTB, in conjunction with GPUs, provides an accelerated approach to performing large-scale quantum calculations of complex systems. As such, GPU-DFTB provides a powerful tool for exploring the electronic and thermodynamic properties of complex materials and chemical systems, with potential applications in materials science, physics, chemistry, and biochemistry.

Appendix A

A New Interpretation of the Structure and Solvent Dependence of the Far UV Circular Dichroism Spectrum of Short Oligopeptides

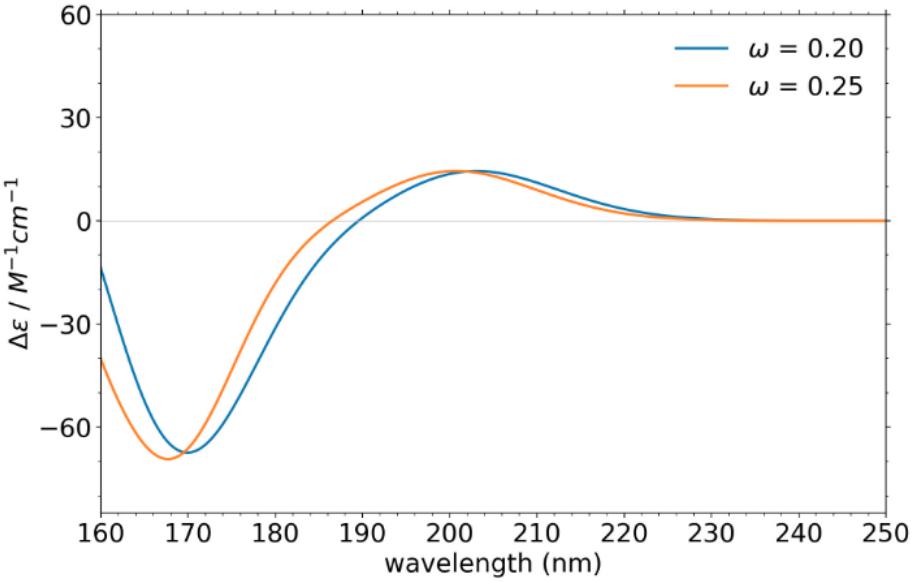


Figure A.1: ECD spectra of explicitly-solvated GAG in the optimized pPII conformation calculated with different range-separation values of $\omega = 0.20$ and 0.25.

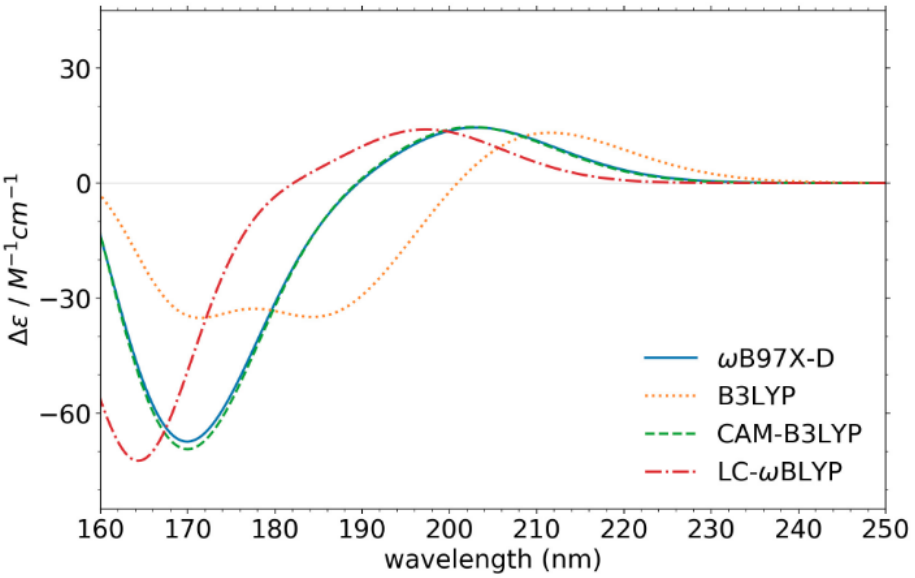


Figure A.2: ECD spectra of explicitly-solvated GAG in the optimized pPII conformation calculated with different exchange-correlation functionals

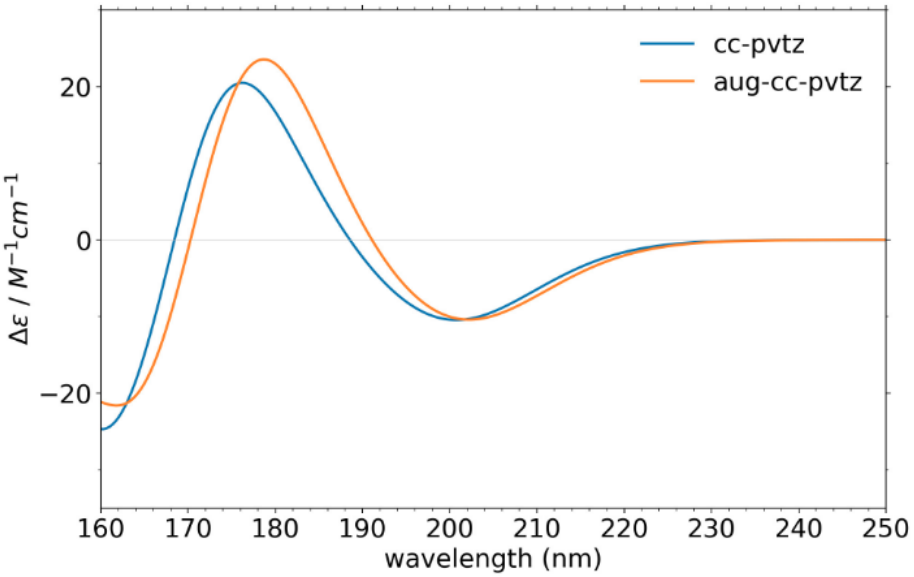


Figure A.3: ECD spectra of explicitly-solvated GAG in the optimized β-strand conformation calculated with different basis sets.

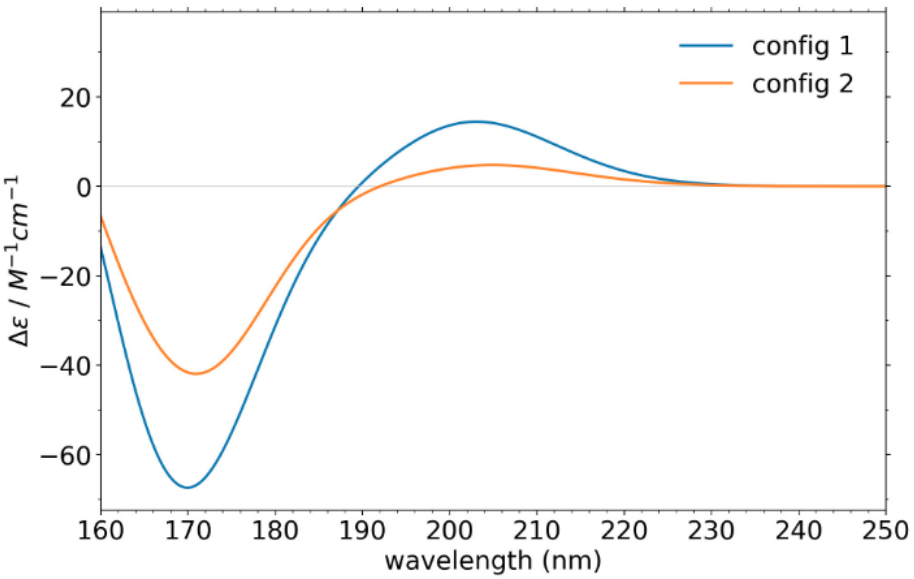


Figure A.4: ECD spectra of explicitly-solvated GAG in the optimized pPII conformation with different configurations of explicit water molecules

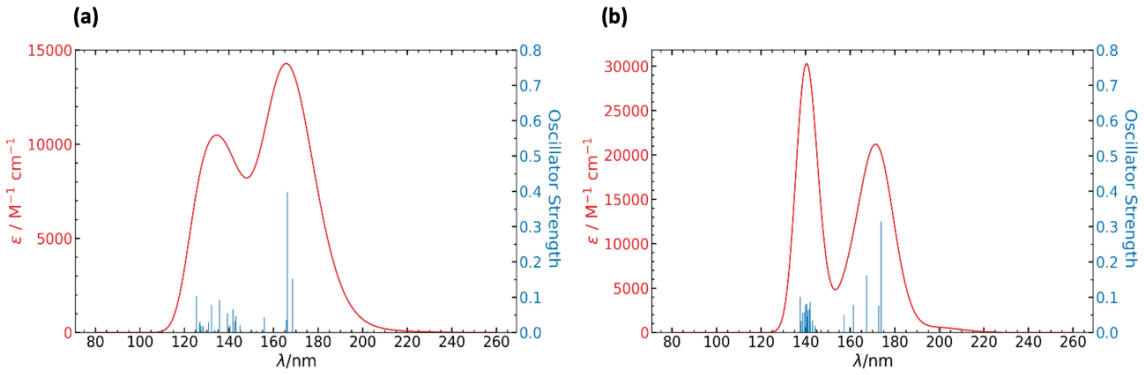


Figure A.5: Absorption spectra of GAG in the β conformation calculated at the ωB97XD/cc-pVTZ level of theory in (a) implicit and (b) explicit water.

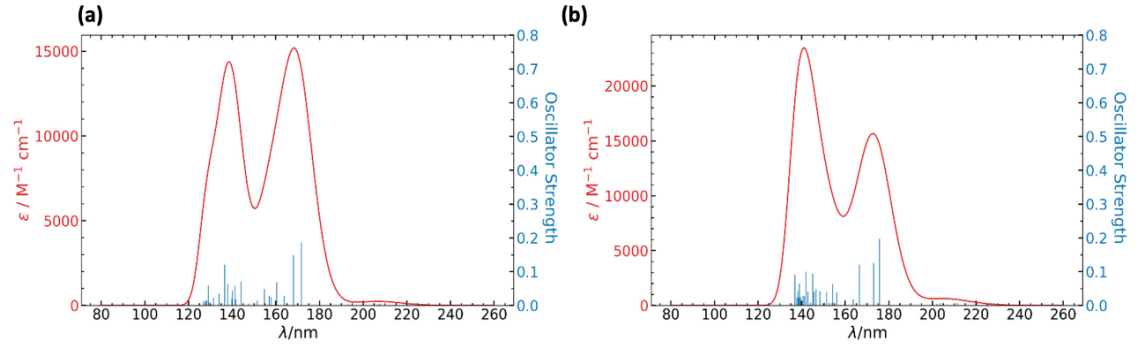


Figure A.6: Absorption spectra of GAG in the pPII conformation calculated at the ωB97XD/cc-pVTZ level of theory in (a) implicit and (b) explicit water.

Appendix B

Water-Mediated Electronic Structure of Oligopeptides Probed by Their UV Circular Dichroism, Absorption Spectra, and Time-Dependent DFT Calculations

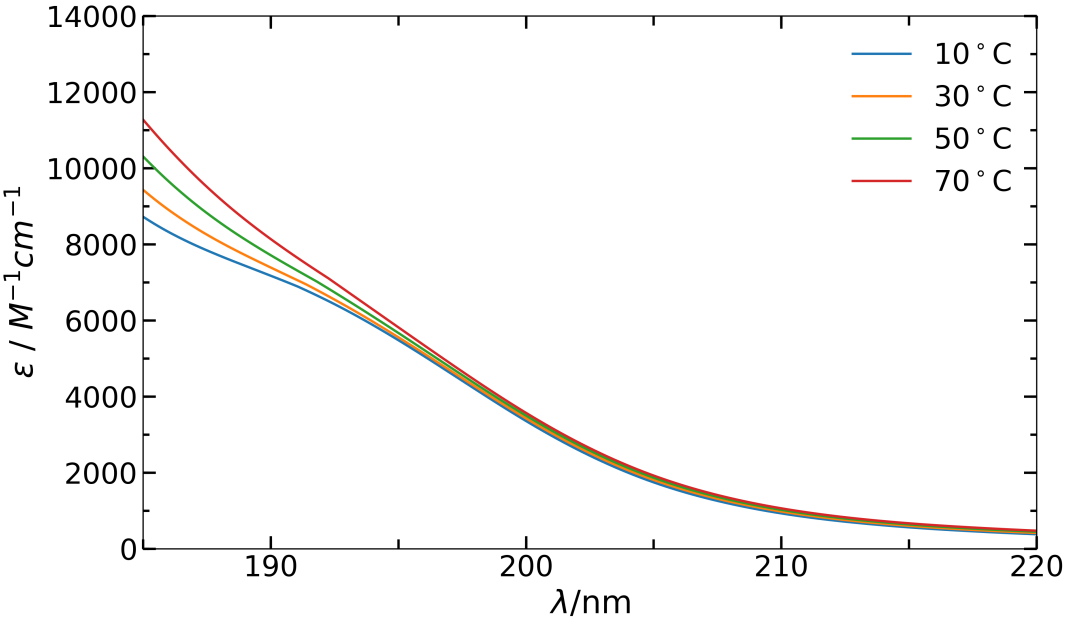


Figure B.1: UV absorption spectra of cationic GAG in water at pH = 2 without background subtraction taken at 10, 30, 50, and 70°C.

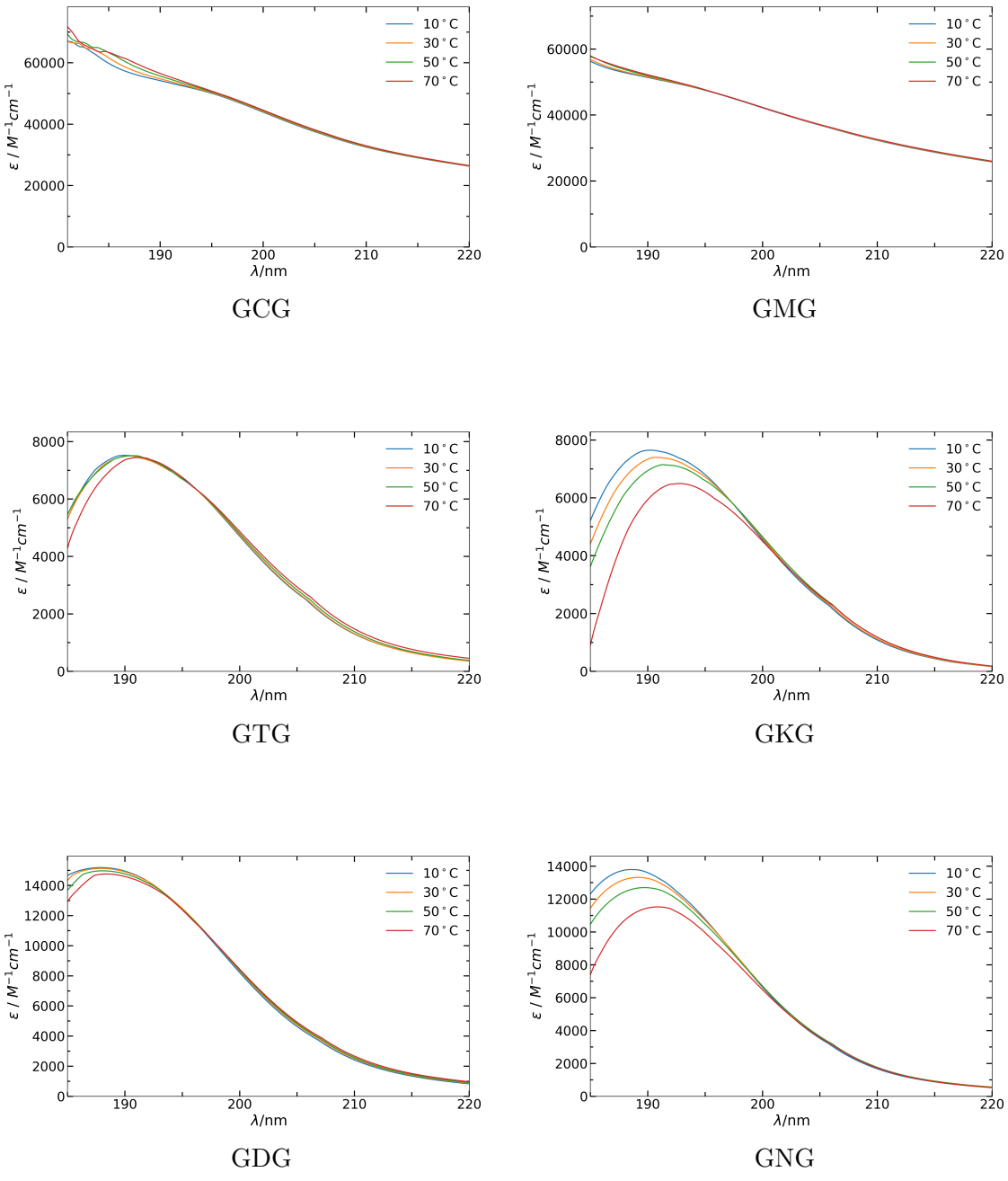


Figure B.2: Background-corrected UV absorption spectra of GxG (x = C, M, T, K, D, and N) taken at 10, 30, 50, and 70°C. The spectra for GKG, GNG, and GRG were measured with a pathlength of 50 μm, whereas a pathlength of 100 μm was used for the GDG spectra. The peptide concentration was 10 mM.

