Doctoral Thesis in Theoretical Chemistry and Biology

Accurate Force Fields for Spectroscopic Studies of Protein–Ligand Interactions and Self-Assembly Structures

YOGESH TODARWAL

Stockholm, Sweden 2024
Accurate Force Fields for Spectroscopic Studies of Protein–Ligand Interactions and Self-Assembly Structures

YOGESH TODARWAL

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 2nd February 2024, at 9:00 a.m. in F3, Lindstedtsvägen 26, Stockholm.
To the friends who guided me, to the mentors who inspired me.
Abstract

The computational prediction of complex molecular behaviors is an essential component of modern chemistry, as it provides a faster and more cost-effective way to explore molecular interactions that may be difficult or even impossible to study experimentally. Molecular dynamics (MD) simulations often serve as a valuable tool for such predictions; however, their accuracy is inherently dependent on the force field (FF) parameters employed. While the general amber force field (GAFF) is designed to provide reasonable results for a broad array of small molecules, it often requires further refinement when using it for a specific small organic molecule. Especially for ligands of the oligothiophene class, the dihedral potential representing the rotatable bond between the two thiophene rings (of the SCCS type) is inadequately described.

An objective of this dissertation is to refine FF parameters for producing meaningful MD trajectories that capture key molecular interactions, binding modes, and thermodynamic properties, and subsequent accurate calculations of spectroscopic properties. The refined FF parameters were first tested by comparing the dihedral potential derived from the FF method to the density functional theory (DFT) based dihedral potential. They were then validated by assessing the relative energies of conformers optimized using both FF and DFT methods, and by comparing the transition wavelengths calculated based on geometries optimized with both FF and DFT approaches. Importantly, the errors in dihedral potential were kept below 1 kcal/mol, and the discrepancies in transition energies were less than 0.1 eV for molecular transitions around 5 eV.

This FF parametrization methodology was used in research studies focusing on two classes of supramolecular systems: host-guest chemistry related to neurodegenerative diseases, and self-assembly systems for material development. Specifically, we examine host-guest interactions involving proteins such as amyloid-beta, tau, and transthyretin (TTR), which are associated with neurodegenerative diseases. Various fluorescent ligands are used for the detection of these proteins in pathological samples. Our results for these protein–ligand systems propose strong binding sites and modes, and include estimations of binding energies for different ligands interacting with the targeted proteins. Additionally, comparative studies among the ligands have been conducted. Interestingly, no fluorescence was observed when low binding energy ligands interacted with amyloid fibrils. In the case of bTVBT4 binding to tau associated with Alzheimer’s disease (AD), a unique binding site was identified. This site was not accessible in the tau fold found in Pick’s disease (PiD), thus explaining the specificity of bTVBT4 for AD-related tau. For self-assembly systems, our findings encompass spectral profiles altered by tyrosine substitutions in oligothiophenes, a stable self-assembly model formed by chiral sulfonimidamides.
that explains the involved interactions, and comparisons of experimental circular dichroism (CD) profiles to assign isomers of [4]cyclonaphthodithiophene diimides to specific spectral profiles. We also investigated the solvent effects on the spectroscopic properties of symmetric and asymmetric azaoxahelicenes.

In conclusion, the methodological development of FF parameters provides a robust framework for accurately modeling the behavior of complex supramolecular systems. The improvements in the dihedral potential align closely with DFT-based calculations, thereby elevating the predictive power of MD simulations for both binding modes and subsequent spectroscopic properties. The research has direct applications in the detection of neurodegenerative diseases at an early stage by designing fluorescent ligands that specifically bind to targeted proteins. It also contributes to the creation of advanced materials with finely-tuned properties. Furthermore, the methodology employed can serve as a blueprint for future studies aiming to refine computational models for other classes of molecules.
Svensk sammanfattning

Den beräkningsmässiga förutsägelsen av komplext molekylärt beteende är en viktig komponent i modern kemi, eftersom det ger ett snabbare och mer kostnadseffektivt sätt att utforska molekylära interaktioner som kan vara svåra eller till och med omöjliga att studera experimentellt. Molecular dynamics (MD)-simuleringar fungerar ofta som ett värdefullt verktyg för sådana förutsägelser; deras noggrannhet är emellertid i sig beroende av de använda kraftfältsparametrarna (FF). Medan det allmänna bårnstenskraftfältet (GAFF) är utformat för att ge rimliga resultat för ett brett spektrum av små molekyler, kräver det ofta ytterligare förfining när det används för en specifik liten organisk molekyl. Speciellt för ligander av oligofotenklassen är den dihedrala potentialen som representerar den roterbara bindningen mellan de två tiofentoringarna (av SCCS-typ) otillräckligt beskriven.

Ett syfte med denna avhandling är att förfina FF-parametrar för att producera meningsfulla MD-banor som fångar viktiga molekylära interaktioner, bindningslägen och termodynamiska egenskaper, och efterföljande noggranna beräkningar av spektroskopiska egenskaper. De förfinade FF-parametrarna testades först genom att jämföra den dihedrala potentialen härledd från FF-metoden med den densitetsfunktionella teorin (DFT)-baserade dihedrala potentialen. De validerades sedan genom att bedöma de relativa energierna för konformers optimerade med både FF- och DFT-metoder, och genom att jämföra övergångsvåglängderna som beräknats baserat på geometrier optimerade med både FF- och DFT-metoder. Viktigt är att felen i dihedral potential hölls under 1 kcal/mol, och avvikelserna i övergångsenergier var mindre än 0,1 eV för molekylära övergångar runt 5 eV.

tyrosinsubstitutioner i oligotiofener, en stabil självmonteringsmodell bildad av kirala sulfonimidamider som förklarar de inblandade interaktionerna, och jämförelser av experimentella cirkulär dikroism (CD) profiler för att tilldela isomerer av [4] cyklonaftoditiofendiimider till specifika spektrala profiler. Vi undersökte också lösningsmedelseffekterna på de spektroskopiska egenskaperna hos symmetriska och asymmetriska azaoxahelicener.

Acknowledgements

My PhD journey has been filled with immense professional and spiritual growth. Words often fall short when attempting to express the gratitude I feel towards many individuals who have been an integral part of this transformative chapter of my life. Nonetheless, I will try my best to express my heartfelt appreciation.

First and foremost, I am extremely thankful and want to express my profound gratitude to my supervisor, Patrick Norman, for giving me this golden opportunity to pursue research under his guidance. His detailed approach and constructive feedback have been invaluable in shaping my academic path. He always remained calm during research discussions, consistently providing me with different perspectives to interpret the same results. He was always patient and understanding when I faced research challenges. I’m grateful for his support and for providing the best possible resources, which have enhanced my work at every step. I couldn’t ask for a better mentor, as his constructive criticism and personalized attention have immensely helped me hone my skills.

Next, my co-supervisor, Mathieu Linares, has been a force of energy and drive. His lively and strong-willed approach has fueled my desire to reach higher and strengthened my commitment to science. His push and constant challenges were not mere obstacles but spark that moulded me into a more focused and sharp researcher. His role has been instrumental in guiding me towards my scientific goals. I am also grateful to Jacob Kongsted for his guidance during my PhD secondment at SDU, Denmark.

Additionally, I’d like to express my gratitude to my collaborators: Peter Nilsson, Kurt Mikkelsen, Carolin König, Mikael Lindgren, Nghia Nguyen, and Andreas Bendtsen. I extend my thanks to Thomas Fransson, Karan Ahmadzadeh, and Yaoquan Tu for proofreading this thesis and assisting in identifying errors.

I would like to convey my heartfelt gratitude to my colleagues who have also become dear friends. Their support, guidance, and companionship have made this journey both memorable and rewarding. Camilla Gustavsson deserves special mention for her constant assistance since my first day in Sweden. Her insights have profoundly transformed my presentation style, a change that has been immensely beneficial throughout my journey. With Manuel Brand, our shared moments, filled with laughter and joy, are forever engraved in my memory. His uplifting spirit was a constant source of positivity. Karan Ahmadzadeh consistently provided profound intellectual discussions about career and growth, conversations that have been instrumental in shaping my perspectives and ambitions. I am thankful to Dusanka Golo for those marathon group study sessions; her dedication and companionship made those long
hours bearable. I’m also grateful to Viktoriia Savchenko for adding a refreshing touch of fun and adventures outside of our academic setting. Spiritual discussions with Juan de Gracia and Ke Ye have provided me with much-needed depth and perspective, grounding me in moments of doubt.

Equally important are Lluis Artús, Gabriel Rodrigues, Zilvinas Rinkevicius, Mårten Ahlquist, Victor Kimberg, Olav Vahtras, Faris Gelmukhanov, Xiaoyu Chen, Michal Biler, Markéta Paloncýová, Pratip Chakraborty, Qing Liu, Ge Li, Pouria Farahani, Manuel Hodecker, Lulia Brumboiu, Valentin Lindfeld, and Lukas Lampe. Throughout various stages of my journey, each of you added a unique vibrancy with your presence, making the department feel like a second home.

As I look back on the past five years at the TCB, they seem to have passed in the blink of an eye, leaving me with many memories and learnings that I will cherish forever.

Yogesh Todorwal
Stockholm, October 2023
# CONTENTS

1 Introduction 3

2 Quantum Mechanics 13
   2.1 Fundamentals .............................................. 13
   2.2 Born–Oppenheimer Approximation ......................... 14
   2.3 Variational Principle .................................... 16
   2.4 Hartree–Fock Method .................................... 17
   2.5 Density Functional Theory ............................... 20

3 Force Fields 23
   3.1 Principles .................................................. 23
   3.2 Parameterization .......................................... 26
   3.3 Validation .................................................. 31
   3.4 Limitations ................................................ 33

4 Molecular Dynamics Simulations 35
   4.1 Configurational Sampling ................................ 35
   4.2 Periodic Boundary Conditions ........................... 37
   4.3 Interatomic Interactions ................................. 37
   4.4 Thermostat and Barostat ................................ 39
   4.5 Simulation Setup and Protocols ......................... 39
   4.6 Analysis ..................................................... 41
   4.7 Limitations ................................................ 45

5 Spectroscopic Properties of Molecules 47
   5.1 Response Theory ............................................ 47
   5.2 Polarizable Embedding .................................... 51
   5.3 Conformational Averaging ............................... 54

6 Overview of Article Findings 57

Bibliography 67
Papers

List of papers and my contributions

Paper I 83
  **Tau Protein Binding Modes in Alzheimer’s Disease for Cationic Luminescent Ligands**

Paper II 95
  **Comparative Analysis of Luminescent Cationic, Anionic, and Neutral Ligands Binding to Amyloid Fibrils in Alzheimer’s Disease: A Computational Insight**
  Manuscript

Paper III 135
  **Distinct Heterocyclic Moieties Govern the Selectivity of Thiophene-Vinylene-based Ligands Towards Aβ or Tau Pathology in Alzheimer’s Disease**

Paper IV 151
  **Binding of a Pyrene-Based Fluorescent Amyloid Ligand to Transthyretin: A Combined Crystallographic and Molecular Dynamics Study**

Paper V 161
  **Tyrosine Side-Chain Functionalities at Distinct Positions Determine the Chirooptical Properties and Supramolecular Structures of Pentameric Oligothiophenes**
  *ChemistryOpen*, 9: 1100–1108, 2020

Paper VI 173
  **Naphthodithiophene Diimide Based Chiral π-Conjugated Nanopillar Molecules**

Paper VII 181
  **Ultralight aerogels via supramolecular polymerization of a new chiral perfluoropyridine-based sulfonimidamide organogelator**
  Manuscript Submitted to Nanoscale

Paper VIII 219
  **Modeling Absorption and Emission Spectroscopies of Symmetric and Asymmetric Azaoxahelicenes in Vacuum and Solution**
Simulating Reality: Where ordinary molecules mingle, party hard, and emerge as sophisticated Supramolecules!
INTRODUCTION

In the rapidly evolving modern science, the exploration of complex supramolecular systems and the molecular interactions fostering them is of crucial importance. These studies not only deepen our understanding of fundamental molecular mechanisms but also fuel the development of material innovation, advanced therapeutic and diagnostic strategies. Supramolecular structures are formed when multiple molecules or groups of molecules interact with each other—non-covalently through hydrogen bonding, electrostatic interactions, hydrophobic interactions and van der Waals forces. These interactions play a critical role in forming host-guest complexes, self-assembly systems, and supramolecular polymers. Such systems have the potential for creating new materials with novel properties that are not present in the individual constituent molecules. This inherently complex field is often described as "chemistry beyond the molecule" and it forms an interface with both biological and materials science. Within this vast scientific landscape, the primary focus of the thesis is on two key aspects: host-guest chemistry and self-assembly systems.