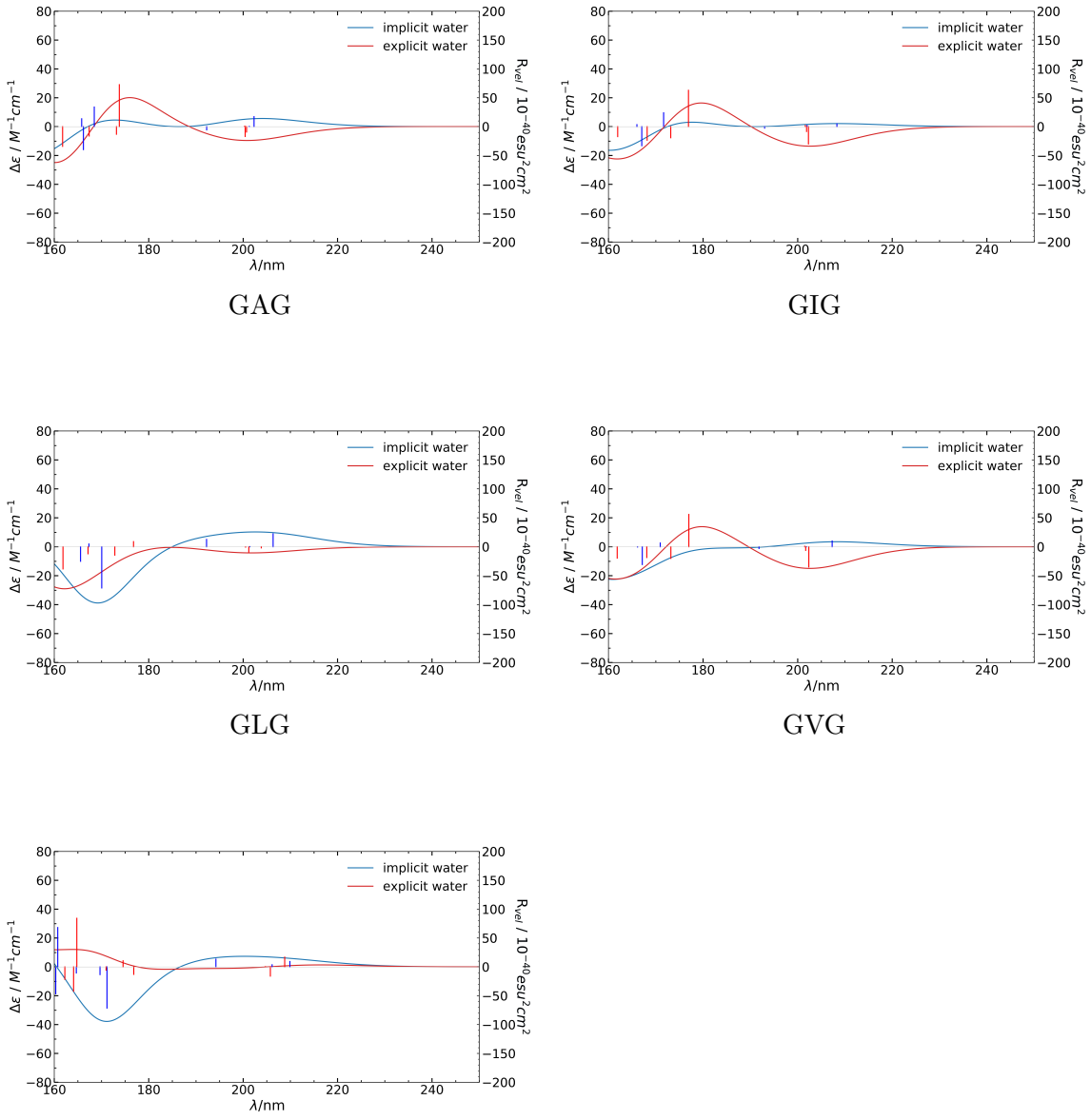


Figure B.3: Calculated UVCD spectra of GxG (x = A, I, L, V, and R) in the β conformation with implicit and explicit water. The sticks in the respective figures represent the rotatory strength (velocity) of the individual electronic transitions.

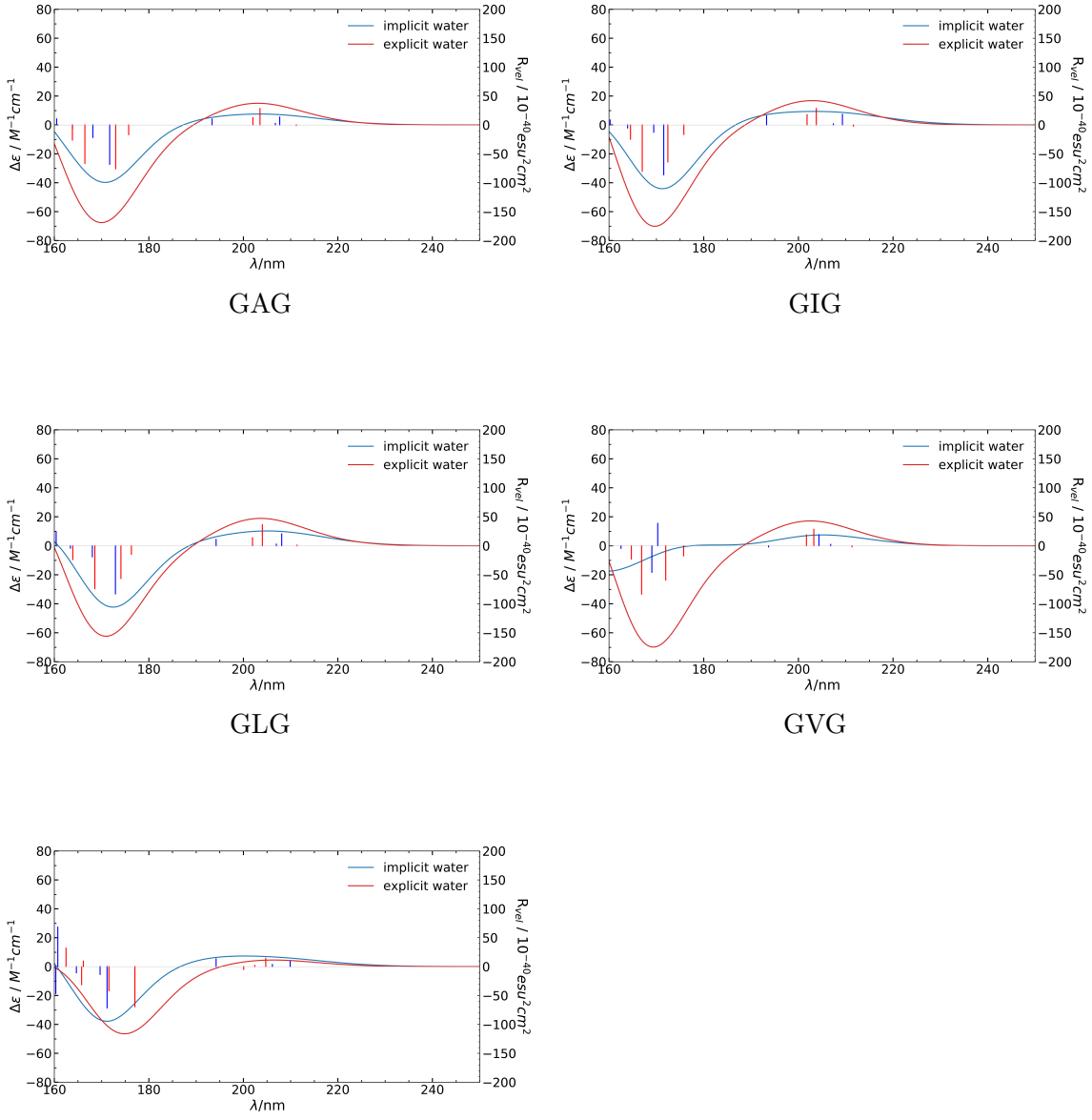
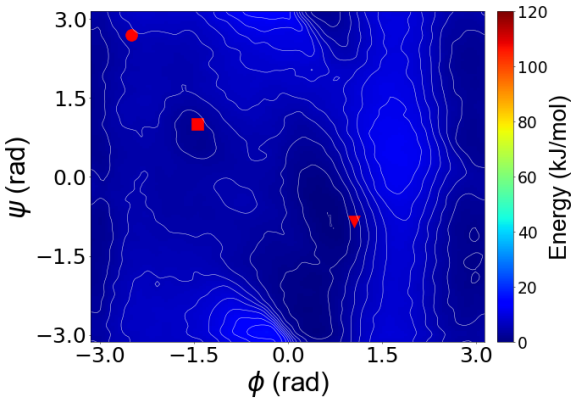


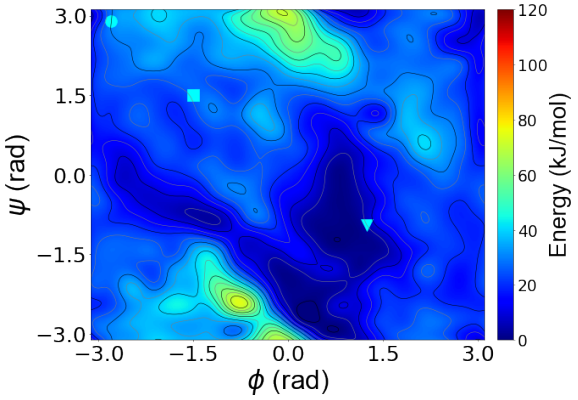
Figure B.4: Calculated UVCD spectra of GxG (x = A, I, L, V, and R) in the pPII conformation with implicit and explicit water. The sticks in the respective figures represent the rotatory strength (velocity) of the individual electronic transitions.

Appendix C

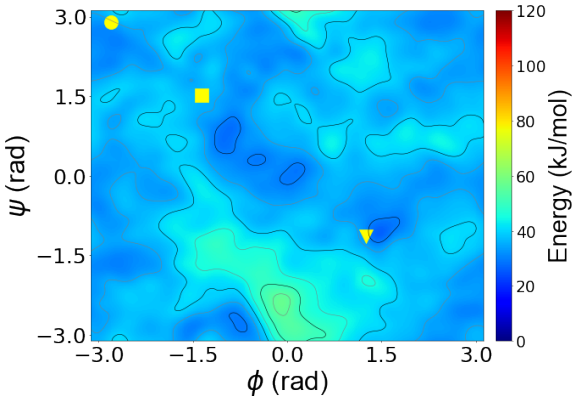
GPU-Enhanced DFTB for Efficiently Predicting Free Energies of Biochemical Systems



(a) Amber99sb

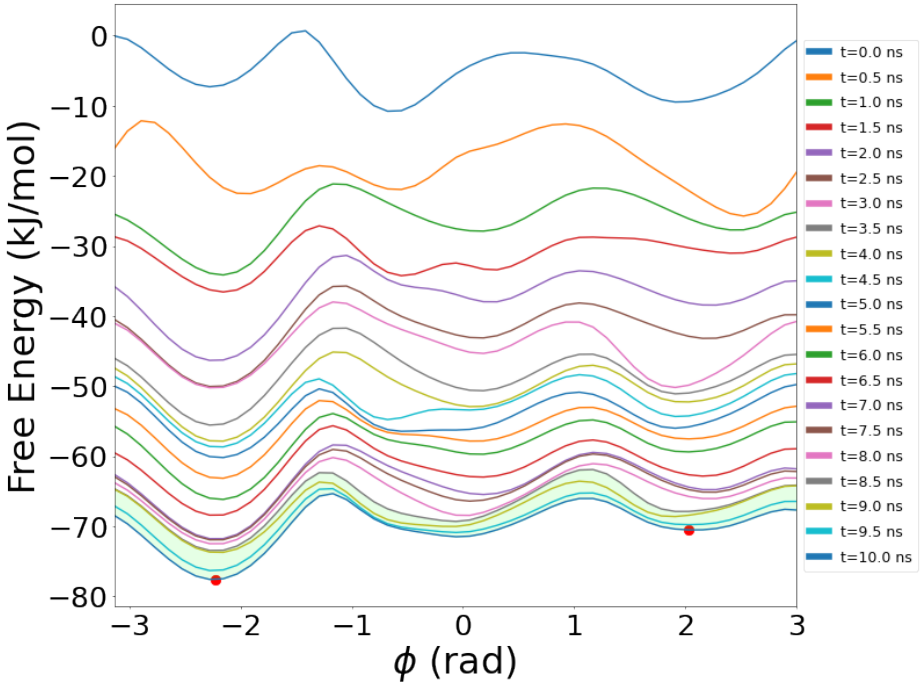


(b) PBE0

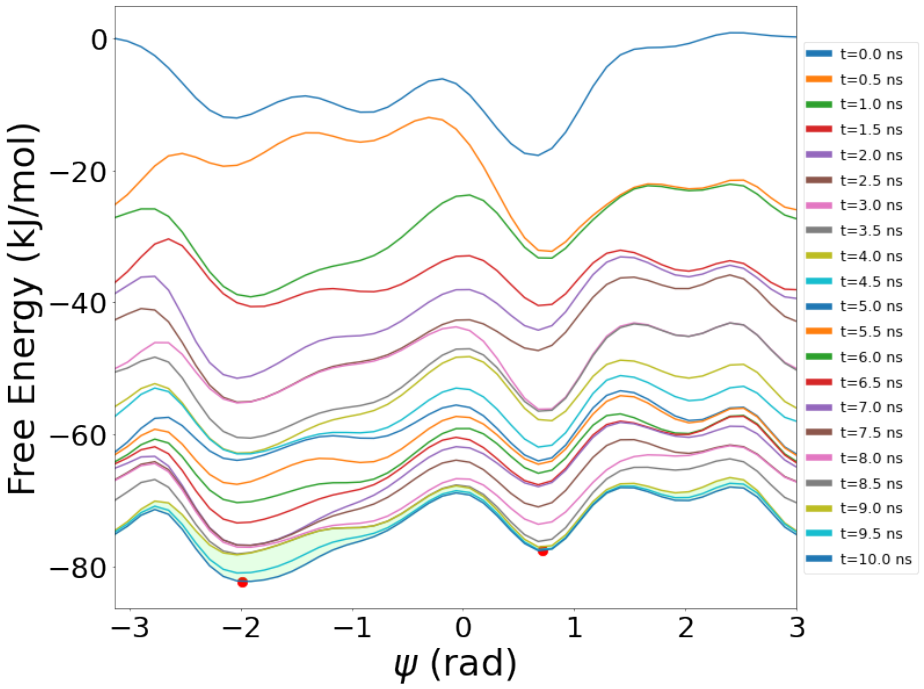


(c) DFTB3

Figure C.1: Comparison of the entropic term, $T\Delta S$, of alanine dipeptide as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (a) classical MD with the Amber99sb force field, (b) DFT-PBE0 calculations, and (c) SCC-DFTB3 calculations. The red, cyan, and yellow points in panels (a), (b), and (c) represent the local minima on the FES (cf. Fig. 8.2) obtained using the Amber99sb force field, PBE0, and SCC-DFTB3, respectively. ●, ■, and ▼ denote the β, C_{7eq}, and C_{7ax} metastable structures, respectively.

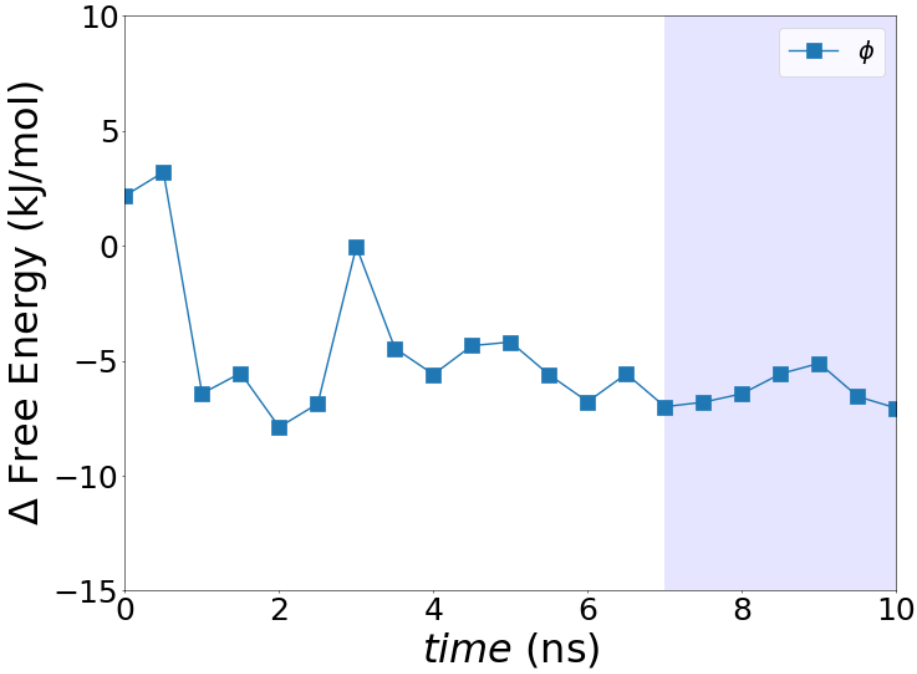


(a)

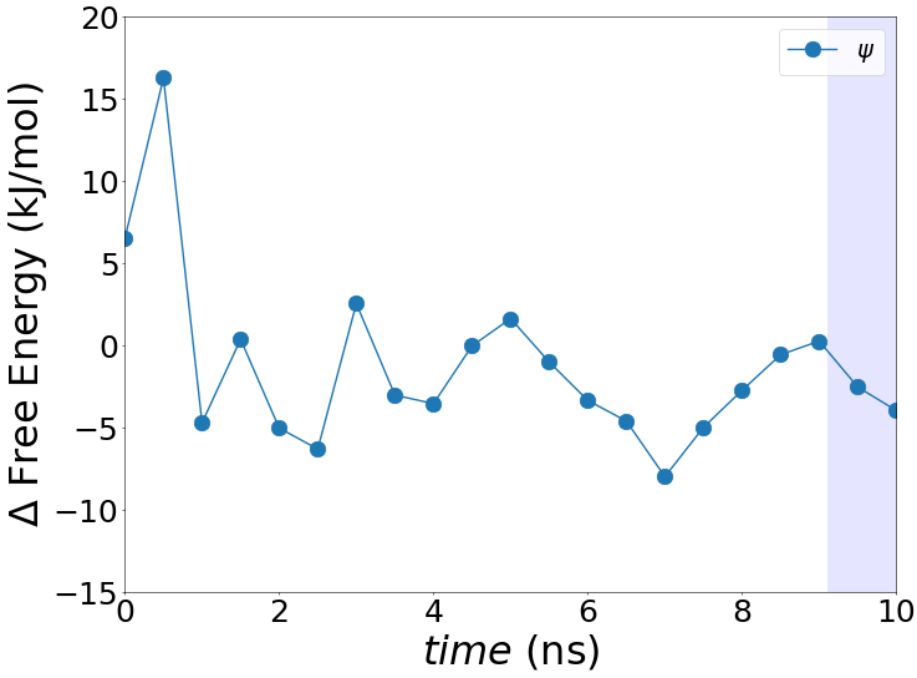


(b)

Figure C.2: Convergence of the free energy surface of remdesivir as a function of time with respect to the dihedral angles, (a) ϕ and (b) ψ, using well-tempered metadynamics. The light-green shaded region in each plot indicates converged values. The red dots in panels (a) and (b) represent the local minima at which the free energy difference is calculated in Fig. C.3.



(a)

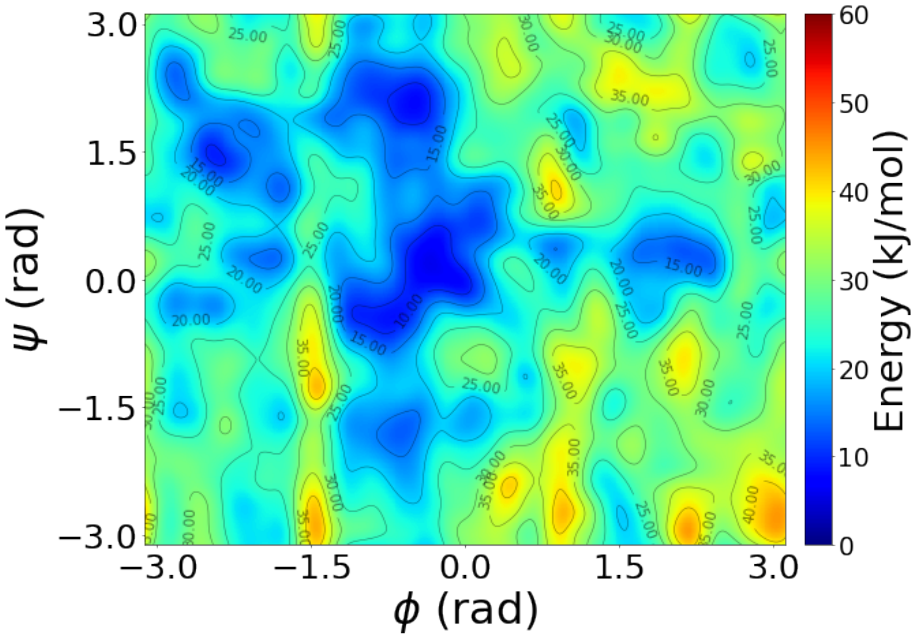


(b)

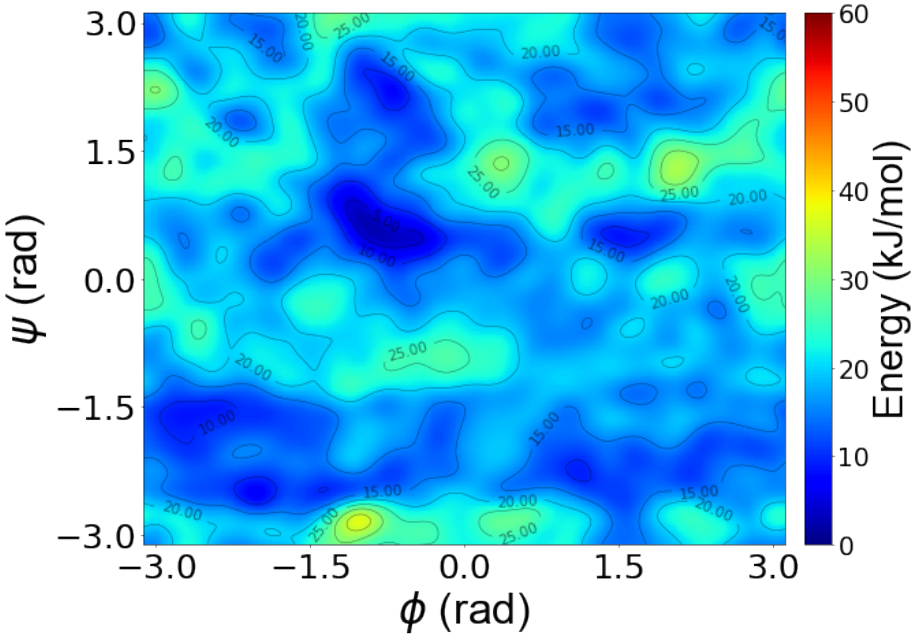
Figure C.3: (a) Free energy difference between the basins near ϕ = -2 and 2 radians (shown as red dots at t = 10 ns Fig. C.2(a)) as a function of simulation time. (b) Free energy difference between the basins near ψ = -2 and 1 radians (shown as red dots at t = 10 ns in Fig. C.2(b)) as a function of simulation time. The light-blue shaded region in each plot indicates converged values.

Table C.1: Relative energies of local minima (points A, B, C, and D in Fig. 8.5) calculated at the PBE0 and B3LYP levels of theory using the 6-311++g(d,p) basis set.

	PBE0 (kJ/mol)	B3LYP (kJ/mol)
E_A	1.26	1.49
E_B	1.46	2.69
E_C	0	0
E_D	3.22	3.27



(a) Amber-ff19SB



(b) DFTB3

Figure C.4: Comparison of the entropic term, $T\Delta S$, of remdesivir as a function of the backbone dihedral angles, ϕ and ψ, obtained from well-tempered metadynamics simulations using (a) classical MD from the Amber-ff19SB force field and (b) SCC-DFTB3 calculations.

Some of the codes used in this work can be downloaded from Github.

Appendix D

Harnessing Deep Neural Networks to Solve Inverse Problems in Quantum Dynamics: Machine-Learned Predictions of Time-Dependent Optimal Control Fields

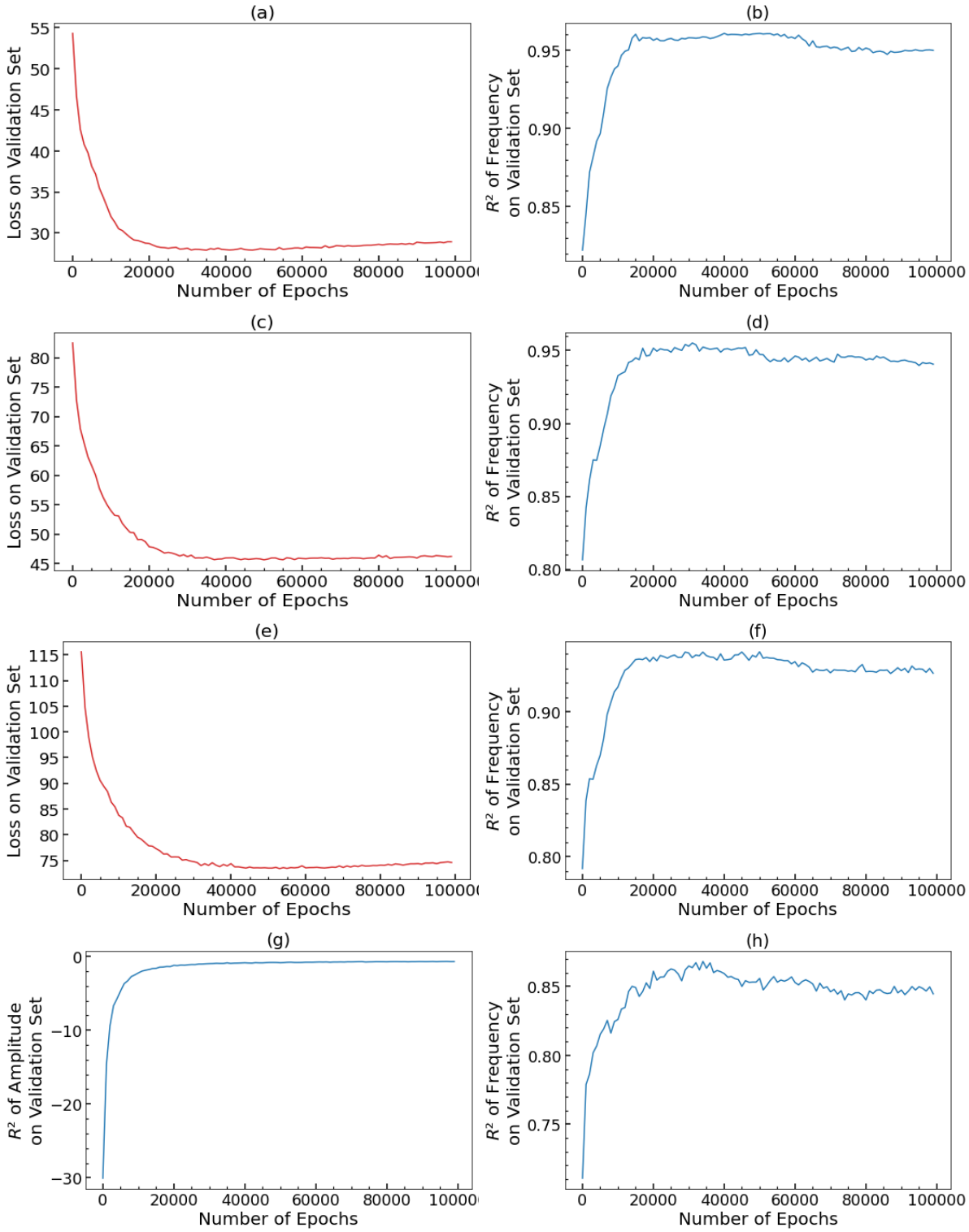


Figure D.1: Plots of the loss vs. the number of epochs for the (a) 600-, (c) 800-, and (e) 1000-output-layer-unit model for the validation dataset. Plots of R^2 values for the predicted frequency vs. the number of epochs for the (b) 600-, (d) 800-, and (f) 1000-output-layer-unit model for the validation dataset. (g) Plot of R^2 values for the predicted amplitude vs. the number of epochs for the 1000-output-layer-unit model for the validation dataset. Using the R^2 definition in the main text, the amplitude R^2 becomes negative-valued, which also manifests itself in the the 600- and 800-output-layer-unit models (not plotted here). (h) Plot of R^2 values for the predicted frequency vs. the number of epochs for the 1000-output-layer-unit model when only 2 hidden layers are used.

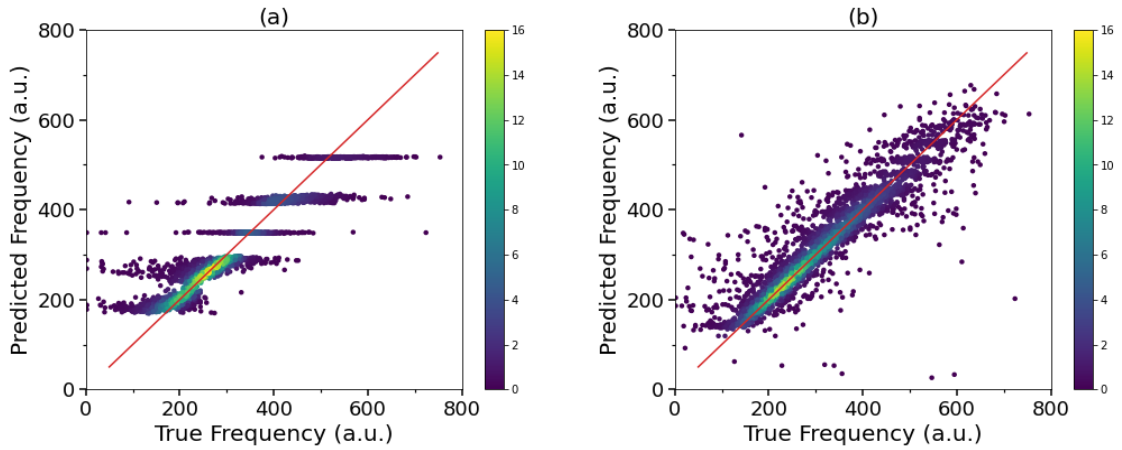


Figure D.2: Density plot of the predicted and true resonance frequencies for the (a) 600-output-layer-unit model for the unscaled amplitude (i.e., when the amplitude is not multiplied by 80), and (b) 1,000-output-layer-unit model with only 2 hidden layers. The diagonal line in each plot represents a perfect match between the machine-learned predictions and the true values. The vertical color bar in each sub-plot indicates the density of the data points.

Bibliography

[1] **Anshuman Kumar**, Reinhard Schweitzer-Stenner, and Bryan M Wong. A new interpretation of the structure and solvent dependence of the far uv circular dichroism spectrum of short oligopeptides. *Chemical Communications*, 55(40):5701–5704, 2019.

[2] **Anshuman Kumar**, Siobhan E Toal, David DiGuiseppi, Reinhard Schweitzer-Stenner, and Bryan M Wong. Water-mediated electronic structure of oligopeptides probed by their uv circular dichroism, absorption spectra, and time-dependent dft calculations. *The Journal of Physical Chemistry B*, 124(13):2579–2590, 2020.

[3] **Anshuman Kumar**, Xian Wang, Christian R Shelton, and Bryan M Wong. Harnessing deep neural networks to solve inverse problems in quantum dynamics: machine-learned predictions of time-dependent optimal control fields. *Physical Chemistry Chemical Physics*, 22(40):22889–22899, 2020.

[4] **Anshuman Kumar**, Zulfikhar A. Ali, and Bryan M. Wong. Efficient predictions of formation energies and convex hulls from density functional tight binding calculations. *J. Mater. Sci. Technol.*, 141:236–244, 2023.

[5] **Anshuman Kumar**, Pablo R Arantes, Aakash Saha, Giulia Palermo, and Bryan M Wong. Gpu-enhanced dftb metadynamics for efficiently predicting free energies of biochemical systems. *Molecules*, 28(3):1277, 2023.

[6] Mingguang Chen, Wangxiang Li, **Anshuman Kumar**, Guanghui Li, Mikhail E Itkis, Bryan M Wong, and Elena Bekyarova. Covalent atomic bridges enable unidirectional enhancement of electronic transport in aligned carbon nanotubes. *ACS Applied Materials & Interfaces*, 11(21):19315–19323, 2019.

[7] Lihua Xu, **Anshuman Kumar**, and Bryan M Wong. Linear polarizabilities and second hyperpolarizabilities of streptocyanines: Results from broken-symmetry dft and new ccsd (t) benchmarks. *Journal of Computational Chemistry*, 39(28):2350–2359, 2018.

[8] Sarah I Allec, **Anshuman Kumar**, and Bryan M Wong. Linear-response and real-time, time-dependent density functional theory for predicting optoelectronic properties of dye-sensitized solar cells. In *Dye-Sensitized Solar Cells*, pages 171–201. Elsevier, 2019.