Host-guest chemistry exemplifies molecular recognition systems, where the unique electronic and structural characteristics of the guest and host dictate their interaction and function. A prime example is the protein-ligand system, with the protein serving as the host and the ligand as the guest. Proteins possess binding sites with specially shaped cavities, providing an ideal environment for the guest ligand to bind. For a protein host, a guest ligand is designed to fit in specific ways, much like a key fits into a lock. The shape, size, charge, and functional groups of a ligand can be tailored to specifically bind to the target protein host. This precise fitting is crucial in various biomedical research areas, including detecting misfolded proteins, as seen in Alzheimer’s disease. Here, the ligand is not only designed to fit the protein but also possesses unique spectroscopic properties that change upon binding. These alterations in spectroscopic properties can be employed to specifically identify misfolded proteins.

On the other hand, self-assembly is a process wherein simple molecular units organize into complex structures under the influence of intermolecular non-covalent interactions. This process can also be seen as "bottom-up" self-assembly, as it allows for building a complex structure using molecular subunits. Molecules can self-assemble into diverse structures such as micelles, vesicles, nanofibers, and various nano- and macrostructures, and they have applications in a wide range of scientific fields, such as drug delivery, optoelectronics, and biosensing.

Combining experimental and computational methods provides a comprehensive approach to explore the complex domain
of host–guest and self-assemble supramolecular systems. The experimental methods such as fluorescence spectroscopy for host–guest complexes provides information about understanding of disease progression, examine binding interactions, and help in detecting the disease at an early stage. On the other hand, circular dichroism (CD) spectroscopy on self-assembled supramolecular systems helps identify structural features, study chirality, monitor system stability and dynamic behavior, and detect conformational changes in response to stimuli.10

Complementing the experimental studies, computational studies offer deeper insight into atomic-level details of these systems. In addition, they allows us to test-run potentially hazardous or expensive experiments in silico, significantly reducing both risk and cost. For protein–ligand systems (host–guest chemistry), computational methods enable the exploration of the conformational landscape, binding energies, and dynamics of these systems. Methods such as molecular dynamics (MD) simulations, spatial distribution function and umbrella sampling can predict binding poses, analyze protein–ligand interactions, and quantify binding affinities,11 thus offering insights into how ligands interact with their protein targets. In addition, theoretical spectra can also be calculated and compared with experimental spectra to validate molecular models and for providing information about shifts in absorption spectra of ligand in different surroundings (for studying environment influences on optical properties of the ligand). Computational methods can also be used to interpret the experimental results. For instance, in paper I, we explain why bTVBT4 binds exclusively to the tau fold from Alzheimer’s disease, but not to the tau fold in Pick’s disease. For molecules undergoing self-assembly, computational chemistry allows the study of assembly mechanisms at a level of detail that is difficult to achieve experimentally. It can model how simple components can spontaneously organize into complex structures, offering predictions and guiding principles to design and control self-assembly processes. In addition, CD spectroscopy gives information of a secondary structure, conformational changes, and assembly process of chiral molecules and aggregates. When aligned with experimental CD spectra, these computational results can serve to validate the predicted structures and assembly mechanisms.

Despite the advantages of computational methods, we must recognize that they are not intended to replace experimental methods but to complement them. Computational and experimental chemistry are two sides of the same coin — the former offers depth of detail and the ability to predict and explain, while the latter provides evidence and practical applications. They work together, complementing each other to overcome their respective limitations.
The size and complexity of the systems studied in this thesis can range from a simple molecule (containing < 200 atoms) to the massively complex systems (containing up to 3 million atoms) such as protein–ligand or self-assembled structures. To efficiently model this broad spectrum of system size and complexity necessitates to use a variety of computational methods, with each having their own strengths and limitations. At one end of the spectrum we have quantum mechanical (QM) methods, which are designed to accurately model the behavior of electrons in a molecular system. These methods, which include density functional theory (DFT), provide an accurate description of molecular structures and electronic properties. However, due to their computational expense, QM methods are often limited to relatively small systems. For larger systems, like proteins or large supramolecular assemblies, the use of QM methods becomes computationally prohibitive. To address this, molecular mechanics (MM) methods are employed, which use a classical physics approach to model atoms and their interactions. While MM methods are less accurate than QM methods because they do not explicitly consider electronic structures, they are computationally efficient and can handle large systems comprising thousands to millions of atoms.

However, there are scenarios where both the large size of the system and the accurate treatment of electronic structure are crucial, such as studying the electronic spectra of a ligand within a large protein environment. To tackle such cases, hybrid quantum mechanics/molecular mechanics (QM/MM) methods are employed. In QM/MM methods, the region of interest (like the ligand) is treated with a QM method to accurately represent its electronic structure, while the rest of the system (like the surrounding protein environment) is modeled using MM methods. This approach allows for the accurate modeling of key parts of large systems without an overwhelming computational cost, providing a pragmatic compromise between accuracy and efficiency.

Central to the effectiveness of MM methods, and by extension QM/MM methods, is the accuracy of force field (FF) parameters. The FF defines the potential energy of a system as a function of its atomic positions, this allow for calculating force acting on each atom and thus simulating the behavior of the system. The positions, orientations, and interactions of atoms are all influenced by the parameters of the FF. So if the FF is inaccurate, the simulated structures from dynamics will be unreliable, and potentially leading to erroneous conclusions. Moreover, accurate FFs allow for a more realistic description of the energy landscape of the system. This is particularly important when studying phenomena like conformational changes, binding events, or self-assembly processes, where the system transitions between different energy states. In addition, estimating spectroscopic properties based on snapshots derived from dynamics requires
accurate FF parameters. Imprecise FF parameters can lead to calculated spectra that are likely to be off the mark. Therefore, a primary aim of this thesis is to develop and optimize FF parameters that are tailored to yield meaningful MD trajectories and subsequent accurate calculations of spectroscopic properties. The process involves careful parameterization of initial FF parameters by relying on high-quality data from QM method, to ensure that they reflect the electronic structure and corresponding spectroscopic properties of the molecules under study. Subsequently, testing and validation against known QM results to ensure the reliability and applicability of the developed FF in the study of protein–ligand and self-assembled complex systems.

Systems under Study

In the articles attached to this thesis, two primary supramolecular systems are examined: the host (protein)–guest (ligand) system and the self-assembly system. Within the context of protein–ligand systems, the discussion is centered on the amyloid-beta (Aβ(1–42)), tau, and TTR proteins. Whereas for the self-assembly systems, we explore nanopillars, low molecular weight gels, and helicenes. Additionally, details on how the initial models for these systems were constructed are provided.

HOST–GUEST SYSTEMS

Aβ(1–42) and tau protein fibrils

Even after decades of enhancement in technology and research, Alzheimer’s disease (AD) still remains an elusive challenge without a definitive solution. The Aβ(1–42) (main composition of extracellular plaques) and tau fibrils (main composition of intraneuronal tangles) are the primary pathological hallmark of the disease. Initially, the Aβ(1–42) deposits are found in the neocortex region while the tau fibrils in the limbic system. In the next phase of the disease, some tau fibrils propagate towards the neocortex region and Aβ(1–42) towards the limbic system. According to one of the hypotheses, this propagation causes a change from a preclinical to a clinical state, making it necessary to detect these fibrils at an early stage.

Detecting the Aβ(1–42) and tau fibril deposits at an early stage is important for timely intervention and potentially slowing down the progression of AD. The non invasive techniques such as positron emission tomography (PET) and fluorescence spectroscopy have emerged as increasingly valuable tools in this regard. For fluorescence studies, ligands — small hydrophobic molecules — are used to detect the deposits in a pathological samples. They help visualize, track the progression of fibrils, and biomarker development. To be used as an amyloid ligand,
they should also exhibit distinct spectroscopic properties in different environments. For in vivo studies, these effective ligands can subsequently be transformed into PET tracers by labeling one of the atoms within the ligand.\textsuperscript{29}

In order to model and simulate this protein–ligand complex, we need initial coordinates for the amyloid fibrils. Modeling the starting protein structure poses significant challenges due to the highly complex and dynamic nature of protein folding. The vast conformational space, which contains a large number of possible structures, makes it computationally expensive to explore and identify the native folded state. Additionally, proteins folding events happens at long time scales, ranging from nanoseconds to milliseconds, further increasing the computational cost.\textsuperscript{30} Hence for the initial protein structures used in MD simulations, we rely on experimental approaches such as X-ray crystallography and cryo-electron microscopy (cryo-EM) data.

A study by Fitzpatrick \textit{et al.} successfully deduced the cryo-EM structure of the tau protein,\textsuperscript{31} revealing two ultrastructural polymorphs: paired helical filaments (PHFs) and straight filaments (SFs). The key distinction between these polymorphs lies in the repeating units. Although they are composed of two identical protofilaments, they differ in their inter-protofilament packing. Each protofilament structure is comprised of 73 amino acids. In the PHFs, the repeating units stack together to form an elongated fibril structure. These consecutive repeating units in the fibril structure display an approximate twist of 1.0° and a distance of 4.7 Å between them. (see figure 1.1)

Utilizing the cryo-EM structure (PDB ID 5O3L), the coordinates of one protofilament were extracted. Next, the base unit containing two protofilaments was created; the second protofilament was rotated by 180° relative to the first in the X Y-plane, together forming the base unit. Subsequently, the fibril with the half pitch was systematically constructed by rotating and translating the consecutive base units along the Z-direction. It was ensured that the last base unit had a desired twist of 180° relative to the first. The strategy involved using this half-pitch fibril in combination with periodic boundary conditions. In this setup, the half-pitch fibril perfectly aligned with adjacent periodic images of it in the Z-direction, allowing us to model an infinitely long tau fibril. Using this approach, we constructed multiple tau models with varying twists between consecutive base units (0.8°, 0.9°, 0.97°, 1.0°, and 1.2°) and a translating distance of 5 Å.

For each tau model, simulated annealing was conducted to slowly warm the system under NPT conditions. The temperature of the system was increased from 150 K to 300 K, utilizing 14 annealing points with 25 K increments. The temperature was raised every 200 ps, and at each step, the system was simulated for 2 ns (except for the final 300 K step, which was simulated for 13 ns). It was observed that the tau fibril model with an

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Schematic Representation of Tau Fibril Formation. a) Depicts a top view of a single tau protofilament. b) Showcases the assembly of two tau protofilaments in X Y-plane, resulting in the formation of a base unit. c) Base units stack vertically, creating the tau fibril structure with each unit angled approximately 1.0° relative to the adjacent base unit.}
\end{figure}
adjacent base unit twist of 0.97° exhibited fewer kinks compared to models with different twist angles. The dynamics of the half-pitch length of the tau fibril along its axis is illustrated in figure 1.2 for the model with a 0.97° twist. The simulation at the final 300 K step revealed an equilibrium structure from 14 ns onwards, with an average distance of 4.8 Å between two adjacent base units. In comparison, the cryo-EM experimental structure of the tau fibril has adjacent base unit distance of 4.7 Å at cryogenic temperatures. Despite this minor difference, the model fibril closely resembles the experimentally observed structure. The stability and fewer kinks in the tau model with a 0.97° twist make it a reliable representation of the tau fibril structure under physiological conditions. This tau fibril structure after 25 ns, was used as starting structure to study interaction of various ligands with tau. The model for the Aβ(1–42) fibril was adopted from previous work done by König et al., which was also constructed based on cryo-EM data.

TTR Protein

Transthyretin (TTR) is a tetrameric protein which acts as transporter of the thyroid hormone thyroxine and retinol in a human body. Under certain conditions, TTR can misfold and lead to the formation of insoluble amyloid fibrils. This process alters their native three-dimensional structure, which is necessary for their correct functioning. This can lead to diseases known as TTR amyloidoses. Similar to Alzheimer’s disease detection approach, the fluorescence Py1SA ligand could be used in the detection of TTR protein and the investigation of its misfolding mechanisms.

For modeling the starting structure of TTR interactions with the Py1SA ligand, X-ray crystallography data containing four distinct configurations of the TTR-Py1SA complex were used.
The structures of these complexes were then refined, which involved several steps. First, the structures were cleaned by removing water molecules, then adding missing residues in the two chains of TTR protein, and finally adding hydrogen atoms to amino acid residues while maintaining them at their physiological pH state. This was done with the help of tools like PDBFixer\textsuperscript{35}, GaussView\textsuperscript{36}, and the pdb2gmx routine of GROMACS.\textsuperscript{37–40} The obtained four final structures were used as starting structures for the MD simulation. The simulations were performed by our collaborators.

**SELF-ASSEMBLY SYSTEMS**

Nanopillar

The \([4]\text{cyclo-naphthodithiophene diimides ([4]C-NDTIs)}\) represent a novel class of \(\pi\)-conjugated macrocycles with a unique pillar-shaped topology. The \([4]\text{C-NDTI)}\) unit possesses inherent chirality due to its thiophene–thiophene junctions. Additionally, they display important properties, such as absorption in the near-infrared (NIR) region, strong excitonic coupling, and the capability to encapsulate C60 molecules tightly. These properties and potential applications based on them make \([4]\text{C-NDTIs)}\) a compelling subject for scientific investigation.

Owing to their propensity for \(\pi–\pi\) stacking interactions, they undergo self-assembly, forming an ordered two-dimensional (2D) lattice structure. This 2D molecular packing could enhance the \(\pi–\pi\) stacking further, which might amplify exciton delocalization and optical absorption characteristics. The presence of this 2D lattice structure enhances photon capture efficiency, energy transfer, and directional energy flow, showing their broad range of applications.

For computational studies, we considered all six isomers (A4, B4, A3B, AB3, A2B2, and ABAB) of \([4]\text{C-NDTIs)}\) units. The relatively small number of atoms and rigid structures allowed for modelling the initial structures with molecular point groups D4, C2, C2h, and S4.

(R)-SIA Gelator

Organogels are three-dimensional fibrous networks formed by small organic molecules when subjected to specific environments.\textsuperscript{41} The formation of these gels is largely influenced by non-covalent forces, such as hydrogen bonding and \(\pi–\pi\) interactions. At increased concentrations of organic molecules, these forces drive the formation of a self-assembled fibrous structure in organogels. Their formation is reversible when subjected to specific stimuli, such as heat and light, making them dynamic and adaptable.\textsuperscript{42} Organogels have a broad spectrum of potential applications, spanning from photoelectronics\textsuperscript{43} and drug delivery\textsuperscript{44}
The system studied in this respect is a fibre structure formed in nonpolar solvents by the low-molecular-weight organic molecule sulfonimidamide or (R)-SIA.

Molecular modelling of the (R)-SIA fibre network is a complex process that involves understanding all possible interactions formed by it. The base layer was constructed by assembling four molecules of (R)-SIA to form a hydrogen bonding network. The choice of hydrogen bond type within this layer was based on consideration of the functional groups in the (R)-SIA molecule and the associated test calculations: an intermolecular hydrogen bonding network was constructed between the H atom of NH₂ and the N atom of the pyridine group from two adjacent (R)-SIA molecules (as shown in Figure 1.5).

Next, a vertical stack of these layers was built such that the distance between the adjacent layers in the stack was 4 Å with a −24° twist. This vertical stack involved stacking 16 layers to form an M-type helical fibre structure, mirroring experimental data about the helical structure. This resulted in a vertical hydrogen bond network being formed between the top-layer hydrogen atom of the NH₂ group in (R)-SIA and the bottom-layer oxygen atom in (R)-SIA.

The dipole moment of the layer is non-zero in the Z-direction. Therefore, the dipole moment accumulates upon vertical addition of the layer in the direction of a helical axis (see Figure 1.5b). To lower the system dipole moment, another helical stack was constructed but oriented in the opposite direction (see Figure 1.6b). Similarly, multiple possible configurations of helical stacks were built and tested by performing MD simulations (see Figure 1.6).
Helicenes

Helicenes are polycyclic aromatic compounds in which benzene rings, or other aromatic rings, are angularly fused to form helical molecules. As the number of angularly fused rings increases, the structure becomes non-planar. The planes of consecutive rings are tilted to prevent steric clashes. This helical structure gives them with unique photophysical properties, including strong absorption in the UV-visible range. Their twisted configuration significantly influences their response to light, making these characteristics crucial for many modern electronic devices.\textsuperscript{46} Helicenes are key components in devices that interact with light and its polarization, such as certain types of detectors.\textsuperscript{47,48} However, synthesizing complex helicenes remains challenging, with ongoing ambiguity in understanding the structural features that influence their chiral responses.\textsuperscript{49–51} Computational studies of helicenes could illuminate the structure–property relationships of these compounds and aid in their targeted synthesis. Studied helicene systems include HO[7]OH, MeO[9]OMe, MeO[10]OH, and MeO[13]OMe. The initial model structures for these helicenes were derived from crystallographic data, with the HO[7]OH and MeO[9]OMe helicenes in the m-conformation and the MeO[10]OH and MeO[13]OMe helicenes in the p-conformation.

Outlook of thesis

The subsequent chapters present the foundational theory that drives our investigation of supramolecular systems and their spectroscopic properties. The second chapter provides a basic understanding of quantum mechanics by focusing on important concepts such as the Born–Oppenheimer approximation, the variational principle, the Hartree–Fock method, and density functional theory. Switching from quantum to classical mechanics, the third chapter introduces a key aspect of our research approach. It begins with an overview of force field (FF) principles, then progressively outlines the parameterization of original GAFF parameters for a molecule, followed by its validation. The limitations of FFs are also discussed. We then build upon this understanding in the fourth chapter by diving into the domain of molecular dynamics simulations. Here, we focus on configurational sampling, periodic boundary conditions, and interatomic interactions. This chapter further provides detailed insights into the practical aspects of setting up and executing these simulations. It is followed by techniques for analyzing MD simulations, such as the potential of mean force for assessing binding strength, and density plots to comprehend the spatial distribution of ligands around amyloid fibrils, as well as the limitations of MD methods. The fifth chapter shifts focus to the spec-
trosopic properties of molecules, where we first briefly discuss the framework of response theory. This is followed by the polarizable embedding approach to account for the solvent effect on spectra, and the technique of conformational averaging to incorporate the dynamics of the molecule in spectral features. The subsequent chapter provides a concise overview of pivotal findings from the articles. For a more in-depth understanding, all eight articles are included for comprehensive reading.
The universe is not only stranger than we imagine, it is stranger than we can imagine.
- J.B.S. Haldane

**Fundamentals**

The development of quantum mechanics (QM) in the early 20th century has undoubtedly transformed our perception of the physical universe. It was formulated when classical mechanics failed to explain the underlying physics behind various experiments such as the black body radiation, photoelectric effect, and atomic spectra. The QM theory is used to describe the physical properties at the scale of atomic and sub-atomic particles. The theory shows that a system’s energy levels are quantized at an atomic scale. This energy quantization challenges our classical understanding of energy as a continuum. It reveals the counterintuitive concept that physical properties at the atomic scale vary discretely rather than smoothly. Such peculiar behaviour of particles is described with the help of the QM postulates, which allow us to comprehend their physical properties in a probabilistic manner.

Quantum systems in stationary states (i.e., states where the probability density remains constant over time) are solved by employing the time-independent Schrödinger equation (TISE). The TISE is an eigenvalue equation that represents the wave function of a particle in terms of its energy eigenstates and it is mathematically given by:

$$\hat{H} \Psi = E \Psi$$  \hspace{1cm} (2.1)

$\hat{H}$ is the Hermitian Hamiltonian operator, which ensures that all the eigenvalues are real, and eigenvectors are orthogonal. It includes terms representing the kinetic energy of electrons and nuclei, and the potential energy due to electron–electron, nuclear–nuclear, and electron–nuclear type of interactions. The eigenvalue and eigenfunction of the Hamiltonian is the energy ($E$) and wave function ($\Psi$) of the state of the quantum system, respectively. In essence, $E$ and $\Psi$ connect quantum mathematical models to real-world observable properties. The energies obtained from the TISE equation are used to predict the outcomes of various experimental measurements, such as spectroscopy, scattering, and photoemission experiments. Theoretically, a wave function ($\Psi$) encodes all of the information about a system, including its position, momentum, and energy. The probability density of a particle is then given by the absolute square of its wave function, indicating the likelihood of finding the particle at a specific location.
In contrast, when the system is not in a stationary state then the probability density will evolve over time described by the time-dependent Schrödinger equation (TDSE) (More details on this will be discussed in chapter 5). The general solution to this TDSE is obtained by a linear combination of the eigenstates of the system with time-dependent coefficients. These coefficients gives estimate of the probability of finding the system in a particular eigenstate at a given point in time.