[9] Jin Chong Tan and Anthony K Cheetham. Mechanical properties of hybrid inorganic–organic framework materials: establishing fundamental structure–property relationships. *Chemical Society Reviews*, 40(2):1059–1080, 2011.

[10] Mark S Hybertsen and Latha Venkataraman. Structure–property relationships in atomic-scale junctions: Histograms and beyond. *Accounts of chemical research*, 49(3):452–460, 2016.

[11] Xi Wu, Feiyu Kang, Wenhui Duan, and Jia Li. Density functional theory calculations: A powerful tool to simulate and design high-performance energy storage and conversion materials. *Progress in Natural Science: Materials International*, 29(3):247–255, 2019.

[12] Jens K Nørskov, Frank Abild-Pedersen, Felix Studt, and Thomas Bligaard. Density functional theory in surface chemistry and catalysis. *Proceedings of the National Academy of Sciences*, 108(3):937–943, 2011.

[13] Puru Jena and Qiang Sun. Theory-guided discovery of novel materials. *The Journal of Physical Chemistry Letters*, 12(28):6499–6513, 2021.

[14] Dimple Kumari, Prabhat Ranjan, and Tanmoy Chakraborty. A computational study of potential therapeutics for covid-19 invoking conceptual density functional theory. *Structural Chemistry*, 33(6):2195–2204, 2022.

[15] Hyuna Kwon, **Anshuman Kumar**, and Bryan M Wong. Electron/hole mobilities of periodic dna and nucleobase structures from large-scale dft calculations. *In Submission.*, 2023.

[16] Steve Yang, **Anshuman Kumar**, and Bryan M Wong. Beyond bandgaps: Dimensionality determines efficacy of conductivity on 1d and 2d π-conjugated polymers. *In preparation.*, 2023.

[17] Xian Wang, Mahmut Okyay, **Anshuman Kumar**, and Bryan M Wong. Accelerating quantum optimal control of multi-qubit systems with symmetry-based hamiltonian transformations. *In preparation.*, 2023.

[18] José M. Rodríguez-Borbón, Zulfikhar A. Ali, **Anshuman Kumar**, Prithviraj Yuvaraj, and Bryan M Wong. Enhanced computation of eigenvalues and eigenvectors on field programmable gate arrays (fpgas) for fast and energy-efficient ab initio molecular dynamics simulations. *In preparation.*, 2023.

[19] Akber Raza, Chengkuan Hong, Xian Wang, **Anshuman Kumar**, Christian R Shelton, and Bryan M Wong. Nic-cage: An open-source software package for predicting optimal control fields in photo-excited chemical systems. *Computer Physics Communications*, 258:107541, 2021.

[20] Yongho Joo, Lifeng Huang, Naresh Eedugurala, Alexander E London, **Anshuman Kumar**, Bryan M Wong, Bryan W Boudouris, and Jason D Azoulay. Thermoelectric performance of an open-shell donor–acceptor conjugated polymer doped with a radical-containing small molecule. *Macromolecules*, 51(10):3886–3894, 2018.

[21] Mark E. Casida. *Time-Dependent Density Functional Response Theory for Molecules*, pages 155–192.

[22] Kieron Burke and Lucas O Wagner. Dft in a nutshell. *International Journal of Quantum Chemistry*, 113(2):96–101, 2013.

[23] Kieron Burke. Perspective on density functional theory. *The Journal of Chemical Physics*, 136(15):150901, 2012.

[24] Klaus Capelle. A bird's-eye view of density-functional theory. *Brazilian journal of physics*, 36:1318–1343, 2006.

[25] Carsten A Ullrich. Time-dependent density-functional theory: concepts and applications. 2011.

[26] Miguel AL Marques, Neepa T Maitra, Fernando MS Nogueira, Eberhard KU Gross, and Angel Rubio. *Fundamentals of time-dependent density functional theory*, volume 837. Springer, 2012.

[27] Mark E. Casida. Time-dependent density-functional theory for molecules and molecular solids. *Journal of Molecular Structure: THEOCHEM*, 914(1):3–18, 2009. Time-dependent density-functional theory for molecules and molecular solids.

[28] Ute Werner, Roland Mitric, Toshinori Suzuki, and Vlasta Bonacic-Koutecky. Nonadiabatic dynamics within the time dependent density functional theory: Ultrafast photodynamics in pyrazine. *Chemical Physics*, 349(1):319–324, 2008. Electron Correlation and Molecular Dynamics for Excited States and Photochemistry.

[29] Michael S Strano, Christopher A Dyke, Monica L Usrey, Paul W Barone, Mathew J Allen, Hongwei Shan, Carter Kittrell, Robert H Hauge, James M Tour, and Richard E Smalley. Electronic structure control of single-walled carbon nanotube functionalization. *Science*, 301(5639):1519–1522, 2003.

[30] Christopher A Dyke and James M Tour. Covalent functionalization of single-walled carbon nanotubes for materials applications. *The Journal of Physical Chemistry A*, 108(51):11151–11159, 2004.

[31] Dimitrios Tasis, Nikos Tagmatarchis, Alberto Bianco, and Maurizio Prato. Chemistry of carbon nanotubes. *Chemical reviews*, 106(3):1105–1136, 2006.

[32] Morinobu Endo, Michael S Strano, and Pulickel M Ajayan. Potential applications of carbon nanotubes. In *Carbon nanotubes*, pages 13–62. Springer, 2007.

[33] Matthew J Allen, Vincent C Tung, and Richard B Kaner. Honeycomb carbon: a review of graphene. *Chemical reviews*, 110(1):132–145, 2010.

[34] Guosong Hong, Shuo Diao, Alexander L Antaris, and Hongjie Dai. Carbon nanomaterials for biological imaging and nanomedicinal therapy. *Chemical reviews*, 115(19):10816–10906, 2015.

[35] Bin Wang, Chuangang Hu, and Liming Dai. Functionalized carbon nanotubes and graphene-based materials for energy storage. *Chemical Communications*, 52(100):14350–14360, 2016.

[36] AV Krasheninnikov and FJNM Banhart. Engineering of nanostructured carbon materials with electron or ion beams. *Nature materials*, 6(10):723–733, 2007.

[37] M Terrones, F Banhart, N Grobert, J-C Charlier, H Terrones, and PM Ajayan. Molecular junctions by joining single-walled carbon nanotubes. *Physical review letters*, 89(7):075505, 2002.

[38] Çağlar Ö Girit and Alex Zettl. Soldering to a single atomic layer. *Applied Physics Letters*, 91(19):193512, 2007.

[39] Yonggang Yao, Kun Kelvin Fu, Shuze Zhu, Jiaqi Dai, Yanbin Wang, Glenn Pastel, Yanan Chen, Tian Li, Chengwei Wang, Teng Li, et al. Carbon welding by ultrafast joule heating. *Nano letters*, 16(11):7282–7289, 2016.

[40] Santanu Sarkar, Matthew L Moser, Xiaojuan Tian, Xixiang Zhang, Yas Fadel Al-Hadeethi, and Robert C Haddon. Metals on graphene and carbon nanotube surfaces: from mobile atoms to atomtronics to bulk metals to clusters and catalysts. *Chemistry of Materials*, 26(1):184–195, 2014.

[41] Irina Kalinina, Elena Bekyarova, Santanu Sarkar, Feihu Wang, Mikhail E Itkis, Xiaojuan Tian, Sandip Niyogi, Neetu Jha, and Robert C Haddon. Hexahapto-metal complexes of single-walled carbon nanotubes. *Macromolecular Chemistry and Physics*, 213(10-11):1001–1019, 2012.

[42] Feihu Wang, Mikhail E Itkis, Elena B Bekyarova, Xiaojuan Tian, Santanu Sarkar, Aron Pekker, Irina Kalinina, Matthew L Moser, and Robert C Haddon. Effect of first row transition metals on the conductivity of semiconducting single-walled carbon nanotube networks. *Applied Physics Letters*, 100(22):223111, 2012.

[43] Andreas Thess, Roland Lee, Pavel Nikolaev, Hongjie Dai, Pierre Petit, Jerome Robert, Chunhui Xu, Young Hee Lee, Seong Gon Kim, Andrew G Rinzler, et al. Crystalline ropes of metallic carbon nanotubes. *science*, 273(5274):483–487, 1996.

[44] A Haaland, R Blomstrand, M Reunanen, H Kyyhkynen, RA Hoffman, and A Westerdahl. The molecular structure of gaseous dibenzene chromium (c6h6) 2cr. *Acta Chem. Scand*, 19(1):41–46, 1965.

[45] Keith Prout, A Gourdon, C Couldwell, B Meunier, Fang Ming Miao, and J Woolcock. The structures of mono-and bis-arene tungsten derivatives:(i) bis (η6-toluene) tungsten,(ii) hydridobis (η6-toluene) tungsten hexafluorophosphate,(iii) hydridobis (η6-monofluorobenzene) tungsten hexafluorophosphate,(iv) dichlorohydrido (η6-toluene) bis (trimethylphosphine) tungsten hexafluorophosphate,(v) bis (acetonitrile)(η3-allyl)(η6-toluene) tungsten hexafluorophosphate. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 38(2):456–462, 1982.

[46] Holger Braunschweig, Nele Buggisch, Ulli Englert, Melanie Homberger, Thomas Kupfer, Dirk Leusser, Matthias Lutz, and Krzysztof Radacki. Dilithiation of bis (benzene) molybdenum and subsequent isolation of a molybdenum-containing paracyclophane. *Journal of the American Chemical Society*, 129(15):4840–4846, 2007.

[47] RS Lee, HJ Kim, JE Fischer, A Thess, and Richard E Smalley. Conductivity enhancement in single-walled carbon nanotube bundles doped with k and br. *Nature*, 388(6639):255–257, 1997.

[48] L Grigorian, GU Sumanasekera, AL Loper, S Fang, JL Allen, and PC Eklund. Transport properties of alkali-metal-doped single-wall carbon nanotubes. *Physical Review B*, 58(8):R4195, 1998.

[49] Marc Bockrath, J Hone, A Zettl, Paul L McEuen, Andrew G Rinzler, and Richard E Smalley. Chemical doping of individual semiconducting carbon-nanotube ropes. *Physical Review B*, 61(16):R10606, 2000.

[50] John E Fischer. Chemical doping of single-wall carbon nanotubes. *Accounts of chemical research*, 35(12):1079–1086, 2002.

[51] Xiaojuan Tian, Matthew L Moser, Aron Pekker, Santanu Sarkar, Jason Ramirez, Elena Bekyarova, Mikhail E Itkis, and Robert C Haddon. Effect of atomic interconnects on percolation in single-walled carbon nanotube thin film networks. *Nano letters*, 14(7):3930–3937, 2014.

[52] Feihu Wang, Mikhail E Itkis, Elena Bekyarova, Santanu Sarkar, Xiaojuan Tian, and Robert C Haddon. Solid-state bis-hexahapto-metal complexation of single-walled carbon nanotubes. *Journal of Physical Organic Chemistry*, 25(7):607–610, 2012.

[53] Walt A De Heer, WS Bacsa, A Chatelain, T Gerfin, R Humphrey-Baker, L Forro, and D Ugarte. Aligned carbon nanotube films: production and optical and electronic properties. *Science*, 268(5212):845–847, 1995.

[54] Xiaowei He, Weilu Gao, Lijuan Xie, Bo Li, Qi Zhang, Sidong Lei, John M Robinson, Erik H Hároz, Stephen K Doorn, Weipeng Wang, et al. Wafer-scale monodomain films of spontaneously aligned single-walled carbon nanotubes. *Nature nanotechnology*, 11(7):633–638, 2016.

[55] Volker Blum, Ralf Gehrke, Felix Hanke, Paula Havu, Ville Havu, Xinguo Ren, Karsten Reuter, and Matthias Scheffler. Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications*, 180(11):2175–2196, 2009.

[56] Alexandre Tkatchenko and Matthias Scheffler. Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data. *Physical review letters*, 102(7):073005, 2009.

[57] Yongho Joo, Myungwoong Kim, Catherine Kanimozhi, Peishen Huang, Bryan M Wong, Susmit Singha Roy, Michael S Arnold, and Padma Gopalan. Effect of dipolar

molecule structure on the mechanism of graphene-enhanced raman scattering. *The Journal of Physical Chemistry C*, 120(25):13815–13824, 2016.

[58] Supriyo Datta. *Electronic transport in mesoscopic systems*. Cambridge university press, 1997.

[59] P Havu, MJ Hashemi, M Kaukonen, ET Seppälä, and RM Nieminen. Effect of gating and pressure on the electronic transport properties of crossed nanotube junctions: formation of a schottky barrier. *Journal of Physics: Condensed Matter*, 23(11):112203, 2011.

[60] Tomi Ketolainen, Ville Havu, and Martti J Puska. Enhancing conductivity of metallic carbon nanotube networks by transition metal adsorption. *The Journal of Chemical Physics*, 142(5):054705, 2015.

[61] San-Huang Ke, Harold U Baranger, and Weitao Yang. Electron transport through molecules: Self-consistent and non-self-consistent approaches. *Physical Review B*, 70(8):085410, 2004.

[62] Seong Jun Kang, Coskun Kocabas, Taner Ozel, Moonsub Shim, Ninad Pimparkar, Muhammad A Alam, Slava V Rotkin, and John A Rogers. High-performance electronics using dense, perfectly aligned arrays of single-walled carbon nanotubes. *Nature nanotechnology*, 2(4):230–236, 2007.

[63] Phaedon Avouris et al. Thin film nanotube transistors based on self-assembled. *Aligned, Semiconducting Carbon Nanotube ArraysACS Nano*, 2(12):2445–2452, 2008.

[64] Melburne C LeMieux, Mark Roberts, Soumendra Barman, Yong Wan Jin, Jong Min Kim, and Zhenan Bao. Self-sorted, aligned nanotube networks for thin-film transistors. *Science*, 321(5885):101–104, 2008.

[65] Yilei Wang, Suresh Kumar Raman Pillai, and Mary B Chan-Park. High-performance partially aligned semiconductive single-walled carbon nanotube transistors achieved with a parallel technique. *Small*, 9(17):2960–2969, 2013.

[66] Gerald J Brady, Austin J Way, Nathaniel S Safron, Harold T Evensen, Padma Gopalan, and Michael S Arnold. Quasi-ballistic carbon nanotube array transistors with current density exceeding si and gaas. *Science advances*, 2(9):e1601240, 2016.

[67] Jana Zaumseil, Xinning Ho, Jeffrey R Guest, Gary P Wiederrecht, and John A Rogers. Electroluminescence from electrolyte-gated carbon nanotube field-effect transistors. *ACS nano*, 3(8):2225–2234, 2009.

[68] Koungmin Ryu, Alexander Badmaev, Chuan Wang, Albert Lin, Nishant Patil, Lewis Gomez, Akshay Kumar, Subhasish Mitra, H-S Philip Wong, and Chongwu Zhou. Cmos-analogous wafer-scale nanotube-on-insulator approach for submicrometer devices and integrated circuits using aligned nanotubes. *Nano Letters*, 9(1):189–197, 2009.

[69] Coskun Kocabas, Simon Dunham, Qing Cao, Kurt Cimino, Xinning Ho, Hoon-Sik Kim, Dale Dawson, Joseph Payne, Mark Stuenkel, Hong Zhang, et al. High-frequency performance of submicrometer transistors that use aligned arrays of single-walled carbon nanotubes. *Nano letters*, 9(5):1937–1943, 2009.

[70] Tejas A Shastry, Jung-Woo T Seo, Josue J Lopez, Heather N Arnold, Jacob Z Kelter, Vinod K Sangwan, Lincoln J Lauhon, Tobin J Marks, and Mark C Hersam. Large-area, electronically monodisperse, aligned single-walled carbon nanotube thin films fabricated by evaporation-driven self-assembly. *Small*, 9(1):45–51, 2013.

[71] Mingguang Chen, Wangxiang Li, Thaís Eloá da Silveira Venzel, Guanghui Li, Mikhail E Itkis, Robert C Haddon, and Elena Bekyarova. Effect of constructive rehybridization on transverse conductivity of aligned single-walled carbon nanotube films. *Materials Today*, 21(9):937–943, 2018.

[72] Chao Sui, Yingchao Yang, Robert J Headrick, Zixuan Pan, Jianyang Wu, Jing Zhang, Shuai Jia, Xinwei Li, Weilu Gao, Oliver S Dewey, et al. Directional sensing based on flexible aligned carbon nanotube film nanocomposites. *Nanoscale*, 10(31):14938–14946, 2018.

[73] Ling Zhang, Yang Wu, Lei Deng, Yi Zhou, Changhong Liu, and Shoushan Fan. Photodetection and photoswitch based on polarized optical response of macroscopically aligned carbon nanotubes. *Nano Letters*, 16(10):6378–6382, 2016.