The TISE offers a comprehensive framework for understanding the quantum states of particles. However, it becomes increasingly difficult to solve this equation for systems with many interacting particles, such as molecules. For example, if we consider a water molecule, it contains 10 electrons and three nuclei, yielding a total of 39 spatial coordinates. If we were to solve the TISE for the water molecule, and even if we consider ten grid points per coordinate, leading to a staggering number of $10^{39}$ complex numbers required for representing wave function in sparse Hilbert space (A Hilbert space is an infinite-dimensional vector space used to describe the state of a quantum system). This makes the problem unmanageable, even for the most powerful computers. Hence, the approximations are introduced to simplify the equation and make it solvable.

**Born–Oppenheimer Approximation**

When a system consists of multiple electrons and nuclei, applying the Born–Oppenheimer approximation is an efficient way of solving the TISE given in previous section. The main idea behind the approximation is that motions of electrons in a molecule are faster than those of nuclei. Therefore, at a given time, the state of electrons is determined by the instantaneous positions of nuclei. In other words, from perspective of electrons, nuclei appear to be at rest, while from viewpoint of nuclei, electrons seem delocalized (illustrated in Figure 2.1 and 2.2). The Born–Oppenheimer approximation takes advantage of this separation of time scales by treating the electrons and nuclei as independent degrees of freedom. This allows the total wave function of the system to be factored into the product of the wave functions of the electronic and nuclear degrees of freedom:

$$|\Psi\rangle = \Psi_e|\Psi_n\rangle$$  \hspace{1cm} (2.2)

Here, $|\Psi_e\rangle$ is the electronic wave function, which depends upon electronic coordinates and parametrically on nuclear coordinates (see equation (2.6)). However, the nuclear wave function $|\Psi_n\rangle$ is a function of the nuclear coordinates only (see equation (2.7)). The total Hamiltonian operator can be written as a sum of the electronic and nuclear operators:

$$\hat{H} = \hat{H}_e + \hat{H}_n$$  \hspace{1cm} (2.3)
the electronic Hamiltonian operator, $\hat{H}_e$, is given by:

$$
\hat{H}_e = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{N} \nabla^2_{r_i} - \frac{e^2}{4\pi \epsilon_0} \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{|r_i - R_A|} + \frac{e^2}{4\pi \epsilon_0} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{1}{|r_i - r_j|}
$$

(2.4)

and the nuclear Hamiltonian operator, $\hat{H}_n$, is given by:

$$
\hat{H}_n = -\frac{\hbar^2}{2} \sum_{A=1}^{M} \frac{\nabla^2_{R_A}}{M_A} + \frac{e^2}{4\pi \epsilon_0} \sum_{A=1}^{M-1} \sum_{B=A+1}^{M} \frac{Z_A Z_B}{|R_A - R_B|}
$$

(2.5)

where $\hbar$ is the reduced Planck constant, $m_e$ is the mass of the electron, $M_A$ is the mass of the $A$th nucleus, $Z_A$ is the atomic number of the $A$th nucleus, $e$ is the elementary charge, and $\epsilon_0$ is the vacuum permittivity. $\nabla_{r_i}$ and $\nabla_{R_A}$ represent the gradient operators with respect to the electronic and nuclear coordinates, respectively. In $\hat{H}_e$ (equation (2.4)), the first term represents the kinetic energy of electrons, the second term denotes the electron–nucleus Coulomb interaction energy, and the third term represents the electron–electron Coulomb interaction energy. In $\hat{H}_n$ (equation (2.5)), the first term denotes the kinetic energy of nuclei, and the second term represents the nucleus–nucleus Coulomb interaction energy.

The Born–Oppenheimer approximation allows us to solve the Schrödinger equation for the electronic wave function $|\psi_e\rangle$ and the nuclear wave function $|\psi_n\rangle$ separately. We first solve the Schrödinger equation for the electronic wave function as:

$$
\hat{H}_e \psi_e(r; R) = E_e(R) \psi_e(r; R)
$$

(2.6)

Here, $r$ represents the coordinates of all the electrons in the system, while $R$ represents the set of fixed nuclear coordinates. The electronic wave function $\psi_e(r; R)$ is parametrically dependent on the nuclear coordinates. This implies that while the nuclear positions influence the electronic wave function, these positions remain fixed when solving the electronic Schrödinger equation. The term $E_e(R)$ denotes the electronic energy. This energy is subsequently used in solving for the nuclear wave function $|\psi_n\rangle$.

$$
\left(\hat{H}_n + E_e(R)\right) \psi_n(R) = E_{\text{tot}} \psi_n(R)
$$

(2.7)

Combining $\hat{H}_n$ and $E_e(R)$, we obtain a total effective Hamiltonian for the nuclear wave function. The term $E_{\text{tot}}$ represents the cumulative energy of the system, including contributions from both the electronic and nuclear components.
When equation (2.6) is solved for the electronic wave function at different nuclear configurations, it yields the electronic energy $E_e(R)$ for each configuration. The potential energy surface (PES) is then constructed by adding the nuclear–nuclear repulsion energy to these electronic energies. The PES, therefore, represents the total potential energy of the system as a function of the nuclear coordinates (illustrated in Figure 2.3). The nuclear configuration that corresponds to the minimum energy point on the PES is referred to as the equilibrium geometry, which represents the most stable arrangement of molecules in the system. The ground state equilibrium geometry significantly influences the chemical and physical properties of the molecule because it is the most statistically populated state. An understanding of PES holds the key to unraveling molecular behaviors and properties.

**Variational Principle**

Determining the exact wave function of a quantum system with multiple particles is often not feasible. This is because the complexity and computational cost associated with solving the Schrödinger equation increase with the number of particles. This complexity is often referred to as the “many-body problem” in quantum mechanics. For instance, even for a helium atom, which is the second simplest atom after hydrogen, an exact solution to the Schrödinger equation is not known, due to electron–electron correlation effects. The situation becomes even more complicated for larger atoms and molecules, which contain many electrons.

Because of these challenges, we often use methods to find approximate solutions. These solutions might not be exact, but they are usually sufficiently accurate for practical purposes. One such method is the variational principle. This method allows finding an approximate electronic wave function called trial wave functions. According to the variational principle, the trial energy represents a upper bound for the true energy of the system eigenvalue. The trial energy is defined as the expectation value of the Hamiltonian operator with respect to the trial wave function and given by the following equation:

$$E = \langle \tilde{\Psi} | \hat{H} | \tilde{\Psi} \rangle \geq E_0 \quad (2.8)$$

The electronic Hamiltonian operator, previously represented as $\hat{H}_e$, will henceforth be denoted as $\hat{H}$. $E$ is the energy associated with the trial wave function, while $E_0$ is the true energy eigenvalue. The trial wave function, denoted as $\tilde{\Psi}$, is expressed in terms of Slater determinants for $N$ electrons, that in turn can be described by the following expression:
This determinant is constructed as a sum of products of single-electron wave functions, also known as spin orbitals. Each product includes one orbital for each electron in the system and uses distinct orbitals. Moreover, the set of orbitals used in the Slater determinant are often chosen to be orthonormal. When constructing a Slater determinant for a molecule, each element of the determinant, which is a spin orbital \( \psi(r) \), can be expressed as the product of a molecular orbital \( \phi(r) \) and a spin function, denoted here as \( \sigma \). This spin function \( \sigma \) takes the value of either \( \alpha \) or \( \beta \), representing the two possible spin states.

\[
\psi(r) = \phi(r) \sigma
\]  
(2.10)

Next, these molecular orbitals \( \phi(r) \) can be written as linear combinations of atomic orbitals (LCAO) \( \chi_\alpha(r) \), where \( c_\alpha \) are the coefficients of these combinations.

\[
\phi(r) = \sum_\alpha c_\alpha \chi_\alpha(r)
\]  
(2.11)

Therefore, to find the approximate energy eigenvalue, we must minimize trial energy expression by finding the optimal coefficients \( c_\alpha \). This can be achieved by setting the derivatives of the trial energy with respect to these coefficients to zero, under the constraint that the orbitals remain orthonormal. Solving such equations yield the coefficients, which give us an approximate energy that is typically very close to the true energy of the system. The accuracy of the variational method largely depends on the selection of the trial wave function. Therefore, it should be chosen carefully.

Hartree–Fock Method

The Schrödinger equation becomes complex and nearly impossible to solve for systems consisting of multiple electrons in an atom or molecule due to the electron–electron repulsion term. The Hartree–Fock method provides a practical approach for solving this equation by approximating the ground state wave function and energy of a many-electron system.\(^{57-59} \) In this approach, the electronic wave function \( |\psi_e\rangle \) for the many-electron system is approximated using a single Slater determinant. While the Hartree–Fock method accounts for the average field experienced by electrons, it does not capture electron correlation effects, which arise from the instantaneous interactions between
electrons. These effects are treated explicitly in more advanced quantum chemistry methods, such as Møller–Plesset perturbation theory and coupled-cluster theory, which build upon the Hartree–Fock approximation.

The starting point is to consider the expectation value of the electronic Hamiltonian operator in equation (2.8). Since $|\psi_e\rangle$ is approximated as a Slater determinant, attention is turned to its constituent spin orbitals. Each spin orbital describes the state of a single electron in the system. To minimize the energy $E$ associated with the total wave function, the individual spin orbitals are varied. The conditions for this minimization lead to the Hartree–Fock equations for the spin orbitals:

$$\hat{f}\psi_a(\mathbf{r}) = \varepsilon_a \psi_a(\mathbf{r})$$

(2.12)

Where $\psi_a(\mathbf{r})$ is the spin orbital, this describes both the spatial location and the spin of an electron in the system. The $\varepsilon_a$ is the orbital energy associated with $\psi_a(\mathbf{r})$. $\hat{f}$ is the Fock operator that can be written as the sum of the one-electron operators $\hat{h}$, $\hat{J}$ and $\hat{K}$:

$$\hat{f} = \hat{h}(\mathbf{r}) + \sum_{i=1}^{N} \left[ \hat{J}_i(\mathbf{r}) - \hat{K}_i(\mathbf{r}) \right]$$

(2.13)

The operator $\hat{h}(\mathbf{r})$ includes the kinetic energy and nuclear attraction terms for an electron at position $\mathbf{r}$. The operators $\hat{J}_i(\mathbf{r})$ and $\hat{K}_i(\mathbf{r})$ represent the Coulomb and exchange operators, respectively, given by:

$$\hat{J}_i(\mathbf{r}_1)\psi_a(\mathbf{r}_1) = \frac{e^2}{4\pi\varepsilon_0} \int \frac{\psi_a^\dagger(\mathbf{r}_2)\psi_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d^3\mathbf{r}_2 \psi_a(\mathbf{r}_1)$$

(2.14)

$$\hat{K}_i(\mathbf{r}_1)\psi_a(\mathbf{r}_1) = \frac{e^2}{4\pi\varepsilon_0} \int \frac{\psi_a^\dagger(\mathbf{r}_2)\psi_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d^3\mathbf{r}_2 \psi_i(\mathbf{r}_1)$$

(2.15)

The Coulomb operator $\hat{J}_i(\mathbf{r}_1)$ quantifies the repulsion experienced by an electron at position $\mathbf{r}_1$ due to the electron density associated with another electron in spin orbital $\psi_i$. The integral defining this operator runs over all space, accounting for the electron density at every point $\mathbf{r}_2$, and computes the average Coulombic repulsion experienced by the electron at $\mathbf{r}_1$. On the other hand, the exchange operator $\hat{K}_i(\mathbf{r}_1)$ accounts for the quantum mechanical phenomenon that arises from the indistinguishability of electrons. While the Coulomb operator deals with the direct repulsion between electrons, the exchange operator encapsulates the effects of the possible exchange of two electrons occupying spin orbitals $\psi_a$ and $\psi_i$. This exchange reflects the
correlation effects due to the Pauli exclusion principle and the anti-symmetry requirement of the electronic wavefunction.