[74] Ivan Sanchez Esqueda, Xiaodong Yan, Chris Rutherglen, Alex Kane, Tyler Cain, Phil Marsh, Qingzhou Liu, Kosmas Galatsis, Han Wang, and Chongwu Zhou. Aligned carbon nanotube synaptic transistors for large-scale neuromorphic computing. *ACS nano*, 12(7):7352–7361, 2018.

[75] Ertugrul Cubukcu, Fatih Degirmenci, Coskun Kocabas, Mariano A Zimmler, John A Rogers, and Federico Capasso. Aligned carbon nanotubes as polarization-sensitive, molecular near-field detectors. *Proceedings of the National Academy of Sciences*, 106(8):2495–2499, 2009.

[76] Elena Bekyarova, Santanu Sarkar, Feihu Wang, Mikhail E Itkis, Irina Kalinina, Xiaojuan Tian, and Robert C Haddon. Effect of covalent chemistry on the electronic structure and properties of carbon nanotubes and graphene. *Accounts of chemical research*, 46(1):65–76, 2013.

[77] Mildred S Dresselhaus and Gene Dresselhaus. Intercalation compounds of graphite. *Advances in physics*, 51(1):1–186, 2002.

[78] NB Hannay, TH Geballe, BT Matthias, K Andres, P Schmidt, and D MacNair. Superconductivity in graphitic compounds. *Physical Review Letters*, 14(7):225, 1965.

[79] Yoji Koike, H Suematsu, K Higuchi, and S-i Tanuma. Superconductivity in graphite-alkali metal intercalation compounds. *Physica B+ C*, 99(1-4):503–508, 1980.

[80] Albert F Rigosi, Mattias Kruskopf, Heather M Hill, Hanbyul Jin, Bi-Yi Wu, Philip E Johnson, Siyuan Zhang, Michael Berilla, Angela R Hight Walker, Christina A Hacker, et al. Gateless and reversible carrier density tunability in epitaxial graphene devices functionalized with chromium tricarbonyl. *Carbon*, 142:468–474, 2019.

[81] Benoit Champagne and David M Bishop. Calculations of nonlinear optical properties for the solid state. *Advances in Chemical Physics*, 126:41–92, 2003.

[82] Benoît Champagne, Felipe A Bulat, Weitao Yang, Sean Bonness, and Bernard Kirtman. Density functional theory investigation of the polarizability and second hyperpolarizability of polydiacetylene and polybutatriene chains: treatment of exact exchange and role of correlation. *The Journal of chemical physics*, 125(19):194114, 2006.

[83] PA Franken, Alan E Hill, CW el Peters, and Gabriel Weinreich. Generation of optical harmonics. *Physical Review Letters*, 7(4):118, 1961.

[84] Florian Schindler, John M Lupton, Josef Müller, Jochen Feldmann, and Ullrich Scherf. How single conjugated polymer molecules respond to electric fields. *Nature materials*, 5(2):141–146, 2006.

[85] JA Armstrong, N Bloembergen, J Ducuing, and Peter S Pershan. Interactions between light waves in a nonlinear dielectric. *Physical review*, 127(6):1918, 1962.

[86] Nicolaas Bloembergen and PS Pershan. Light waves at the boundary of nonlinear media. *Physical review*, 128(2):606, 1962.

[87] L Dalton, M Canva, GI Stegeman, U Gubler, C Bosshard, HK Shim, and JI Jin. Polymers for photonics applications i; lee ks; ed. *Springer-Verlag Berlin, Heidelberger Platz*, 3:1–86, 2002.

[88] David J Williams. Nonlinear optical properties of organic and polymeric materials. In *ACS symposium series 233*. ACS, 1983.

[89] Daniel Simon Chemla. *Nonlinear Optical Properties of Organic Molecules and Crystals V1*, volume 1. Elsevier, 2012.

[90] Donald M Burland, Robert D Miller, and Cecilia A Walsh. Second-order nonlinearity in poled-polymer systems. *Chemical Reviews*, 94(1):31–75, 1994.

[91] Thomas R Cundari, Henry A Kurtz, and Tie Zhou. Modeling nonlinear optical properties of transition metal complexes. basis set, effective core potential, and geometry effects. *The Journal of Physical Chemistry A*, 102(17):2962–2966, 1998.

[92] Tae-Dong Kim, Jae-Wook Kang, Jingdong Luo, Sei-Hum Jang, Jae-Won Ka, Neil Tucker, Jason B Benedict, Larry R Dalton, Tomoko Gray, René M Overney, et al. Ultralarge and thermally stable electro-optic activities from supramolecular self-assembled molecular glasses. *Journal of the American Chemical Society*, 129(3):488–489, 2007.

[93] Dimitri A Parthenopoulos and Peter M Rentzepis. Three-dimensional optical storage memory. *Science*, 245(4920):843–845, 1989.

[94] Michael Emmelius, Georg Pawlowski, and Hansjörg W Vollmann. Materials for optical data storage. *Angewandte Chemie International Edition in English*, 28(11):1445–1471, 1989.

[95] NS Makarov, A Rebane, M Drobizhev, H Wolleb, and H Spahni. Optimizing two-photon absorption for volumetric optical data storage. *JOSA B*, 24(8):1874–1885, 2007.

[96] Warren R Zipfel, Rebecca M Williams, and Watt W Webb. Nonlinear magic: multiphoton microscopy in the biosciences. *Nature biotechnology*, 21(11):1369–1377, 2003.

[97] Mikhail Drobizhev, Nikolay S Makarov, Shane E Tillo, Thomas E Hughes, and Aleksander Rebane. Two-photon absorption properties of fluorescent proteins. *Nature methods*, 8(5):393–399, 2011.

[98] Brian H Cumpston, Sundaravel P Ananthavel, Stephen Barlow, Daniel L Dyer, Jeffrey E Ehrlich, Lael L Erskine, Ahmed A Heikal, Stephen M Kuebler, I-Y Sandy Lee, Dianne McCord-Maughon, et al. Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature*, 398(6722):51–54, 1999.

[99] Wojciech Haske, Vincent W Chen, Joel M Hales, Wenting Dong, Stephen Barlow, Seth R Marder, and Joseph W Perry. 65 nm feature sizes using visible wavelength 3-d multiphoton lithography. *Optics express*, 15(6):3426–3436, 2007.

[100] Rebecca L Gieseking, Sukrit Mukhopadhyay, Stephen B Shiring, Chad Risko, and Jean-Luc Bredas. Impact of bulk aggregation on the electronic structure of streptocyanines: implications for the solid-state nonlinear optical properties and all-optical switching applications. *The Journal of Physical Chemistry C*, 118(41):23575–23585, 2014.

[101] Jan ML Martin and Manoj K Kesharwani. Assessment of ccsd (t)-f12 approximations and basis sets for harmonic vibrational frequencies. *Journal of Chemical Theory and Computation*, 10(5):2085–2090, 2014.

[102] JC Sancho-García and AJ Pérez-Jiménez. Improved accuracy with medium cost computational methods for the evaluation of bond length alternation of increasingly long oligoacetylenes. *Physical Chemistry Chemical Physics*, 9(44):5874–5879, 2007.

[103] M. BelÉN Oviedo and Bryan M. Wong. Real-Time Quantum Dynamics Reveals Complex, Many-Body Interactions in Solvated Nanodroplets. *J. Chem. Theory Comput.*, 12(4):1862–1871, 2016.

[104] Gerald Knizia, Thomas B Adler, and Hans-Joachim Werner. Simplified ccsd (t)-f12 methods: Theory and benchmarks. *The Journal of chemical physics*, 130(5):054104, 2009.

[105] Thomas Körzdörfer, Robert M Parrish, John S Sears, C David Sherrill, and Jean-Luc Brédas. On the relationship between bond-length alternation and many-electron self-interaction error. *The Journal of chemical physics*, 137(12):124305, 2012.

[106] Aron J Cohen, Paula Mori-Sánchez, and Weitao Yang. Challenges for density functional theory. *Chemical reviews*, 112(1):289–320, 2012.

[107] Kerry Garrett, XochitlA Sosa Vazquez, Shawn B Egri, Jacob Wilmer, Lewis E Johnson, Bruce H Robinson, and Christine M Isborn. Optimum exchange for calculation of excitation energies and hyperpolarizabilities of organic electro-optic chromophores. *Journal of Chemical Theory and Computation*, 10(9):3821–3831, 2014.

[108] Bernard Kirtman, Sean Bonness, Alejandro Ramirez-Solis, Benoit Champagne, Hironori Matsumoto, and Hideo Sekino. Calculation of electric dipole (hyper) polarizabilities by long-range-correction scheme in density functional theory: A systematic assessment for polydiacetylene and polybutatriene oligomers. *The Journal of chemical physics*, 128(11):114108, 2008.

[109] Marc de Wergifosse, Frédéric Wautelet, Benoît Champagne, Ryohei Kishi, Kotaro Fukuda, Hiroshi Matsui, and Masayoshi Nakano. Challenging compounds for calculating hyperpolarizabilities: p-quinodimethane derivatives. *The Journal of Physical Chemistry A*, 117(22):4709–4715, 2013.

[110] Chao Wang, Yizhong Yuan, and Xiaohui Tian. Assessment of range-separated exchange functionals and nonempirical functional tuning for calculating the static second hyperpolarizabilities of streptocyanines. *Journal of Computational Chemistry*, 38(9):594–600, 2017.

[111] Josef Paldus and Jiří Čížek. Hartree–fock stability and symmetry breaking: oxygen doubly negative ion. *Canadian journal of chemistry*, 63(7):1803–1811, 1985.

[112] C David Sherrill, Michael S Lee, and Martin Head-Gordon. On the performance of density functional theory for symmetry-breaking problems. *Chemical physics letters*, 302(5-6):425–430, 1999.

[113] Yoshihiro Tawada, Takao Tsuneda, Susumu Yanagisawa, Takeshi Yanai, and Kimihiko Hirao. A long-range-corrected time-dependent density functional theory. *The Journal of chemical physics*, 120(18):8425–8433, 2004.

[114] Julien Toulouse, Francois Colonna, and Andreas Savin. Long-range–short-range separation of the electron-electron interaction in density-functional theory. *Physical Review A*, 70(6):062505, 2004.

[115] Melissa J Lucero, Thomas M Henderson, and Gustavo E Scuseria. Improved semiconductor lattice parameters and band gaps from a middle-range screened hybrid exchange functional. *Journal of Physics: Condensed Matter*, 24(14):145504, 2012.

[116] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical physics letters*, 393(1-3):51–57, 2004.

[117] Michael E Foster and Bryan M Wong. Nonempirically tuned range-separated dft accurately predicts both fundamental and excitation gaps in dna and rna nucleobases. *Journal of chemical theory and computation*, 8(8):2682–2687, 2012.

[118] Bryan M Wong, Manuel Piacenza, and Fabio Della Sala. Absorption and fluorescence properties of oligothiophene biomarkers from long-range-corrected time-dependent density functional theory. *Physical Chemistry Chemical Physics*, 11(22):4498–4508, 2009.

[119] Bryan M Wong and Joseph G Cordaro. Coumarin dyes for dye-sensitized solar cells: A long-range-corrected density functional study. *The Journal of chemical physics*, 129(21):214703, 2008.

[120] Bryan M Wong and Timothy H Hsieh. Optoelectronic and excitonic properties of oligoacenes: substantial improvements from range-separated time-dependent density functional theory. *Journal of chemical theory and computation*, 6(12):3704–3712, 2010.

[121] Alexandra E Raeber and Bryan M Wong. The importance of short-and long-range exchange on various excited state properties of dna monomers, stacked complexes, and watson–crick pairs. *Journal of Chemical Theory and Computation*, 11(5):2199–2209, 2015.

[122] Lindsey N Anderson, M Belén Oviedo, and Bryan M Wong. Accurate electron affinities and orbital energies of anions from a nonempirically tuned range-separated density functional theory approach. *Journal of Chemical Theory and Computation*, 13(4):1656–1666, 2017.

[123] Lindsey N Anderson, Fredy W Aquino, Alexandra E Raeber, Xi Chen, and Bryan M Wong. Halogen bonding interactions: Revised benchmarks and a new assessment of exchange vs dispersion. *Journal of Chemical Theory and Computation*, 14(1):180–190, 2018.

[124] Natalia Kuritz, Tamar Stein, Roi Baer, and Leeor Kronik. Charge-transfer-like $\pi \rightarrow \pi^*$ excitations in time-dependent density functional theory: A conundrum and its solution. *Journal of chemical theory and computation*, 7(8):2408–2415, 2011.

[125] Ryan M Richard and John M Herbert. Time-dependent density-functional description of the 1l a state in polycyclic aromatic hydrocarbons: Charge-transfer character in disguise? *Journal of chemical theory and computation*, 7(5):1296–1306, 2011.

[126] Leeor Kronik, Tamar Stein, Sivan Refaely-Abramson, and Roi Baer. Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals. *Journal of Chemical Theory and Computation*, 8(5):1515–1531, 2012.

[127] Tamar Stein, Leeor Kronik, and Roi Baer. Prediction of charge-transfer excitations in coumarin-based dyes using a range-separated functional tuned from first principles. *The Journal of chemical physics*, 131(24):244119, 2009.

[128] Tamar Stein, Leeor Kronik, and Roi Baer. Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory. *Journal of the American Chemical Society*, 131(8):2818–2820, 2009.

[129] Thomas B Adler, Gerald Knizia, and Hans-Joachim Werner. A simple and efficient ccsd (t)-f12 approximation. *The Journal of chemical physics*, 127(22):221106, 2007.

[130] Gerald Knizia and Hans-Joachim Werner. Explicitly correlated rmp2 for high-spin open-shell reference states. *The Journal of chemical physics*, 128(15):154103, 2008.

[131] David P Tew, Wim Klopper, Christian Neiss, and Christof Hättig. Quintuple-ζ quality coupled-cluster correlation energies with triple-ζ basis sets. *Physical Chemistry Chemical Physics*, 9(16):1921–1930, 2007.

[132] Sam-Shajing Sun and Larry R Dalton. *Introduction to organic electronic and optoelectronic materials and devices*. CRC press, 2008.

[133] Rebecca L Gieseking, Chad Risko, and Jean-Luc Bredas. Distinguishing the effects of bond-length alternation versus bond-order alternation on the nonlinear optical properties of π-conjugated chromophores. *The journal of physical chemistry letters*, 6(12):2158–2162, 2015.

[134] HJ Werner, PJ Knowles, G Knizia, FR Manby, and M Schütz. Wiley interdiscip. *Rev.: Comput. Mol. Sci*, 2(2):242–253, 2012.

[135] MJEA Frisch, Gary W Trucks, H Bernhard Schlegel, Gustavo E Scuseria, Michael A Robb, James R Cheeseman, Giovanni Scalmani, Vincenzo Barone, Benedetta Mennucci, GAea Petersson, et al. Gaussian 09, revision d. 01, 2009.

[136] Tian Lu and Feiwu Chen. Multiwfn: a multifunctional wavefunction analyzer. *Journal of computational chemistry*, 33(5):580–592, 2012.

[137] Carlos A Jiménez-Hoyos, R Rodríguez-Guzmán, and Gustavo E Scuseria. Polyradical character and spin frustration in fullerene molecules: An ab initio non-collinear hartree–fock study. *The Journal of Physical Chemistry A*, 118(42):9925–9940, 2014.

[138] Artëm E Masunov, Dane Anderson, Alexandra Ya Freidzon, and Alexander A Bagaturyants. Symmetry-breaking in cationic polymethine dyes: part 2. shape of electronic absorption bands explained by the thermal fluctuations of the solvent reaction field. *The Journal of Physical Chemistry A*, 119(26):6807–6815, 2015.

[139] Lee M Thompson and Hrant P Hratchian. Modeling the photoelectron spectra of monbo2–accounting for spin contamination in density functional theory. *The Journal of Physical Chemistry A*, 119(32):8744–8751, 2015.

[140] Kada Yesudas. Cationic cyanine dyes: impact of symmetry-breaking on optical absorption and third-order polarizabilities. *Physical Chemistry Chemical Physics*, 15(44):19465–19477, 2013.

[141] Mary A Rohrdanz and John M Herbert. Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory. *The Journal of chemical physics*, 129(3):034107, 2008.

[142] Mary A Rohrdanz, Katie M Martins, and John M Herbert. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *The Journal of chemical physics*, 130(5):054112, 2009.

[143] TJ Lee. Ge scuseria in quantum mechanical electronic structure calculations with chemical accuracy, ed. sr langhoff, 1995.

[144] Robert W Woody. Theory of circular dichroism of proteins. In *Circular dichroism and the conformational analysis of biomolecules*, pages 25–67. Springer, 1996.

[145] Narasimha Sreerama and Robert W Woody. Estimation of protein secondary structure from circular dichroism spectra: comparison of contin, selcon, and cdsstr methods with an expanded reference set. *Analytical biochemistry*, 287(2):252–260, 2000.

[146] Siobhan E Toal, Daniel J Verbaro, and Reinhard Schweitzer-Stenner. Role of enthalpy–entropy compensation interactions in determining the conformational propensities of amino acid residues in unfolded peptides. *The Journal of Physical Chemistry B*, 118(5):1309–1318, 2014.

[147] Andrew Hagarman, Thomas J Measey, Daniel Mathieu, Harald Schwalbe, and Reinhard Schweitzer-Stenner. Intrinsic propensities of amino acid residues in gxg peptides inferred from amide i' band profiles and nmr scalar coupling constants. *Journal of the American Chemical Society*, 132(2):540–551, 2010.

[148] Siobhan Toal and Reinhard Schweitzer-Stenner. Local order in the unfolded state: conformational biases and nearest neighbor interactions. *Biomolecules*, 4(3):725–773, 2014.

[149] M Lois Tiffany and S Krimm. Circular dichroism of poly-l-proline in an unordered conformation. *Biopolymers: Original Research on Biomolecules*, 6(12):1767–1770, 1968.

[150] Robert W Woody. Circular dichroism spectrum of peptides in the poly (pro) ii conformation. *Journal of the American Chemical Society*, 131(23):8234–8245, 2009.

[151] Zhengshuang Shi, Robert W Woody, and Neville R Kallenbach. Is polyproline ii a major backbone conformation in unfolded proteins? *Advances in protein chemistry*, 62:163–240, 2002.

[152] Narasimha Sreerama and Robert W Woody. Structural composition of βi-and βii-proteins. *Protein Science*, 12(2):384–388, 2003.

[153] Zhigang Liu, Kang Chen, Angela Ng, Zhengshuang Shi, Robert W Woody, and Neville R Kallenbach. Solvent dependence of pii conformation in model alanine peptides. *Journal of the American Chemical Society*, 126(46):15141–15150, 2004.

[154] Jürgen Graf, Phuong H Nguyen, Gerhard Stock, and Harald Schwalbe. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/nmr study. *Journal of the American Chemical Society*, 129(5):1179–1189, 2007.

[155] Siobhan Toal, Derya Meral, Daniel Verbaro, Brigita Urbanc, and Reinhard Schweitzer-Stenner. ph-independence of trialanine and the effects of termini blocking in short peptides: a combined vibrational, nmr, uvcd, and molecular dynamics study. *The Journal of Physical Chemistry B*, 117(14):3689–3706, 2013.

[156] William Moffitt. Optical rotatory dispersion of helical polymers. *The Journal of Chemical Physics*, 25(3):467–478, 1956.

[157] EW Ronish and Samuel Krimm. The calculated circular dichroism of polyproline ii in the polarizability approximation. *Biopolymers: Original Research on Biomolecules*, 13(8):1635–1651, 1974.

[158] Jiří Šebek, Zdeněk Kejík, and Petr Bouř. Geometry and solvent dependence of the electronic spectra of the amide group and consequences for peptide circular dichroism. *The Journal of Physical Chemistry A*, 110(14):4702–4711, 2006.

[159] Jakub Kaminský, Jan Kubelka, and Petr Bour. Theoretical modeling of peptide α-helical circular dichroism in aqueous solution. *The Journal of Physical Chemistry A*, 115(9):1734–1742, 2011.

[160] Jiří Šebek, Bela Gyurcsik, Jaroslav Šebestík, Zdeněk Kejík, Lucie Bednarova, and Petr Bouř. Interpretation of synchrotron radiation circular dichroism spectra of anionic, cationic, and zwitterionic dialanine forms. *The Journal of Physical Chemistry A*, 111(14):2750–2760, 2007.

[161] Niranjan V Ilawe, Alexandra E Raeber, Reinhard Schweitzer-Stenner, Siobhan E Toal, and Bryan M Wong. Assessing backbone solvation effects in the conformational propensities of amino acid residues in unfolded peptides. *Physical Chemistry Chemical Physics*, 17(38):24917–24924, 2015.

[162] Niranjan V Ilawe, Reinhard Schweitzer-Stenner, David DiGuiseppi, and Bryan M Wong. Is a cross-β-sheet structure of low molecular weight peptides necessary for the formation of fibrils and peptide hydrogels? *Physical Chemistry Chemical Physics*, 20(27):18158–18168, 2018.

[163] Arsênio PV Neto, Daniel F Scalabrini Machado, Thiago O Lopes, Ademir J Camargo, and Heibbe CB de Oliveira. Explicit aqueous solvation treatment of epinephrine

from car–parrinello molecular dynamics: Effect of hydrogen bonding on the electronic absorption spectrum. *The Journal of Physical Chemistry B*, 122(35):8439–8450, 2018.

[164] S Brahms, J Brahms, G Spach, and A Brack. Identification of β, β-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. *Proceedings of the National Academy of Sciences*, 74(8):3208–3212, 1977.

[165] Richard L Martin. Natural transition orbitals. *The Journal of chemical physics*, 118(11):4775–4777, 2003.

[166] Siobhan E Toal, Nina Kubatova, Christian Richter, Verena Linhard, Harald Schwalbe, and Reinhard Schweitzer-Stenner. Randomizing the unfolded state of peptides (and proteins) by nearest neighbor interactions between unlike residues. *Chemistry–A European Journal*, 21(13):5173–5192, 2015.

[167] Reinhard Schweitzer-Stenner and Siobhan E Toal. Anticooperative nearest-neighbor interactions between residues in unfolded peptides and proteins. *Biophysical journal*, 114(5):1046–1057, 2018.

[168] Sven Hovmöller, Tuping Zhou, and Tomas Ohlson. Conformations of amino acids in proteins. *Acta Crystallographica Section D: Biological Crystallography*, 58(5):768–776, 2002.

[169] Ramachandran GN Ramakrishnan C Sasisekharan. Stereochemistry of polypeptide chain configurations j mol biol 79599. ramachandran gn, ramakrishnan c, and sasisekharan (1963). stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7:95–99, 1963.

[170] David A Brant and Paul J Flory. The configuration of random polypeptide chains. i. experimental results. *Journal of the American Chemical Society*, 87(13):2788–2791, 1965.

[171] Paul J Flory and M Volkenstein. Statistical mechanics of chain molecules, 1969.

[172] Francisco F García-Prieto, Ignacio Fdez Galván, Manuel A Aguilar, and M Elena Martín. Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology. *The Journal of chemical physics*, 135(19):194502, 2011.

[173] Michael Feig. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *Journal of Chemical Theory and Computation*, 4(9):1555–1564, 2008.

[174] Bosco K Ho, Annick Thomas, and Robert Brasseur. Revisiting the ramachandran plot: Hard-sphere repulsion, electrostatics, and h-bonding in the α-helix. *Protein Science*, 12(11):2508–2522, 2003.

[175] Bosco K Ho and Robert Brasseur. The ramachandran plots of glycine and pre-proline. *BMC structural biology*, 5(1):1–11, 2005.

[176] Zhengshuang Shi, Kang Chen, Zhigang Liu, Angela Ng, W Clay Bracken, and Neville R Kallenbach. Polyproline ii propensities from ggxgg peptides reveal an anticorrelation with β-sheet scales. *Proceedings of the National Academy of Sciences*, 102(50):17964–17968, 2005.

[177] Joze Grdadolnik, Vlasta Mohacek-Grosev, Robert L Baldwin, and Franc Avbelj. Populations of the three major backbone conformations in 19 amino acid dipeptides. *Proceedings of the National Academy of Sciences*, 108(5):1794–1798, 2011.

[178] Jože Grdadolnik, Simona Golić Grdadolnik, and Franc Avbelj. Determination of conformational preferences of dipeptides using vibrational spectroscopy. *The Journal of Physical Chemistry B*, 112(9):2712–2718, 2008.

[179] Daniel J Verbaro, Daniel Mathieu, Siobhan E Toal, Harald Schwalbe, and Reinhard Schweitzer-Stenner. Ionized trilysine: A model system for understanding the nonrandom structure of poly-l-lysine and lysine-containing motifs in proteins. *The Journal of Physical Chemistry B*, 116(28):8084–8094, 2012.

[180] Reinhard Schweitzer-Stenner and Siobhan E Toal. Entropy reduction in unfolded peptides (and proteins) due to conformational preferences of amino acid residues. *Physical Chemistry Chemical Physics*, 16(41):22527–22536, 2014.

[181] Václav Parchaˊnský, Josef Kapitán, Jakub Kaminský, Jaroslav Šebestík, and Petr Bouř. Ramachandran plot for alanine dipeptide as determined from raman optical activity. *The Journal of Physical Chemistry Letters*, 4(16):2763–2768, 2013.

[182] James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

[183] Chuan Tian, Koushik Kasavajhala, Kellon AA Belfon, Lauren Raguette, He Huang, Angela N Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, et al. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.*, 16(1):528–552, 2019.

[184] Alexander D Mackerell Jr, Michael Feig, and Charles L Brooks III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, 25(11):1400–1415, 2004.

[185] Derya Meral, Siobhan Toal, Reinhard Schweitzer-Stenner, and Brigita Urbanc. Water-centered interpretation of intrinsic ppii propensities of amino acid residues: in vitro-driven molecular dynamics study. *The Journal of Physical Chemistry B*, 119(42):13237–13251, 2015.

[186] Shuting Zhang, Reinhard Schweitzer-Stenner, and Brigita Urbanc. Do molecular dynamics force fields capture conformational dynamics of alanine in water? *Journal of chemical theory and computation*, 16(1):510–527, 2019.

[187] Melissa A Kelly, Brian W Chellgren, Adam L Rucker, Jerry M Troutman, Michael G Fried, Anne-Frances Miller, and Trevor P Creamer. Host- guest study of left-handed polyproline ii helix formation. *Biochemistry*, 40(48):14376–14383, 2001.

[188] Adam L Rucker and Trevor P Creamer. Polyproline ii helical structure in protein unfolded states: lysine peptides revisited. *Protein science*, 11(4):980–985, 2002.

[189] Fatma Eker, Kai Griebenow, and Reinhard Schweitzer-Stenner. Stable conformations of tripeptides in aqueous solution studied by uv circular dichroism spectroscopy. *Journal of the American Chemical Society*, 125(27):8178–8185, 2003.

[190] Zhengshuang Shi, Kang Chen, Zhigang Liu, and Neville R Kallenbach. Conformation of the backbone in unfolded proteins. *Chemical reviews*, 106(5):1877–1897, 2006.

[191] Wei Yuan Yang, Edgar Larios, and Martin Gruebele. On the extended β-conformation propensity of polypeptides at high temperature. *Journal of the American Chemical Society*, 125(52):16220–16227, 2003.

[192] Ignacio Tinoco Jr. Theoretical aspects of optical activity. *Polymers. Adv. Chem. Phys., 4 (1962)*, pages 113–160, 2009.

[193] Joe S Ham and JR Platt. Far uv spectra of peptides. *The Journal of Chemical Physics*, 20(2):335–336, 1952.

[194] WB Gratzer, W Rhodes, and GD Fasman. Optical properties of the poly-l-proline and collagen helices. *Biopolymers: Original Research on Biomolecules*, 1(4):319–330, 1963.

[195] Eigil B Nielsen and John A Schellman. The absorption spectra of simple amides and peptides. *The Journal of physical chemistry*, 71(7):2297–2304, 1967.

[196] Leo J Saidel, A Robert Goldfarb, and Sheldon Waldman. The absorption spectra of amino acids in the region two hundred to two hundred and thirty millimicrons. *Journal of Biological Chemistry*, 197(1):285–291, 1952.

[197] Michael J Frisch. Gaussian 03 rev. e. 01. *http://www. gaussian. com/*, 2004.

[198] Niranjan V. Ilawe, M. Belén Oviedo, and Bryan M. Wong. Effect of quantum tunneling on the efficiency of excitation energy transfer in plasmonic nanoparticle chain waveguides. *J. Mater. Chem. C*, 6:5857–5864, 2018.

[199] Giuseppe Lanza and Maria A Chiacchio. Ab initio mp2 and density functional theory computational study of acalanh2 peptide hydration: A bottom-up approach. *ChemPhysChem*, 15(13):2785–2793, 2014.

[200] Karin Rybka, Siobhan E Toal, Daniel J Verbaro, Daniel Mathieu, Harald Schwalbe, and Reinhard Schweitzer-Stenner. Disorder and order in unfolded and disordered peptides and proteins: A view derived from tripeptide conformational analysis. ii. tripeptides with short side chains populating asx and β-type like turn conformations. *Proteins: Structure, Function, and Bioinformatics*, 81(6):968–983, 2013.

[201] Jun Jiang, Darius Abramavicius, Benjamin M Bulheller, Jonathan D Hirst, and Shaul Mukamel. Ultraviolet spectroscopy of protein backbone transitions in aqueous solution: Combined qm and mm simulations. *The Journal of Physical Chemistry B*, 114(24):8270–8277, 2010.

[202] Jonathan D Hirst, Karl Colella, and Andrew TB Gilbert. Electronic circular dichroism of proteins from first-principles calculations. *The Journal of Physical Chemistry B*, 107(42):11813–11819, 2003.

[203] Michael J Robertson, Julian Tirado-Rives, and William L Jorgensen. Improved peptide and protein torsional energetics with the opls-aa force field. *Journal of chemical theory and computation*, 11(7):3499–3509, 2015.

[204] Robert B Best and Gerhard Hummer. Optimized molecular dynamics force fields applied to the helix- coil transition of polypeptides. *The journal of physical chemistry B*, 113(26):9004–9015, 2009.

[205] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo, Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18):3586–3616, 1998.

[206] Chen-Yang Zhou, Fan Jiang, and Yun-Dong Wu. Residue-specific force field based on protein coil library. rsff2: modification of amber ff99sb. *The journal of physical chemistry B*, 119(3):1035–1047, 2015.

[207] Shuxiang Li and Adrian H Elcock. Residue-specific force field (rsff2) improves the modeling of conformational behavior of peptides and proteins. *The journal of physical chemistry letters*, 6(11):2127–2133, 2015.

[208] Giuseppe Lanza and Maria A Chiacchio. Comprehensive and accurate ab initio energy surface of simple alanine peptides. *ChemPhysChem*, 14(14):3284–3293, 2013.

[209] Christoph Freysoldt, Blazej Grabowski, Tilmann Hickel, Jörg Neugebauer, Georg Kresse, Anderson Janotti, and Chris G Van de Walle. First-principles calculations for point defects in solids. *Reviews of modern physics*, 86(1):253, 2014.

[210] Minhua Ai, Jing-Wen Zhang, Yi-Wei Wu, Lun Pan, Chengxiang Shi, and Ji-Jun Zou. Role of vacancies in photocatalysis: a review of recent progress. *Chemistry–An Asian Journal*, 15(22):3599–3619, 2020.

[211] Chris G Van de Walle and Anderson Janotti. Advances in electronic structure methods for defects and impurities in solids. *physica status solidi (b)*, 248(1):19–27, 2011.

[212] Risto M Nieminen. Issues in first-principles calculations for defects in semiconductors and oxides. *Modelling and Simulation in Materials Science and Engineering*, 17(8):084001, 11 2009.

[213] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 11 1964.

[214] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 11 1965.

[215] Th. Frauenheim, F. Weich, Th. Köhler, S. Uhlmann, D. Porezag, and G. Seifert. Density-functional-based construction of transferable nonorthogonal tight-binding potentials for si and sih. *Phys. Rev. B*, 52:11492–11501, 10 1995.

[216] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, 51:12947–12957, 05 1995.

[217] J. Widany, Th. Frauenheim, Th. Köhler, M. Sternberg, D. Porezag, G. Jungnickel, and G. Seifert. Density-functional-based construction of transferable nonorthogonal tight-binding potentials for b, n, bn, bh, and nh. *Phys. Rev. B*, 53:4443–4452, 02 1996.

[218] Sharma S.R.K.C. Yamijala, Ma. Belén Oviedo, and Bryan M. Wong. *Density Functional Tight Binding Calculations for Probing Electronic-Excited States of Large Systems*, chapter 2, pages 45–79. John Wiley & Sons, Ltd, 2022.

[219] Niranjan V. Ilawe, M. BelÉN Oviedo, and Bryan M. Wong. Real-Time Quantum Dynamics of Long-Range Electronic Excitation Transfer in Plasmonic Nanoantennas. *J. Chem. Theory Comput.*, 13(8):3442–3454, 2017.

[220] Niranjan V Ilawe, M BelÉN Oviedo, and Bryan M Wong. Effect of quantum tunneling on the efficiency of excitation energy transfer in plasmonic nanoparticle chain waveguides. *J. Mater. Chem. C*, 6(22):5857–5864, 2018.

[221] Sarah I Allec, Yijing Sun, Jianan Sun, Chia-en A Chang, and Bryan M Wong. Heterogeneous cpu+ gpu-enabled simulations for dftb molecular dynamics of large chemical and biological systems. *J. Chem. Theory. Comput.*, 15(5):2807–2815, 2019.