All spin orbitals are obtained by solving the eigenvalue equation (2.12). The challenge arises because the Fock operator, $\hat{f}$, itself depends on the spin orbitals, which creates a paradox of needing to know the solution beforehand. This is a common problem in quantum chemistry, typically addressed through a self-consistent field (SCF) method, which iteratively refines the spin orbitals. Initially, a guess for the spin orbitals is made, and a provisional Fock operator is constructed based on this guess. Equation (2.12) is then solved using this operator to produce a new set of spin orbitals. These new orbitals serve as the basis for constructing Fock operator in the next iteration, and the process is repeated. With each iteration, the spin orbitals and the Fock operator are updated, gradually converging to a consistent set of spin orbitals that satisfy both the Fock operator and the eigenvalue equation simultaneously.

For practical computations, the Roothaan–Hall approach is designed to numerically solve the HF equations. This method involves representing the wave function as a linear combination of predetermined spatial basis functions. These basis functions are pivotal for the efficiency and accuracy of electronic structure calculations. When focusing on atomic orbital (AO)-based calculations, typically two primary classes of basis sets are used. The first is the Slater-type orbitals (STOs), which closely mimic the solutions of Schrödinger equation solved for the hydrogen atom. Despite their accuracy, STOs are computationally expensive. To overcome this, the second class, Gaussian-type orbitals (GTOs), is employed. While GTOs are computationally efficient, they are not as accurate as STOs. Therefore, to enhance their accuracy, multiple GTOs are often combined to emulate the behavior of a single STO, giving rise to what are termed as contracted Gaussian functions (CGFs).

Among the plethora of basis sets, the Pople and Dunning style basis sets are popular. The Pople-style basis sets, like the 6-31G or 6-311G series, are characterized by their splitting of the valence orbitals into two groups: inner and outer valence electrons. Each group described by a different number of GTOs, thereby aiming to provide a more nuanced description of the valence electrons. On the other hand, Dunning basis sets, such as the cc-pVXZ series (where X indicates quality, e.g., D for double-zeta), are hierarchical in nature, the accuracy improves with every additional GTO layer. Moving from D to T to Q increases both size and accuracy but also requires more computational resources.

To achieve a more accurate representation of molecular systems, especially where electronic distributions can be polarized or extended far from the nuclei, specialized functions are added to basis sets. Polarization functions are introduced to provide additional accuracy.
ditional flexibility in representing electron distributions. These functions allow the molecular orbitals to better adapt to asymmetric electron distributions encountered in chemical environments, especially during bond formations. Their inclusion is essential for a more accurate description of molecular geometries, and various chemical interactions. On the other hand, diffuse functions are crucial for representing electron distributions in regions far away from the atomic nuclei. These GTOs have very small exponents, enabling the orbitals to stretch out and encompass electrons in regions of low electron density. These kinds of functions are crucial for capturing long-range interactions and providing an accurate description of weak interactions. Both the Pople and Dunning basis sets have versions that include these functions, making them adaptable for various computational needs.

Once, the molecular orbitals are represented by mathematical known basis functions, which are then solved iteratively, starting from an initial guess for the coefficients of the basis functions. The iterative process involves solving a series of matrix equations, which are computationally demanding but can be parallelized to reduce the overall computational time. The choice of basis set is critical in the electronic method calculations, as it determines the accuracy and efficiency of the calculation. Generally, a larger basis set will provide more accurate results, but at the cost of increased computational time. Therefore, selecting the appropriate basis set is a crucial step in performing accurate and efficient electronic structure calculations.

**Density Functional Theory**

In 1964, Hohenberg and Kohn proposed density functional theory (DFT) as an alternative method to solve for the electronic structure of molecules and materials. The first Hohenberg–Kohn theorem states that the ground state properties of a many-electron system interacting with an external potential (for example the potential from the nuclei) are uniquely determined by the ground state electron density. It essentially provides one-to-one correspondence between the ground state electron density and the external potential for non-degenerate ground states. This is an important concept because the wave function (which is a function of $3N$ variables for an $N$-electron system) can be considerably complex and computationally intensive. However, by using the electron density (which is a function of just three spatial variables) the computational process is greatly simplified, allowing for more efficient calculations. The second Hohenberg–Kohn theorem establishes that there exists a variational principle for the energy in terms of the electron density. The central idea from both the theorems is that the true ground state energy of a sys-
**DENSITY FUNCTIONAL THEORY**

**FIGURE 2.4:**
Top: Solving Schrödinger equation (SE) for quantum system properties. Bottom: The equivalent but more manageable approach using density functional theory (DFT) Kohn–Sham (KS) equations. Sketch inspired from Mattsson et al. 66

The system can be determined from the true electron density and could be written as follow:

\[
E[n(r)] = T[n] + V_{ee}[n] + \int n(r)v_{ext}(r)\,d^3r \tag{2.16}
\]

where \(T[n]\) represents the kinetic energy of the electrons, \(V_{ee}[n]\) is the electron–electron interaction energy, and \(v_{ext}(r)\) denotes the external potential. The major challenge with this equation arises from our lack of knowledge about the exact forms of the \(T[n]\) and \(V_{ee}[n]\) functionals.

One year later in 1965, Kohn and Sham proposed a way to bypass this problem and solve for ground state energy.67 Similar to the HF approach, they introduced a reference system which consisted of electrons that do not interact with each other and instead move under the influence of an external field. However, the density of this fictitious non-interacting system is kept the same as the true density of interacting electrons. By constructing the Hamiltonian for the reference system, the one-electron Kohn–Sham orbitals can be obtained. The energy of the system can be expressed as:

\[
E[n(r)] = T_{KS}[n] + V_{ee,KS}[n] + E_{xc}[n] + \int n(r)v_{ext}(r)\,d^3r \tag{2.17}
\]

The terms \(T_{KS}[n]\) and \(V_{ee,KS}[n]\) represent the kinetic energy of the non-interacting electrons and the classical Coulomb interaction energy, respectively, in the reference Kohn–Sham system. The \(E_{xc}[n]\) is the exchange–correlation energy. Comparing equations (2.16) and (2.17), we determine \(E_{xc}[n(r)]\), defined as the difference between kinetic and electron interaction energies of the true system and the reference non-interacting particle system.

\[
E_{xc}[n(r)] = (T[n] - T_{KS}[n]) + (V_{ee}[n] - V_{ee,KS}[n]) \tag{2.18}
\]

Due to the unknown nature of the exchange–correlation energy functional, approximations are needed to solve the equation, which comes in a hierarchy defined by Jacob’s ladder of DFT.68 It starts with the local density approximation (LDA), and progresses to more advanced approximations such as the generalized gradient approximation (GGA), meta-generalized gradient approximation (meta-GGA), and hybrid functional. Jacob’s ladder serves as a framework for exploring and comparing the performance and limitations of these different approximations, which help in the selecting the most appropriate approximation for a given system. However, it is crucial to note that ascending the ladder does not necessarily guarantee increased accuracy.

One of the most popular hybrid functionals is the B3LYP (Becke three-parameter, Lee-Yang-Parr) functional, due to its balance between computational cost and accuracy for a wide range
FIGURE 2.5: Jacob’s ladder of exchange–correlation functionals illustrates the progression towards more advanced approximations; however, it does not guarantee an increase in the accuracy of the results.

of molecules.\textsuperscript{69–71} It is a hybrid functional that mixes Hartree–Fock (HF) exchange with DFT exchange–correlation functionals and is primarily used in the calculation of ground state molecular properties. Being a global hybrid functional, B3LYP uses a fixed proportion (20\%) of HF exchange throughout all distances. Using a DFT exchange–correlation functional might be beneficial for short-range interactions, especially in capturing local electron correlation effects, such as those seen in covalent bonding where electrons are strongly correlated. However, this approach has its limitations when long-range interactions become important, as in the case of charge transfer. Hence, B3LYP fails most of the time in describing excitation states and charge-transfer effects.

To address these shortcomings, the Coulomb attenuating method B3LYP (CAM-B3LYP), a range-separated hybrid functional was developed by Yanai, Tew, and Handy.\textsuperscript{72} It also uses a modified form of the HF exchange similar to B3LYP, but a long-range correction to the exchange term is also included to improve the description of long-range effects. Moreover, unlike B3LYP, CAM-B3LYP adapts to the nature of interactions at different ranges. For short-range interactions, it employs about 19\% HF exchange, while for long-range interactions, it significantly increases the proportion of HF exchange to 65\%. As a consequence, CAM-B3LYP tends to predict excitation energies with higher accuracy and offers an improved description of charge-transfer effects compared to B3LYP.

It is important to note that no single functional can accurately predict all molecular properties, and the choice of functional should be carefully considered based on the specific system and property being studied.
FORCE FIELDS

The purpose of computing is insight, not numbers.
- Richard Hamming

This chapter is divided into four main sections, with each section addressing a different aspect of the force field. The first section provides a foundation by giving fundamental information about each term involved in the force field. The second section describes the steps involved in reparameterizing the general Amber force field (GAFF) of a molecule, an important process for optimizing the performance of original force field. The third section is about validating the parameterized force field, which is crucial for ensuring its accuracy and reliability. Finally, the last section discusses the limitations associated with force field method.

Principles

The force field method is used to study molecular properties and was developed in the 1950s and 60s. Initially, it started by describing intermolecular forces between atoms with simple mathematical models and later extended to larger molecules in the 1970s. Over the years, various types of force fields have been developed to capture the complexity of molecular interactions, such as Merck molecular force field (MMFF), chemistry at Harvard macromolecular mechanics (CHARMM), optimized potentials for liquid simulations (OPLS), and amber force field for amino acids. Among these, the amber force field has received significant attention for its reliability in simulating protein structures. GAFF was developed with a focus on simulating organic and drug like molecules. GAFF is compatible with the Amber force field when it comes to simulating protein–ligand type systems.

These various force fields are tailored for specific molecular systems to accurately model the potential energy landscape that governs molecular interactions. This landscape is often captured through force field equation and evaluated based on the nuclear coordinates of the system. The terms in the force field equation can be categorized into two main types: bonding and non-bonding interaction energies. The bonding interaction energy includes bond lengths, bond angles, and torsion angles. Whereas the non-bonded interaction component is calculated through Coulomb and van der Waals interaction terms. All the terms involved in force field equation are modelled using mathematical equations, and the parameters involved in these equations are either based on experimental data or ab-initio quantum mechanics calculations. One of the key advantages of MM methods is the ability to derive parameters for small basic units and
then use these derived parameters to construct a force field for larger molecular structures composed of these basic units. This approach is widely used in biological systems, where force field parameters for amino acids are derived and used to construct force fields for proteins. Because of such transferability property of force field parameters across different molecular systems, MM methods become important in providing valuable insights into the behaviour of molecules.