[222] Shahidul M. Islam and Pierre-Nicholas Roy. Performance of the scc-dftb model for description of five-membered ring carbohydrate conformations: Comparison to force fields, high-level electronic structure methods, and experiment. *J. Chem. Theory Comput.*, 8(7):2412–2423, 2012. PMID: 26588973.

[223] Kirsty Leong, Michael E. Foster, Bryan M. Wong, Erik D. Spoerke, Dara Van Gough, Joseph C. Deaton, and Mark D. Allendorf. Energy and charge transfer by donor–acceptor pairs confined in a metal–organic framework: A spectroscopic and computational investigation. *J. Mater. Chem. A*, 2:3389–3398, 2014.

[224] José M. Rodríguez-Borbón, Amin Kalantar, Sharma S. R. K. C. Yamijala, M. Belén Oviedo, Walid Najjar, and Bryan M. Wong. Field programmable gate arrays for enhancing the speed and energy efficiency of quantum dynamics simulations. *J. Chem. Theory Comput.*, 16(4):2085–2098, 2020.

[225] A Van der Ven, JC Thomas, Qingchuan Xu, and J Bhattacharya. Linking the electronic structure of solids to their thermodynamic and kinetic properties. *Mathematics and computers in simulation*, 80(7):1393–1410, 2010.

[226] Anton Van der Ven, John C Thomas, Qingchuan Xu, Benjamin Swoboda, and Dane Morgan. Nondilute diffusion from first principles: Li diffusion in li_xtis_2. *Physical Review B*, 78(10):104306, 2008.

[227] AP Drozdov, MI Eremets, IA Troyan, Vadim Ksenofontov, and Sergii I Shylin. Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system. *Nature*, 525(7567):73–76, 2015.

[228] Takahiro Ishikawa, Takashi Miyake, and Katsuya Shimizu. Materials informatics based on evolutionary algorithms: Application to search for superconducting hydrogen compounds. *Physical Review B*, 100(17):174506, 2019.

[229] Bastien Guigue, Adrien Marizy, and Paul Loubeyre. Direct synthesis of pure h_3s from s and h elements: No evidence of the cubic superconducting phase up to 160 gpa. *Physical Review B*, 95(2):020104, 2017.

[230] Maddury Somayazulu, Muhtar Ahart, Ajay K Mishra, Zachary M Geballe, Maria Baldini, Yue Meng, Viktor V Struzhkin, and Russell J Hemley. Evidence for superconductivity above 260 k in lanthanum superhydride at megabar pressures. *Physical review letters*, 122(2):027001, 2019.

[231] AP Drozdov, PP Kong, VS Minkov, SP Besedin, MA Kuzovnikov, S Mozaffari, L Balicas, FF Balakirev, DE Graf, VB Prakapenka, et al. Superconductivity at 250 k in lanthanum hydride under high pressures. *Nature*, 569(7757):528–531, 2019.

[232] Christopher R Weinberger, Xiao-Xiang Yu, Hang Yu, and Gregory B Thompson. Ab initio investigations of the phase stability in group ivb and vb transition metal nitrides. *Computational Materials Science*, 138:333–345, 2017.

[233] Xiao-Xiang Yu, Christopher R Weinberger, and Gregory B Thompson. Ab initio investigations of the phase stability in group ivb and vb transition metal carbides. *Computational Materials Science*, 112:318–326, 2016.

[234] Artem R Oganov and Colin W Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of chemical physics*, 124(24):244704, 2006.

[235] Artem R Oganov, Yanming Ma, Andriy O Lyakhov, Mario Valle, and Carlo Gatti. Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Reviews in Mineralogy and Geochemistry*, 71(1):271–298, 2010.

[236] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Crystal structure prediction via particle-swarm optimization. *Physical Review B*, 82(9):094116, 2010.

[237] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Calypso: A method for crystal structure prediction. *Computer Physics Communications*, 183(10):2063–2070, 2012.

[238] Dang-Hyok Yoon and Ivar E Reimanis. A review on the joining of sic for high-temperature applications. *Journal of the Korean Ceramic Society*, 57(3):246–270, 2020.

[239] Xu She, Alex Q Huang, Oscar Lucia, and Burak Ozpineci. Review of silicon carbide power devices and their applications. *IEEE Transactions on Industrial Electronics*, 64(10):8193–8205, 2017.

[240] Roland Madar. Silicon carbide in contention. *Nature*, 430(7003):974–975, 2004.

[241] JB Casady and R Wayne Johnson. Status of silicon carbide (sic) as a wide-bandgap semiconductor for high-temperature applications: A review. *Solid-State Electronics*, 39(10):1409–1422, 1996.

[242] Christof Wöll. The chemistry and physics of zinc oxide surfaces. *Progress in surface science*, 82(2-3):55–120, 2007.

[243] A. Santoro and A. D. Mighell. Properties of crystal lattices: the derivative lattices and their determination. *Acta Crystallographica Section A*, 28(3):284–287, 05 1972.

[244] A. Santoro and A. D. Mighell. Coincidence-site lattices. *Acta Crystallographica Section A*, 29(2):169–175, 03 1973.

[245] Gus L. W. Hart and Rodney W. Forcade. Algorithm for generating derivative structures. *Phys. Rev. B*, 77:224115, 06 2008.

[246] Gus LW Hart, Lance J Nelson, and Rodney W Forcade. Generating derivative structures at a fixed concentration. *Computational Materials Science*, 59:101–107, 2012.

[247] B. Puchala and A. Van der Ven. Thermodynamics of the zr-o system from first-principles calculations. *Phys. Rev. B*, 88:094108, 09 2013.

[248] G. Kresse and J. Hafner. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, 01 1993.

[249] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, 10 1996.

[250] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, 12 1994.

[251] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, 01 1999.

[252] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 10 1996.

[253] Sha Liu, E Martínez, and Javier LLorca. Prediction of the al-rich part of the al-cu phase diagram using cluster expansion and statistical mechanics. *Acta Materialia*, 195:317–326, 2020.

[254] Dirk Porezag, Th Frauenheim, Th Köhler, Gotthard Seifert, and R Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, 51(19):12947, 1995.

[255] G Seifert, D Porezag, and Th Frauenheim. Calculations of molecules, clusters, and solids with a simplified lcao-dft-lda scheme. *Int. J. Quantum Chem.*, 58(2):185–192, 1996.

[256] Marcus Elstner, Dirk Porezag, G Jungnickel, J Elsner, M Haugk, Th Frauenheim, Sandor Suhai, and Gotthard Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58(11):7260, 1998.

[257] Th Frauenheim, G Seifert, M Elsterner, Z Hajnal, G Jungnickel, D Porezag, S Suhai, and R Scholz. A self-consistent charge density-functional based tight-binding method for predictive materials simulations in physics, chemistry and biology. *Phys. Status Solidi B*, 217(1):41–62, 2000.

[258] Thomas Frauenheim, Gotthard Seifert, Marcus Elstner, Thomas Niehaus, Christof Köhler, Marc Amkreutz, Michael Sternberg, Zoltán Hajnal, Aldo Di Carlo, and Sándor Suhai. Atomistic simulations of complex materials: Ground-state and excited-state properties. *J. Phys. Condens. Matter.*, 14(11):3015, 2002.

[259] Augusto F Oliveira, Gotthard Seifert, Thomas Heine, and Hélio A Duarte. Density-functional based tight-binding: An approximate dft method. *J. Braz. Chem. Soc.*, 20(7):1193–1205, 2009.

[260] Gotthard Seifert and Jan-Ole Joswig. Density-functional tight binding—an approximate density-functional theory method. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2(3):456–465, 2012.

[261] Michael Gaus, Qiang Cui, and Marcus Elstner. Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *J. Chem. Theory Comput.*, 7(4):931–948, 2011.

[262] Masoud Bezi Javan. Optical properties of sic nanocages: ab initio study. *Applied Physics A*, 113(1):105–113, 2013.

[263] RMA Khalil, Fayyaz Hussain, and Niaz Ahmad Niaz. Studies of molecular dynamics and electronic structure in cubic-sic by using density functional tight binding approach. *Journal of Materials and Physical Sciences*, 2(2):88–94, 2021.

[264] Onsuda Arayawut, Teerakiat Kerdcharoen, and Chatchawal Wongchoosuk. Structures, electronic properties, and gas permeability of 3d pillared silicon carbide nanostructures. *Nanomaterials*, 12(11):1869, 2022.

[265] Matti Hellström, Kjell Jorner, Maria Bryngelsson, Stefan E Huber, Jolla Kullgren, Thomas Frauenheim, and Peter Broqvist. An scc-dftb repulsive potential for various zno polymorphs and the zno–water system. *The Journal of Physical Chemistry C*, 117(33):17004–17015, 2013.

[266] Vu Ngoc Tuoc, Tran Doan Huan, Nguyen Viet Minh, and Nguyen Thi Thao. Density functional theory based tight binding study on theoretical prediction of low-density nanoporous phases zno semiconductor materials. In *Journal of Physics: Conference Series*, volume 726, page 012022. IOP Publishing, 2016.

[267] Ahmad W Huran, Conrad Steigemann, Thomas Frauenheim, Bálint Aradi, and Miguel AL Marques. Efficient automatized density-functional tight-binding parametrizations: application to group iv elements. *Journal of Chemical Theory and Computation*, 14(6):2947–2954, 2018.

[268] Christof Köhler, Zoltán Hajnal, Péter Deák, Thomas Frauenheim, and Sándor Suhai. Theoretical investigation of carbon defects and diffusion in α-quartz. *Physical Review B*, 64(8):085333, 2001.

[269] Ney H Moreira, Grygoriy Dolgonos, Bálint Aradi, Andreia L da Rosa, and Thomas Frauenheim. Toward an accurate density-functional tight-binding description of zinc-containing compounds. *Journal of chemical theory and computation*, 5(3):605–614, 2009.

[270] Vladimir S Naumov, Anastasiia S Loginova, Alexander A Avdoshin, Stanislav K Ignatov, Alexey V Mayorov, Bálint Aradi, and Thomas Frauenheim. Structural, electronic, and thermodynamic properties of tio_2/organic clusters: performance of dftb method with different parameter sets. *International Journal of Quantum Chemistry*, 121(2):e26427, 2021.

[271] Thomas A Niehaus, Sigismund TAG Melissen, Balint Aradi, and S Mehdi Vaez Allaei. Towards a simplified description of thermoelectric materials: accuracy of approximate density functional theory for phonon dispersions. *Journal of Physics: Condensed Matter*, 31(39):395901, 2019.

[272] Shuaiwei Wang, Zhaochuan Fan, Rik S Koster, Changming Fang, Marijn A Van Huis, Anil O Yalcin, Frans D Tichelaar, Henny W Zandbergen, and Thijs JH Vlugt. New ab initio based pair potential for accurate simulation of phase transitions in zno. *The Journal of Physical Chemistry C*, 118(20):11050–11061, 2014.

[273] Mahlaga P Molepo and Daniel P Joubert. Computational study of the structural phases of zno. *Physical Review B*, 84(9):094110, 2011.

[274] Paul Erhart, Andreas Klein, and Karsten Albe. First-principles study of the structure and stability of oxygen defects in zinc oxide. *Physical Review B*, 72(8):085213, 2005.

[275] Chris G Van de Walle. Hydrogen as a cause of doping in zinc oxide. *Physical review letters*, 85(5):1012, 2000.

[276] Eun-Cheol Lee, Y-S Kim, Y-G Jin, and Kee-Joo Chang. Compensation mechanism for n acceptors in zno. *Physical Review B*, 64(8):085120, 2001.

[277] Paul Erhart and Karsten Albe. First-principles study of migration mechanisms and diffusion of oxygen in zinc oxide. *Physical Review B*, 73(11):115207, 2006.

[278] Tula R Paudel and Walter RL Lambrecht. First-principles calculation of the o vacancy in zno: A self-consistent gap-corrected approach. *Physical Review B*, 77(20):205202, 2008.

[279] Mark Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2010.

[280] Mark E Tuckerman and Glenn J Martyna. Understanding modern molecular dynamics: Techniques and applications, 2000.

[281] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, volume 1. Elsevier, 2001.

[282] John D Chodera and Frank Noé. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.*, 25:135–144, 2014.

[283] Yinglong Miao, Ferran Feixas, Changsun Eun, and J Andrew McCammon. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.*, 36(20):1536–1549, 2015.

[284] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[285] Bingqing Cheng and Michele Ceriotti. Bridging the gap between atomistic and macroscopic models of homogeneous nucleation. *J. Chem. Phys.*, 146(3):034106, 2017.

[286] Federico Giberti, Matteo Salvalaglio, and Michele Parrinello. Metadynamics studies of crystal nucleation. *IUCrJ*, 2(2):256–266, 2015.

[287] Federico Giberti, Matteo Salvalaglio, Marco Mazzotti, and Michele Parrinello. 1, 3, 5-tris (4-bromophenyl)-benzene nucleation: From dimers to needle-like clusters. *Cryst. Growth Des.*, 17(8):4137–4143, 2017.

[288] Daniel R Reid, Ivan Lyubimov, MD Ediger, and Juan J De Pablo. Age and structure of a model vapour-deposited glass. *Nat. Commun.*, 7(1):1–9, 2016.

[289] Julian Helfferich, Ivan Lyubimov, Daniel Reid, and Juan J de Pablo. Inherent structure energy is a good indicator of molecular mobility in glasses. *Soft Matter*, 12(27):5898–5904, 2016.

[290] Tuan Anh Pham, Yuan Ping, and Giulia Galli. Modelling heterogeneous interfaces for solar water splitting. *Nat. Mater.*, 16(4):401–408, 2017.

[291] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12562–12566, 2002.

[292] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100(2):020603, 2008.

[293] Ayori Mitsutake, Yoshiharu Mori, and Yuko Okamoto. Enhanced sampling algorithms. In *Biomolecular Simulations*, pages 153–195. Springer, 2013.

[294] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.*, 118(45):11225–11236, 1996.

[295] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Genet.*, 65(3):712–725, 2006.

[296] João Henriques and Marie Skepö. Molecular dynamics simulations of intrinsically disordered proteins: On the accuracy of the tip4p-d water model and the representativeness of protein disorder models. *J. Chem. Theory Comput.*, 12(7):3407–3415, 2016.

[297] Alexander P Lyubartsev and Alexander L Rabinovich. Force field development for lipid membrane simulations. *Biochim. Biophys. Acta - Biomembr.*, 1858(10):2483–2497, 2016.

[298] Richard Car and Mark Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471, 1985.

[299] Dominik Marx and Jürg Hutter. *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[300] Emre Sevgen, Federico Giberti, Hythem Sidky, Jonathan K Whitmer, Giulia Galli, Francois Gygi, and Juan J de Pablo. Hierarchical coupling of first-principles molecular dynamics with advanced sampling methods. *J. Chem. Theory Comput.*, 14(6):2881–2888, 2018.

[301] Michael Gaus, Qiang Cui, and Marcus Elstner. Density functional tight binding: Application to organic and biological molecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 4(1):49–61, 2014.

[302] Sharma S.R.K.C. Yamijala, Ma. Belén Oviedo, and Bryan M. Wong. *Density Functional Tight Binding Calculations for Probing Electronic-Excited States of Large Systems*, chapter 2, pages 45–79. John Wiley & Sons, Ltd, 2022.

[303] **Anshuman Kumar**, Zulfikhar A. Ali, and Bryan M. Wong. Efficient predictions of formation energies and convex hulls from density functional tight binding calculations. *J. Mater. Sci. Technol.*, 141:236–244, 2023.

[304] Mohammad Wahiduzzaman, Augusto F Oliveira, Pier Philipsen, Lyuben Zhechkov, Erik Van Lenthe, Henryk A Witek, and Thomas Heine. Dftb parameters for the periodic table: Part 1, electronic structure. *J. Chem. Theory. Comput.*, 9(9):4006–4017, 2013.

[305] Marcus Elstner and Gotthard Seifert. Density functional tight binding. *Philos. Trans. A Math. Phys. Eng. Sci.*, 372(2011):20120483, 2014.

[306] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[307] Yang Yang, Haibo Yu, Darrin York, Qiang Cui, and Marcus Elstner. Extension of the self-consistent-charge density-functional tight-binding method: Third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J. Phys. Chem. A*, 111(42):10861–10873, 2007.

[308] Ben Hourahine, Bálint Aradi, Volker Blum, Frank Bonafe, Alex Buccheri, Cristopher Camacho, Caterina Cevallos, MY Deshaye, T Dumitrică, A Dominguez, et al. Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.*, 152(12):124101, 2020.

[309] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. SIAM, Philadelphia, Pennsylvania, USA, third edition, 1999.

[310] Stanimire Tomov, Jack Dongarra, and Marc Baboulin. Towards dense linear algebra for hybrid gpu accelerated manycore systems. *Parallel Comput.*, 36(5-6):232–240, 2010.

[311] Jan JM Cuppen. A divide and conquer method for the symmetric tridiagonal eigenproblem. *Numer. Math.*, 36(2):177–195, 1980.

[312] Xia Liao, Shengguo Li, Yutong Lu, and Jose E Roman. A parallel structured divide-and-conquer algorithm for symmetric tridiagonal eigenvalue problems. *IEEE Trans. Parallel Distrib. Syst.*, 32(2):367–378, 2020.

[313] Baswanth Oruganti and Ran Friedman. Activation of abl1 kinase explored using well-tempered metadynamics simulations on an essential dynamics sampled path. *J. Chem. Theory Comput.*, 17(11):7260–7270, 2021.

[314] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 1(5):826–843, 2011.

[315] Daniel Sucerquia, Cristian Parra, Pilar Cossio, and Olga Lopez-Acevedo. Ab initio metadynamics determination of temperature-dependent free-energy landscape in ultrasmall silver clusters. *J. Chem. Phys.*, 156(15):154301, 2022.

[316] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[317] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174, 2004.

[318] Jilei Wang, Nan Zhang, Yurong Tan, Fanfan Fu, Gengxin Liu, Yin Fang, Xin-xing Zhang, Minsu Liu, Yuan Cheng, and Jing Yu. Sweat-resistant silk fibroin-based double network hydrogel adhesives. *ACS Applied Materials & Interfaces*, 2022.

[319] Erik G Marklund, Daniel SD Larsson, David van der Spoel, Alexandra Patriksson, and Carl Caleman. Structural stability of electrosprayed proteins: Temperature and hydration effects. *Physical Chemistry Chemical Physics*, 11(36):8069–8078, 2009.

[320] CT Reimann, I Velázquez, M Bittner, and O Tapia. Proteins in vacuo: A molecular dynamics study of the unfolding behavior of highly charged disulfide-bond-intact lysozyme subjected to a temperature pulse. *Physical Review E*, 60(6):7277, 1999.

[321] Christopher I Bayly, Piotr Cieplak, Wendy Cornell, and Peter A Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model. *J. Phys. Chem.*, 97(40):10269–10280, 1993.

[322] Pierre Turq, Frédéric Lantelme, and Harold L Friedman. Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.*, 66(7):3039–3044, 1977.

[323] David A Case, H Metin Aktulga, Kellon Belfon, Ido Ben-Shalom, Scott R Brozell, David S Cerutti, Thomas E Cheatham III, Vinícius Wilian D Cruzeiro, Tom A Darden, Robert E Duke, et al. *Amber 2021*. University of California, San Francisco, 2021.

[324] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[325] Berk Hess. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, 4(1):116–122, 2008.

[326] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185(2):604–613, 2014.

[327] Scott McMillan. Making containers easier with hpc container maker. In *In HPCSYSPROS18: HPC System Professionals Workshop*, Dallas, TX, November 2018.

[328] Peiyu Zhang, Geoffrey PF Wood, Jian Ma, Mingjun Yang, Yang Liu, Guangxu Sun, Yide A Jiang, Bruno C Hancock, and Shuhao Wen. Harnessing cloud architecture for crystal structure prediction calculations. *Cryst. Growth Des.*, 18(11):6891–6900, 2018.

[329] Yuhang Wang, Sruthi Murlidaran, and David A Pearlman. Quantum simulations of sars-cov-2 main protease mpro enable high-quality scoring of diverse ligands. *J. Comput. Aided Mol. Des.*, 35(9):963–971, 2021.

[330] Demian Riccardi, Peter König, Hua Guo, and Qiang Cui. Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor. *Biochem.*, 47(8):2369–2378, 2008.

[331] Yang Yang, Haibo Yu, and Qiang Cui. Extensive conformational transitions are required to turn on atp hydrolysis in myosin. *J. Mol. Biol.*, 381(5):1407–1420, 2008.

[332] Yang Yang and Qiang Cui. Does water relay play an important role in phosphoryl transfer reactions? insights from theoretical study of a model reaction in water and tert-butanol. *J. Phys. Chem. B*, 113(14):4930–4939, 2009.

[333] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys.*, 132(15):154104, 2010.

[334] Stefan Grimme, Stephan Ehrlich, and Lars Goerigk. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011.

[335] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics*, 81(1):511–519, 1984.

[336] William G Hoover. Canonical dynamics: Equilibrium phase-space dstributions. *Physical review A*, 31(3):1695, 1985.

[337] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[338] John L Gustafson. Reevaluating amdahl's law. *Commun. ACM*, 31(5):532–533, 1988.

[339] Birgit Strodel and David J. Wales. Free energy surfaces from an extended harmonic superposition approach and kinetics for alanine dipeptide. *Chem. Phys. Lett.*, 466(4):105–115, 2008.

[340] Pu Liu, Byungchan Kim, Richard A. Friesner, and B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.*, 102(39):13749–13754, 2005.

[341] Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125(2):024106, 2006.

[342] Simone Marsili, Alessandro Barducci, Riccardo Chelli, Piero Procacci, and Vincenzo Schettino. Self-healing umbrella sampling: A non-equilibrium approach for quantitative free energy calculations. *J. Phys. Chem. B*, 110(29):14011–14013, 2006.

[343] Hythem Sidky and Jonathan K. Whitmer. Learning free energy landscapes using artificial neural networks. *J. Chem. Phys.*, 148(10):104111, 2018.

[344] Jerome Cuny, Kseniia Korchagina, Chemseddine Menakbi, and Tzonka Mineva. Metadynamics combined with auxiliary density functional and density functional tight-binding methods: Alanine dipeptide as a case study. *J. Mol. Model.*, 23(3):72, 2017.

[345] Jakob J Malin, Isabelle Suárez, Vanessa Priesner, Gerd Fätkenheuer, and Jan Rybniker. Remdesivir against covid-19 and other viral diseases. *Clin. Microbiol. Rev.*, 34(1):e00162–20, 2020.

[346] Lisa E Gralinski and Vineet D Menachery. Return of the coronavirus: 2019-ncov. *Viruses*, 12(2):135, 2020.

[347] Timothy P Sheahan, Amy C Sims, Sarah R Leist, Alexandra Schäfer, John Won, Ariane J Brown, Stephanie A Montgomery, Alison Hogg, Darius Babusis, Michael O Clarke, et al. Comparative therapeutic efficacy of remdesivir and combination lopinavir, ritonavir, and interferon beta against mers-cov. *Nat. Commun.*, 11(1):1–14, 2020.

[348] Petr Hošek and Vojtěch Spiwok. Metadyn view: Fast web-based viewer of free energy surfaces calculated by metadynamics. *Comput. Phys. Commun.*, 198:222–229, 2016.

[349] Garng M. Huang, T. J. Tarn, and John W. Clark. On the controllability of quantum-mechanical systems. *J. Math. Phys.*, 24(11):2608–2618, 1983.

[350] M. Belén Oviedo and Bryan M. Wong. Real-time quantum dynamics reveals complex, many-body interactions in solvated nanodroplets. *J. Chem. Theory Comput.*, 12(4):1862–1871, 2016.

[351] Niranjan V. Ilawe, M. Belén Oviedo, and Bryan M. Wong. Real-time quantum dynamics of long-range electronic excitation transfer in plasmonic nanoantennas. *J. Chem. Theory Comput.*, 13(8):3442–3454, 2017.

[352] Margherita Maiuri, Maria B. Oviedo, Jacob C. Dean, Michael Bishop, Bryan Kudisch, Zi S. D. Toa, Bryan M. Wong, Stephen A. McGill, and Gregory D. Scholes. High magnetic field detunes vibronic resonances in photosynthetic light harvesting. *J. Phys. Chem. Lett.*, 9(18):5548–5554, 2018.

[353] Bryan Kudisch, Margherita Maiuri, Luca Moretti, Maria B. Oviedo, Leon Wang, Daniel G. Oblinsky, Robert K. Prud'homme, Bryan M. Wong, Stephen A. McGill, and Gregory D. Scholes. Ring currents modulate optoelectronic properties of aromatic chromophores at 25 t. *Proc. Natl. Acad. Sci. U.S.A.*, 117(21):11289–11298, 2020.

[354] K. C. Nowack, F. H. L. Koppens, Yu. V. Nazarov, and L. M. K. Vandersypen. Coherent control of a single electron spin with electric fields. *Science*, 318(5855):1430–1433, 2007.

[355] Michael Kues, Christian Reimer, Piotr Roztocki, Luis Romero Cortés, Stefania Sciara, Benjamin Wetzel, Yanbing Zhang, Alfonso Cino, Sai T. Chu, Brent E. Little, David J. Moss, Lucia Caspani, José Azaña, and Roberto Morandotti. On-chip generation of high-dimensional entangled quantum states and their coherent control. *Nature*, 546(7660):622–626, Jun 2017.

[356] Evan M. Fortunato, Marco A. Pravia, Nicolas Boulant, Grum Teklemariam, Timothy F. Havel, and David G. Cory. Design of strongly modulating pulses to implement precise effective hamiltonians for quantum information processing. *J. Chem. Phys.*, 116(17):7599–7606, 2002.

[357] H. J. Williams, L. Caldwell, N. J. Fitch, S. Truppe, J. Rodewald, E. A. Hinds, B. E. Sauer, and M. R. Tarbutt. Magnetic trapping and coherent control of laser-cooled molecules. *Phys. Rev. Lett.*, 120:163201, Apr 2018.

[358] Allon Bartana, Ronnie Kosloff, and David J. Tannor. Laser cooling of molecules by dynamically trapped states. *Chem. Phys.*, 267(1):195 – 207, 2001.

[359] Benjamin L. Brown, Alexander J. Dicks, and Ian A. Walmsley. Coherent control of ultracold molecule dynamics in a magneto-optical trap by use of chirped femtosecond laser pulses. *Phys. Rev. Lett.*, 96:173002, May 2006.

[360] M. J. Wright, J. A. Pechkis, J. L. Carini, S. Kallush, R. Kosloff, and P. L. Gould. Coherent control of ultracold collisions with chirped light: Direction matters. *Phys. Rev. A*, 75:051401, May 2007.

[361] Paul Brumer and Moshe Shapiro. Coherence chemistry: controlling chemical reactions [with lasers]. *Acc. Chem. Res.*, 22(12):407–413, 1989.

[362] József Somlói, Vladimir A. Kazakov, and David J. Tannor. Controlled dissociation of i2 via optical transitions between the x and b electronic states. *Chem. Phys.*, 172(1):85 – 98, 1993.

[363] Wusheng Zhu, Jair Botina, and Herschel Rabitz. Rapidly convergent iteration methods for quantum optimal control of population. *J. Chem. Phys.*, 108(5):1953–1963, 1998.

[364] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser. Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms. *J. Magn. Reson.*, 172(2):296 – 305, 2005.

[365] Wusheng Zhu and Herschel Rabitz. A rapid monotonically convergent iteration algorithm for quantum optimal control over the expectation value of a positive definite operator. *J. Chem. Phys.*, 109(2):385–391, 1998.

[366] Akber Raza, Sharmistha Bardhan, Lihua Xu, Sharma S. R. K. C. Yamijala, Chao Lian, Hyuna Kwon, and Bryan M. Wong. A machine learning approach for predicting defluorination of per- and polyfluoroalkyl substances (pfas) for their efficient treatment and removal. *Environ. Sci. Technol. Lett.*, 6(10):624–629, 2019.

[367] Konstantin Gubaev, Evgeny V. Podryabinkin, and Alexander V. Shapeev. Machine learning of molecular properties: Locality and active learning. *J. Chem. Phys.*, 148(24):241727, 2018.

[368] Kiran K. Yalamanchi, Vincent C. O. van Oudenhoven, Francesco Tutino, M. Monge-Palacios, Abdulelah Alshehri, Xin Gao, and S. Mani Sarathy. Machine learning to predict standard enthalpy of formation of hydrocarbons. *J. Phys. Chem. A*, 123(38):8305–8313, 2019.

[369] Kevin Ryan, Jeff Lengyel, and Michael Shatruk. Crystal structure prediction via deep learning. *J. Amer. Chem. Soc.*, 140(32):10158–10168, 2018.

[370] Bing Huang and O. Anatole von Lilienfeld. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *J. Chem. Phys.*, 145(16):161102, 2016.

[371] Yang Liu, Nicholas Marcella, Janis Timoshenko, Avik Halder, Bing Yang, Lakshmi Kolipaka, Michael. J. Pellin, Soenke Seifert, Stefan Vajda, Ping Liu, and Anatoly I. Frenkel. Mapping xanes spectra on structural descriptors of copper oxide clusters using supervised machine learning. *J. Chem. Phys.*, 151(16):164201, 2019.

[372] Thaer M. Dieb, Zhufeng Hou, and Koji Tsuda. Structure prediction of boron-doped graphene by machine learning. *J. Chem. Phys.*, 148(24):241716, 2018.

[373] Burak Himmetoglu. Tree based machine learning framework for predicting ground state energies of molecules. *J. Chem. Phys.*, 145(13):134101, 2016.

[374] Pavlo O. Dral. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.*, 11(6):2336–2347, 2020.

[375] Jörg Behler. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.*, 145(17):170901, 2016.

[376] Philipp von den Hoff, Sebastian Thallmair, Markus Kowalewski, Robert Siemering, and Regina de Vivie-Riedle. Optimal control theory – closing the gap between theory and experiment. *Phys. Chem. Chem. Phys.*, 14:14460–14485, 2012.

[377] Sebastian Thallmair, Daniel Keefer, Florian Rott, and Regina de Vivie-Riedle. Simulating the control of molecular reactions via modulated light fields: from gas phase to solution. *J. Phys. B At. Mol. Opt. Phys.*, 50(8):082001, mar 2017.

[378] Tobias Brixner and Gustav Gerber. Quantum control of gas-phase and liquid-phase femtochemistry. *ChemPhysChem*, 4(5):418–438, 2003.

[379] Marcos Dantus and Vadim V. Lozovoy. Experimental coherent laser control of physicochemical processes. *Chem. Rev.*, 104(4):1813–1860, 2004.

[380] Bryan M. Wong and Sumathy Raman. Thermodynamic calculations for molecules with asymmetric internal rotors—application to 1,3-butadiene. *J. Comput. Chem.*, 28(4):759–766, 2007.

[381] Bryan M. Wong, Maria M. Fadri, and Sumathy Raman. Thermodynamic calculations for molecules with asymmetric internal rotors. ii. application to the 1,2-dihaloethanes. *J. Comput. Chem.*, 29(3):481–487.

[382] Hans A Bechtel, Adam H Steeves, Bryan M Wong, and Robert W Field. Evolution of chemical bonding during hcn⇌hnc isomerization as revealed through nuclear quadrupole hyperfine structure. *Angew. Chem. Int. Ed.*, 47(16):2969–2972.

[383] Bryan M. Wong. Nuclear quadrupole hyperfine structure in hc14n/h14nc and dc15n/d15nc isomerization: a diagnostic tool for characterizing vibrational localization. *Phys. Chem. Chem. Phys.*, 10:5599–5606, 2008.

[384] Bryan M. Wong, Ryan L. Thom, and Robert W. Field. Accurate inertias for large-amplitude motions: Improvements on prevailing approximations. *J. Phys. Chem. A*, 110(23):7406–7413, 2006.

[385] Guillaume Reinisch, Kenji Miki, Gérard L. Vignoles, Bryan M. Wong, and Chris S. Simmons. An efficient and accurate formalism for the treatment of large amplitude intramolecular motion. *J. Chem. Theory Comput.*, 8(8):2713–2724, 2012.

[386] Bryan M. Wong, Adam H. Steeves, and Robert W. Field. Electronic signatures of large amplitude motions: Dipole moments of vibrationally excited local-bend and local-stretch states of s0 acetylene. *J. Phys. Chem. B*, 110(38):18912–18920, 2006.

[387] Kirill Prozument, Rachel Glyn Shaver, Monika A. Ciuba, John S. Muenter, G. Barratt Park, John F. Stanton, Hua Guo, Bryan M. Wong, David S. Perry, and Robert W. Field. A new approach toward transition state spectroscopy. *Faraday Discuss.*, 163:33–57, 2013.

[388] Akber Raza, Chengkuan Hong, Xian Wang, **Anshuman Kumar**, Christian R. Shelton, and Bryan M. Wong. NIC-CAGE: An open-source software package for predicting optimal control fields in photo-excited chemical systems. *Comput. Phys. Commun.*, 258:107541, 2021.

[389] Li Li, John C. Snyder, Isabelle M. Pelaschier, Jessica Huang, Uma-Naresh Niranjan, Paul Duncan, Matthias Rupp, Klaus-Robert Müller, and Kieron Burke. Understanding machine-learned density functionals. *Int. J. Quantum Chem.*, 116(11):819–833, 2016.

[390] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[391] Nathan Whitehead and Alex Fit-Florea. Precision and performance: Floating point and IEEE 754 compliance for NVIDIA GPUs, 2019. CUDA libraries available from www.nvidia.com.

[392] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke. Finding density functionals with machine learning. *Phys. Rev. Lett.*, 108:253002, Jun 2012.

[393] Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the kohn-sham equations with machine learning. *Nat. Commun.*, 8(1):872, Oct 2017.