The total potential energy $E_{\text{total}}$ of a molecular system can be calculated as the sum of the potential energies of both bonding interactions ($E_{\text{bond}}, E_{\text{angle}}, E_{\text{dihedral}}$) and non-bonding interactions ($E_{\text{Coulomb}}, E_{\text{LJ}}$). It is represented by the following equation:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{Coulomb}} + E_{\text{LJ}} \quad (3.1)$$

Bond lengths are the distance between two atoms that are covalently bonded together and are typically fixed to a specific equilibrium value. The force constant for a bond is then used to describe the strength of the bond and the energy required to stretch or compress the bond length from its equilibrium value. It is given by the following equation:

$$E_{\text{bond}} = \frac{1}{2} k_{\text{b}} (r - r_{\text{eq}})^2 \quad (3.2)$$

The $E_{\text{bond}}$ is the potential energy associated with the bond, $k_{\text{b}}$ is the force constant controlling the strength of the bond, $r$ is the current bond length, and $r_{\text{eq}}$ is the equilibrium bond length. The term $(r - r_{\text{eq}})^2$ denotes the positive, quadratic increase of potential energy as the bond deviates from its equilibrium length. A factor $\frac{1}{2}$ is included to simplify subsequent derivation of forces on atoms, which are proportional to a negative gradient of potential energy.

Bond angles, on the other hand, are the angle between two bonds that are connected to the same atom and are also typically fixed to a specific value ($\theta_{\text{eq}}$). The force constant ($k_{\theta}$) of bond angle is then used to describe the strength of the angle and $E_{\text{angle}}$ is potential energy required to bend the angle away from its equilibrium value.
Torsion angles describe the rotation of one bond with respect to another bond. They are described using a periodic function which allows for multiple minima and maxima in the energy profile. The torsion angle potential energy ($E_{\text{dihedral}}$) is given by:

$$E_{\text{dihedral}} = k_d \left[ 1 + \cos(n\theta - \gamma) \right]$$

where $k_d$ is dihedral force constant, $\theta$ is phase angle, and the $n$ represent number of complete oscillations in the function. The amplitude of this cosine function determines the magnitude of the potential energy associated with a particular torsion angle, while the phase angle determines the position of the minimum energy point in the energy profile. A large $n$ allows modelling more complex energy profiles with multiple minima and maxima. In Figure 3.4, the dihedral potential for an ethane molecule is shown, with $n = 3$ indicating three minima across 360° range.

Non-bonded interactions are typically described using Lennard-Jones (LJ) and Coulombic potentials. The LJ potential describes the interaction between two atoms as a function of their separation distance. The potential has two parameters, the well depth ($\epsilon$) and the equilibrium distance ($\sigma$), which are determined through parameterization to reproduce experimental data. The $\sigma$ parameter represents the finite distance at which the inter-particle potential is zero. In other words, when $r = \sigma$, $E_{\text{LJ}} = 0$. The $\epsilon$ parameter sets the depth of the potential well. It represents the strength of the attraction when two particles come close together. When $r$ is such that $E_{\text{LJ}}$ is minimized, the potential energy is $\epsilon$. The LJ potential is used to describe van der Waals interactions between atoms, and given by the following equation:

$$E_{\text{LJ}} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right]$$

The first term in LJ potential, raised to the 12th power, represents a repulsive interactions between atoms. As the distance between two atoms decreases (i.e., $r$ becomes smaller), the magnitude of this term rapidly increases, and leading to an increase in the potential energy. This strong repulsion prevents the atoms from overlapping, especially dominant at very short distances. Whereas the second term in LJ potential, raised to the 6th power, represents a attractive interactions between atoms. As the atoms move further apart (i.e., $r$ becomes larger), the value of this term decreases, implying weakening of attractive interaction. This term is significant at intermediate distances but dominated by the repulsive component as atoms come very close.
On the other hand, Coulombic interactions, arise from the electrostatic forces between charged atoms. These interactions are typically described using a Coulombic potential, which depends on the charges of the atoms \((q_i\) and \(q_j)\), and their separation distance \((r_{ij})\). The Coulombic potential can be attractive or repulsive, depending on the charges involved. These interactions can have a significant impact on the behavior of molecules, particularly in the case of charged species like ligands and amino acids. A Coulombic potential energy \((E_{\text{Coulomb}})\) is given by the following equation:

\[
E_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \tag{3.6}
\]

One important parameter in the treatment of non-bonded interactions is the choice of cutoff distance. The cutoff distance is the distance beyond which non-bonded interactions are neglected in the calculation of the potential energy. The choice of cutoff distance can have a significant impact on the accuracy and efficiency of the force field. Increasing the cutoff distance can result in a more accurate representation of the interactions, but also increase the computational cost of the calculation. Conversely, reducing the cutoff distance decreases costs but can lead to significant errors in the interaction energy. To overcome this issue, MM method divide the system into a short-range and long-range interactions and handles them differently. More details on this topic is discussed in following chapter 4).

**Parameterization**

Throughout this thesis, GAFF database has been used to get initial force field parameters to model non-standard residues. While GAFF is designed to model the behavior of many organic molecules, it has limitations, especially in accurately predicting the dihedral potential around rotatable bonds. This aspect is crucial for understanding molecular conformations, and often poorly described by GAFF. Moreover, we are interested in conformational statistics and calculating the optical properties on snapshots obtained using original GAFF, which requires correct structural details. Therefore, reparameterization of GAFF is necessary to improve its accuracy and to ensure that the simulations provide reliable results about molecular conformations.

Reparameterizing involves adjusting the parameters of a force field to improve its agreement with reference experimental or theoretical data. In this section, we will demonstrate the reparameterization process using the HS-276 ligand as an example (see Figure 3.6). First, we will generate original GAFF parameters for the HS-276 and then adjust its dihedral parameters by fitting two dihedral potentials with reference DFT potential data. This can be easily done interactively with the help of the VeloxChem
PARAMETERIZATION

The first step is to perform a conformational analysis to identify the ligand conformer that corresponds to the minima on the PES. It involves evaluating relative energies of different possible structures that a molecule can take based on the rotations around its bonds by employing DFT/B3LYP functional. This analysis is important because molecular conformations can affect chemical properties and interactions with other molecules. Moreover, the FF parameters are constructed based on the most stable conformation, as these parameters are likely to be the most representative of the behavior of the molecule under normal conditions.

Next, we determine the RESP charges by calculating the electrostatic potential (ESP) at the HF/6-31G* level of theory for the conformer already identified as representing a minimum on the PES in the gas phase. The goal is to find a set of atomic charges that reproduces the calculated ESP as closely as possible. However, simply fitting atomic charges to the ESP can result in non-physical charges, particularly for buried atoms, which are less sensitive to variations in charges during ESP fitting. To address this issue, the RESP method introduces a hyperbolic penalty function. This function helps to avoid nonphysically high magnitudes of charges on buried atoms and ensures uniformity in the charges of equivalent atoms.

In the context of Amber force fields, it is important that all building blocks or residues, on which macromolecules are built have integer charges so that atomic charges can be transferred.

```python
# Calculate RESP charges
import veloxchem as vlx

molecule = vlx.Molecule.read_xyz("hs276.xyz")
basis = vlx.MolecularBasis.read(molecule, "6-31G*")

resp_drv = vlx.RespChargesDriver()
resp_drv.update_settings({"equal_charges":"33=34, 33=35"})

resp_charges = resp_drv.compute(molecule, basis, "resp")
```

The above code snippet calculates RESP charges for the HS-276. The optimized molecular geometry is read from file, and the 6-31G* basis set in combination with the HF method are applied. Using the vlx.RespChargesDriver() class, constraints are set to equalize the charges of atoms at indices 33, 34, and 35. Finally, the RESP charges are computed and stored for further use.

After calculating the RESP charges, the next step is to assign atom types based on GAFF convention. This is done by comparing the chemical environment of each atom in the molecule to the pre-defined GAFF atom types.
After assigning the atom types, the initial force field for the HS-276 ligand can be constructed using GAFF database. The constructed topology of HS-276 contains parameters for bonds, angles, dihedrals, and other non-bonded interactions that occur between atoms or molecules. The following two code snippets showcase how this process for HS-276 ligand is carried out.

```python
# Initialize force field generator
ff_gen = vlx.ForceFieldGenerator()
ff_gen.scan_dih_angles = []
ff_gen.scan_energies = []
ff_gen.scan_geometries = []
ff_gen.target_dihedrals = []
ff_gen.workdir = Path(".")
ff_gen.ffversion = 0
ff_gen.molecule_name = "hs276"
ff_gen.molecule = molecule
ff_gen.atom_types = atom_types
ff_gen.force_field_data = "gaff_2_1.dat"
ff_gen.force_field_data_extension = "gaff_2_1_extension.dat"

# Generate topology for HS-276 ligand
ff_gen.workdir.mkdir(parents=True, exist_ok=True)
original_itp_file = ff_gen.workdir / (ff_gen.molecule_name + f"_{ff_gen.ffversion:02d}.itp")
original_top_file = original_itp_file.with_suffix(".top")
ff_gen.write_original_itp(original_itp_file, list(ff_gen.atom_types), resp_chg)
ff_gen.write_top(original_top_file, original_itp_file)
```

The first code snippet initializes a ForceFieldGenerator object (ff_gen) by setting attributes like the working directory, force field version, and molecule name. It uses GAFF 2 database for extracting the required parameters. In cases where GAFF 2 database does not contain the necessary parameters, the ff_gen.force_field_data_extension attribute can be used to externally supplemen the missing data. The second snippet generates the topology of the ligand, outputting the topology (original_top_file) and include topology (original_itp_file) files. The output files are populated with bonds, angles, dihedrals, RESP charges, and LJ parameters.

After constructing the original force field, it is imperative to rigorously test the parameters, particularly those governing the dihedral potential of rotatable bonds. This verification step involves a side-by-side comparison with QM dihedral potentials.
Specifically, a relaxed dihedral scan for each of the relevant dihedral angles is performed using DFT/B3LYP functional and 6-31+G(d,p) basis set, and original GAFF force field of HS-276. The QM scan data could be supplied to "ff_gen" object in xyz format files such as 1_14_15_16.xyz. These files contain not only the atom indices for the dihedrals (like 1, 14, 15, and 16), but also the corresponding QM energies and geometries. The code snippet for reading QM data along with performing the MM scans is shown below, and the results obtained from it are presented in the Figure 3.7.

```python
# reading QM scan data
ff_gen.read_qm_scan_xyz_files(['1_14_15_16.xyz', '27_22_21_16.xyz'])

# dihedral scanning utilizing original GAFF force field
res_phi1 = ff_gen.validate_force_field(original_top_file, 0)
res_phi2 = ff_gen.validate_force_field(original_top_file, 1)

# visualization of dihedral potentials
ff_gen.visualize(res_phi1)
ff_gen.visualize(res_phi2)
```

Comparing the potential energy curves obtained from DFT and MM in Figure 3.7, it is clear that the initial topology generated by GAFF is not adequate for describing the SCCS ($\phi_1$) potential accurately. Due to the artificially high barrier represented in GAFF, the interconversion between the cis and trans conformations of HS-276 would never happen in the MD simulation. Similarly, the other important dihedral angle SCCN ($\phi_2$) must be accurately represented to sample conformation at 300 K.

To align the results of the MM force field more closely with those obtained using DFT, original GAFF parameters must be refined. This refinement is helpful for improving the accuracy of the simulation results and for calculating photophysical properties from frames sampled from a simulated trajectory. For this, an accurate description of bonds, angles, and dihedrals is required. The detailed procedure of doing this is as follows:

**Adjusting Bond and Angle Parameters**
The equilibrium bond distances and bond angles in the initial topology of HS-276 are replaced by those obtained from ground state DFT optimized structure, while keeping same force constant values.

**Parameterizing Dihedrals**
One of the central challenges in force field refinement is to accurately parameterize the dihedral angles. This is a multi-step process:

- **Removing Target Dihedral Angle Energy Contribution:**
  To fit the dihedral of interest accurately, the contribution of
torsion energy can be excluded from the total MM potential energy. This is achieved by setting the appropriate torsion parameters, corresponding to the fitting dihedral, to zero in the force field file. With this modified topology, a relaxed dihedral scan is performed in MM, which is referred as MM_0.

- **Computing Missing Energy:**
  Subsequently, the energy profile obtained from the MM_0 scan ($V_{MM_0}$) is subtracted from the DFT energy profile ($V_{DFT}$). This yields the missing energy contribution, which will then be used to update the torsion parameters of the dihedral angle of interest.

$$V_T = V_{DFT} - V_{MM_0}$$  \hspace{1cm} (3.7)

- **Fitting to DFT Potential:**
  The differences between DFT and MM_0 potentials ($V_T$) is then fitted to the Ryckaert–Bellemans (RB) function:

$$V(\phi) = \sum_{n=0}^{5} c_n \cos^n (\phi - 180^\circ)$$  \hspace{1cm} (3.8)

where $\phi$ represents the dihedral angle, and $c_n$ is the force constant for the $n$th cosine term. The fitting is carried out using a least-squares algorithm.

- **Adding and validating fitted parameters:**
  After obtaining the fitted parameters for the dihedral potential using the RB function, they are added to the force field file. Then, a relaxed dihedral scan is performed using the modified topology with the updated dihedral parameters. The resulting energy profile is compared with the DFT potential energy curve to check the accuracy of the fitted parameters. If the new energy profile agrees well with the DFT potential, the force field parameters are considered to be accurate and can be used for further simulations. If not, the fitting process for parameterizing the dihedral is repeated.

For parametrizing two dihedrals in HS-276 ligand, all the above steps can be executed in VeloxChem using the following code snippet.

```python
target_top_file = original_top_file
# fitting the two dihedrals to DFT data
for i in range(len(ff_gen.target_dihedrals)):
    target_top_file = ff_gen.dihedral_correction(target_top_file, i)
# visualization of fitted dihedrals
```
FIGURE 3.8: A graphical comparison of the potential energy profiles for the two dihedral angles before and after reparameterization, along with a comparison of potentials calculated from DFT and MM methods.

Initially, the target_top_file is set to original_top_file, which contains the current force field parameters. A for loop iterates through each targeted dihedral, invoking the dihedral_correction method to fit the MM potential to the QM data. Finally, the validate_force_field method is called to visualize the quality of the fitting for each dihedral, with the results stored in res_phi1_fit and res_phi2_fit. The plots obtained from it are shown in the Figure 3.8. The fitted curves accurately represent not only the minima at the correct torsion angles but also the correct energy barriers.

Validation

Once the force field parameters have been refined and incorporated into the force field file, validating the accuracy of the new fitted parameters is essential. This validation can be achieved by comparing the relative energies of conformers obtained us-
ing different methods such as DFT and the reparameterized force field.

**TABLE 3.1:** Comparison of the relative energies of HS-276 conformations optimized using the fitted force field versus the original GAFF, in reference to DFT values. All energies are given in kcal/mol.

<table>
<thead>
<tr>
<th>Conformer</th>
<th>$E_{\text{DFT}}$</th>
<th>$E_{\text{fit}}$</th>
<th>Error</th>
<th>$E_{\text{GAFF}}$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>TC</td>
<td>0.26</td>
<td>0.29</td>
<td>0.03</td>
<td>5.36</td>
<td>4.49</td>
</tr>
<tr>
<td>CT</td>
<td>0.87</td>
<td>1.19</td>
<td>0.32</td>
<td>-0.41</td>
<td>0.67</td>
</tr>
<tr>
<td>CC</td>
<td>1.34</td>
<td>1.97</td>
<td>0.63</td>
<td>5.55</td>
<td>4.21</td>
</tr>
</tbody>
</table>

Table 3.1 presents a comparison of the relative energies of various HS-276 conformers, optimized using fitted force field, GAFF, and DFT methods. The symbol "T" indicates that the absolute value of the dihedral angle lies between 90 and 180 degrees, while "C" is assigned for values between 0 and 90 degrees. The two-letter codes correspond to dihedrals $\phi_1$ and $\phi_2$. The fitted force field demonstrates a significant increase in accuracy compared to original GAFF force field. This is particularly evident from the relative energy errors of HS-276 conformers, detailed in Table 3.1, where GAFF exhibits significantly higher discrepancies. The error of the fitted force field are notably less than 1 kcal/mol relative to DFT calculations. For example, the error for the TC conformer is a mere 0.03 kcal/mol with the fitted force field, in contrast to a much higher 4.49 kcal/mol under GAFF.

**TABLE 3.2:** Comparison of transition energies, given in electronvolts (eV), for state $S_1$ calculated for geometries optimized using fitted force field, GAFF, and DFT methods.

<table>
<thead>
<tr>
<th>Conformer</th>
<th>$E_{\text{DFT}}$</th>
<th>$E_{\text{fit}}$</th>
<th>Error</th>
<th>$E_{\text{GAFF}}$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>3.51</td>
<td>3.49</td>
<td>0.02</td>
<td>3.28</td>
<td>0.23</td>
</tr>
<tr>
<td>TC</td>
<td>3.55</td>
<td>3.49</td>
<td>0.06</td>
<td>3.29</td>
<td>0.26</td>
</tr>
<tr>
<td>CT</td>
<td>3.57</td>
<td>3.60</td>
<td>0.03</td>
<td>3.31</td>
<td>0.26</td>
</tr>
<tr>
<td>CC</td>
<td>3.66</td>
<td>3.69</td>
<td>0.03</td>
<td>3.31</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Additionally, the comparison of transition energies for state $S_1$ provides a validation on predicting electronic properties. Single point absorption calculations were performed using CAM-B3LYP/aug-cc-pVDZ on the same optimized conformer geometries as mentioned above. As shown in Table 3.2, the transition energies calculated on geometries optimized with the fitted force field, and GAFF exhibit varying levels of accuracy with reference to the DFT method. Notably, the fitted force field shows again remarkable alignment with the DFT results, as evidenced by the minimal errors in transition energies across all conformers. In
contrast, the original GAFF force field exhibits larger discrepancies, with errors ranging up to 0.35 eV, as seen in the CC conformation. Such significant deviations highlight the limitations of the original GAFF in accurately predicting electronic properties. These differences can be attributed to the intrinsic limitations of GAFF on describing the molecular structure, which may not account for the subtle electronic effects influencing transition energies. To sum up, the improved accuracy of the fitted force field in predicting both geometrical and electronic properties underscores the importance of meticulous parameter refinement.

**Limitations**

Though the force field reparameterization process mentioned in the previous section can improve the accuracy of molecular simulations, there are limitations in general with force field methods. The accuracy of the force field is limited by the quality of the reference data used for parameterization. Since, the reference data depends on the level of theory used for geometry optimization and dihedral scan calculations, the errors within the level of theory will be propagated. Furthermore, force field is molecule-specific and may not apply to different molecule. Their transferability to other systems can be limited, and they may need further parameterization for different moiety within a molecule. Also, it is important to note that force fields may inaccurately predict molecular behavior under extreme conditions (such as high pressure or temperature).
MOLECULAR DYNAMICS SIMULATIONS

Everything that living things do can be understood in terms of the jiggling and wiggling of atoms.
- Richard Feynman

This chapter provides a concise overview of the fundamental principles of MD simulations. We begin by discussing essential topics, such as configuration interaction, periodic boundary conditions, interatomic interactions, as well as temperature and pressure control with thermostats and barostats, respectively. Subsequently, we will walk-through the practical steps needed to effectively set up and run MD simulations. Finally, we will explore specific methods for analyzing the simulated trajectories. We will utilize the spatial distribution function to identify and examine binding sites, and employ umbrella sampling to estimate the binding free energy.

Configurational Sampling

MD simulations aim to predict the behavior of molecular systems over time based on their given initial configuration and the governing physical laws. One fundamental aspect of performing simulations is exploring the configuration space, which encompasses all possible configurations of a system. This exploration is commonly referred to as configurational sampling and is crucial for understanding various thermodynamic and kinetic properties of the system.

At the microscopic level, countless atomic arrangements can occur within a molecular system. Each of these distinct arrangements corresponds to a point in the configuration space. The density with which these points are sampled can greatly affect the accuracy and reliability of simulation results. Sparse or uneven sampling can lead to skewed observations, which might potentially result in misinterpretations of molecular behavior. Conversely, thorough sampling ensures that the ensemble of configurations observed in the simulation accurately represents the true behavior of the system under the specified conditions.

Trajectories obtained during MD simulations serve as a detailed chronological record of atomic configurations as the system evolves over time. These trajectories reveal how the system explores its configuration space, the transitions it makes between different states, and the time it spends in various regions of the configuration space. For systems that exhibit multiple thermally accessible minima, the significance of efficient sampling of the configuration space becomes crucial. Each of these minima signifies a distinct stable or metastable state of the system, each with a unique set of properties.
A comprehensive perspective of such system under consideration could then be obtained by calculating the average properties using the following equation:

$$
\langle P \rangle = P_1 \rho_1 + P_2 \rho_2 + \cdots + P_n \rho_n = \int P(x) \rho(x) dx \quad (4.1)
$$

Mainly, this equation represents an aggregation of contributions from various minima on the potential energy surface (PES), each corresponding to a unique system configuration. For each such configuration ($C_i$), the specific property value $P_i$ is combined with its corresponding probability $\rho_i$. This combination provides a weighted contribution to the calculation of the average property value $\langle P \rangle$. This ensures that even infrequently visited states which might have distinctive properties are factored into the overall behavior of the system. As we move to more complex systems where the number of minima becomes either vast or continuous, the discrete summation converts to an integral representation, encompassing all potential configurations.

MD simulations achieve this sampling by solving Newton’s equations of motion numerically, in this thesis, by employing the leapfrog algorithm. This algorithm is engineered to solve equation of motion for the atoms in system while effectively preserving energy conservation. The starting point of MD simulations requires the initial coordinates and velocities for the molecular system under investigation. The choice of initial configuration depends on the specific system being studied and the research question being asked. For systems like amyloid protein fibrils, the initial coordinates can be obtained from experimental techniques like cryo-electron microscopy (cryo-EM). Cryo-EM provides high-resolution structures of macromolecules, so it can be used either directly or through the generation of models based on it as a starting point for MD simulations. For other systems like low molecular weight gels, where experimental data may be limited or non-existent, computational models can be used to find initial starting coordinates. Once the initial coordinates are determined, the velocities of the atoms are usually assigned based on the Maxwell–Boltzmann distribution at a desired simulation temperature. It describes the distribution of velocities of atoms in a gas at a given temperature and is commonly used to sample velocities in MD simulations.

The next step is to select an appropriate ensemble for the MD simulation. Ensembles such as NVE (microcanonical ensemble), NPT (isothermal–isobaric ensemble), and NVT (canonical ensemble) play a crucial role in defining the boundary conditions and thermodynamic properties of the system. The choice of ensemble depends on the specific system being studied and the research question being asked.
In this thesis, we have chosen the NVT ensemble for the production runs of the MD simulations after fixing the density during the equilibration step. The equilibration step involved performing short NVT and NPT simulations, which allowed us to reach the desired density for the system under investigation. When the density converges to a particular value while doing short NPT, it is considered appropriate to switch to the NVT ensemble for the production run. This choice reduces computational costs while maintaining the relevant thermodynamic properties of the system.

**Periodic Boundary Conditions**

MD simulations typically involve thousands to millions of atoms. In such simulations, handling the atoms at the edges of the simulation box is crucial. If not managed properly, these edge atoms can lead to unrealistic representations of a bulk system. To address this and simulate bulk systems more effectively, periodic boundary conditions (PBCs) are employed. PBCs creates the effect of an infinite system by replicating the simulation box endlessly in all directions. This approach ensures a more accurate representation of bulk system behaviour. In this setup, although conceptually an infinite system is created, the actual recorded frame coordinates are confined to the original simulation box. When an atom crosses the boundary of the simulation box, it re-enters from the opposite side, maintaining the consistency of the simulation. This strategy ensures the continuity of the simulation, avoiding any abrupt changes in the positions or velocities of the atoms.

**Interatomic Interactions**

The main bottleneck in MD simulation is the evaluation of interatomic interactions. To reduce computational costs while still accounting for the essential interatomic interactions, a finite radial cut-off distance is used to truncate these non-bonded interactions. This is done under the minimum image convention, ensuring that each atom interacts only with the nearest surrounding atom from the actual simulation box or its periodic image. It is also important that the radial cut-off distance is less than half of the simulation box length ($L/2$). Otherwise, the atom might start interacting with itself and leading to artifacts. The inter-atomic interactions in MD simulations can be broadly classified into two categories: short-range interactions and long-range interactions.

The short-range interactions become negligible beyond a certain distance between atoms. The Lennard-Jones (LJ) interaction is an example of a short-range interaction that rapidly decays with increasing distance between atoms. To reduce the computational cost, it is common practice to truncate the LJ potential...
by applying a cutoff distance beyond which the interaction is assumed to be zero. However, this approach can lead to discontinuities in the potential energy function, which can affect the accuracy of the simulation results (as shown in Figure 4.2). To overcome this issue, a common approach is to apply a potential-switching function before the cutoff distance. As the distance between atoms approaches the cutoff value, this function ensures a smooth transition to zero by gradually reducing the strength of the LJ interaction.

On the other hand, long-range Coulomb interactions decay slowly with distance and their calculation becomes computationally demanding. To address this issue, methods like the particle mesh Ewald (PME) summation are often employed in MD simulations. The PME method divides the Coulomb interactions into two components: a short-range real space component and a long-range reciprocal space component. The real space component is calculated using a cutoff distance (similar to short-range interactions), beyond which the interactions are assumed to be negligible. The reciprocal space component is calculated using the fast Fourier transform (FFT) algorithm, which enables the rapid calculation of long-range electrostatic interactions. The PME method with the FFT algorithm is widely used in MD simulations and ensures that accurate results are obtained even in systems with complex charge distributions.
Thermostat and Barostat

To accurately model the behavior of molecules and materials under different conditions, it is crucial to maintain constant temperature and pressure within the simulation. This is achieved using thermostat and barostat algorithms, which ensure that the system remains in equilibrium during the simulation.

A thermostat adjusts the kinetic energy of atoms in an MD simulation to maintain a constant temperature. To achieve this goal, various thermostat algorithms have been developed. One of them is the velocity-rescale thermostat. It efficiently controls the temperature of the system while maintaining the correct distribution of velocities. It does this by first calculating a scaling factor which is based on the difference between the current and desired temperatures of the system. It is then multiplied by the particle velocities for rescaling. To improve the stability and convergence of the thermostat, a stochastic factor (a random element) is also added to the scaling factor calculation. This ensures that the kinetic energy of the system remains close to the desired value, and thereby maintaining a constant temperature.

However, because of this stochastic factor, one should be cautious when attempting to reproduce ensemble properties. As obtaining identical results from independent simulations can be challenging, even after starting from the same initial conditions. In such cases, it may be necessary to use other thermostats which allow for better reproducibility of ensemble properties.

Similar to thermostats, barostats maintain constant pressure in MD simulations by adjusting the volume of the simulation box. One of the most widely used barostat algorithms is Berendsen. This barostat works by first calculating scaling factor based on the difference between the current and desired pressures, as well as the compressibility of the system. The simulation box vectors and coordinates are then adjusted by multiplying them with the scaling factor, resulting in a change in volume. This process ensures that the pressure remains constant throughout the simulation.

Simulation Setup and Protocols

This section discusses the simulation setup and protocols involved in performing MD simulations of systems mentioned in this thesis. The simulation process is crucial in obtaining reliable and physically relevant results that provide insights into the molecular interactions governing protein–ligand binding or stability of supramolecular complexes. The simulation setup and protocols comprise several stages: system preparation, minimization, equilibration, and production run. Each stage has its unique purpose and methodology to ensure the simulation stability and enhance the overall quality of the simulation. We will
explore the details of these stages and outline best practices for performing MD simulations.

SYSTEM PREPARATION

In the system preparation stage, our primary goal is to set up an accurate and stable representation of the supramolecular systems within the simulation environment. The initial step involves either constructing or adopting the structure of crucial components such as protein–ligand complexes, helicenes, or supramolecular structures from low molecular weight gels, which will be discussed in detail in the subsequent chapter molecular modeling. After establishing the initial structure, we proceed to construct a simulation box around it. It is essential to ensure that the distance between any atom and its periodic image is significantly greater than the cutoffs used for calculating interaction energies to avoid any artifacts in the simulation. The next step involves solvating the remaining space within the simulation box using the desired solvent. In cases where the solvent is not water, we first create a small solvent box with the correct density, which is then used to solvate the system. Lastly, counterions are added to neutralize the system, as we will be employing the PME method for electrostatic interactions in our MD simulations.

MINIMIZATION

Energy minimization is an essential step in the molecular dynamics simulation process, as it helps optimize the initial structure of the system and remove any potential artifacts or instabilities. The primary objectives of energy minimization are the removal of steric clashes, which may arise from close atomic contacts, and the relaxation of the system to ensure that it begins from a stable conformation. To achieve these goals, various minimization algorithms can be employed, such as the steepest descent method and the conjugate gradient method. The minimization process continues until convergence criteria are met, in which the maximum force on any atom should be less than $10 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. For large protein–ligand systems, a value of $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ may also be accepted. The maximum number of steps for the minimization is usually set to 50,000. The finally obtained energy minimized structure ensures a stable and optimized starting point for subsequent simulation stages.

EQUILIBRATION

Equilibration is an essential step in MD simulations, serving to stabilize the system by allowing the solvent and ions to relax around the supramolecular complex. This process ensures that the system reaches an optimal state for subsequent simulation steps, which often involve analyzing interactions or investigating...
the dynamics of the system. To achieve this, position restraints are applied to the supramolecular complex during equilibration, preventing any large-scale movement and allowing the solvent and ions to settle around it. Equilibration can be performed using two primary methods: NVT and NPT equilibration. NVT equilibration maintains a constant number of atoms, volume, and temperature in the system. This approach employs temperature coupling methods, such as the velocity-rescale, to stabilize the temperature. NPT equilibration, on the other hand, maintains a constant number of atoms, pressure, and temperature. In this method, pressure coupling techniques, such as the Berendsen or Parrinello–Rahman barostats, are utilized to stabilize both pressure and temperature.

PRODUCTION RUN

The production run is a crucial phase in molecular dynamics simulations, with its primary objective being the collection of data for analysis and the exploration of the conformational space of the system under study. To set up the production run, it is essential to select appropriate simulation parameters and define the output frequency for trajectories, energy, and other properties, while also considering the duration and timestep of the simulation. Restarts and checkpointing play a significant role in saving intermediate states for recovery and continuation, ensuring consistency in long simulations. Monitoring the simulation progress involves assessing energy conservation, observing protein–ligand interactions and stability, and checking for any anomalies or issues during the simulation. Upon completion, post-processing and trajectory analysis include trajectory concatenation and alignment, removal of periodic boundary condition artifacts, and preparation of data for subsequent analysis and visualization.

Analysis

In this section, we explore two distinct approaches for analyzing protein–ligand interactions. The first approach utilizes the spatial distribution function (SDF), a tool primarily aimed towards qualitatively identifying ligand binding sites on fibrils. It is instrumental in revealing key interaction regions and specific residues critical for binding. The second approach provides a quantitative perspective by employing umbrella sampling to calculate the binding energy within these sites. This method elucidates the free energy landscape, which is crucial for identifying energetically favourable binding modes within protein–ligand complexes. Integrating insights from both methods enables a more holistic approach to understanding molecular interactions.
Protein–ligand interactions in MD simulations are effectively analyzed using the spatial distribution function in the VIAMD program, with the results visualized as density plots. These plots are result of the spatial and temporal aggregation of ligand distribution around a stable reference structure, such as an amyloid protofilament. This aggregation is achieved by superimposing the protofilament chains of the amyloid fibril and populating the 3D histogram bins with each selected ligand instance, through temporal super-positioning across all frames of the MD simulation.

The density plot highlights information about the binding sites and helps in identifying the strongest binding site, the one with the maximum density for a particular ligand. Additionally, when examining multiple ligands, these plots also reveal comparative characteristics of bound ligands at their respective binding sites. For example, one can determine whether a ligand is highly or sparsely populated at the binding site by comparing the spatial distribution of the ligands, as indicated by the density. Moreover, concentrated and diffuse distributions indicate the degree of protein–ligand interaction specificity, with more concentrated distributions suggesting a higher specificity, while diffuse distributions may imply less specificity in the interaction.

For instance, in Paper II, density plots corresponding to four ligands (pFTAA, qFTAA-CN, HS-276, and bTVBT4) with reference to an Aβ(1–42) protofilament are presented in figure 4.3. These plots only consider ligands that bind as monomers to the fibril. As observed in the figure, pFTAA, qFTAA-CN, and HS-276 exhibit high density at their respective binding sites, indicating that these sites are densely populated. In contrast, the lower density of bTVBT4 suggests sparse population at its binding sites. The calculated binding energies (in kcal/mol) for each ligand–fibril interaction are also presented in the figure.
binding site. This observation aligns with experimental findings where pFTAA, qFTAA-CN, and HS-276 exhibit fluorescence, while bTVBT4 does not. This example demonstrates the utility of the spatial distribution of ligands in identifying binding sites and in the qualitative assessment of ligand interactions.

ESTIMATION OF BINDING ENERGY

The multiple binding modes of the protein–ligand complex obtained from simulation can be quantitatively studied to identify the most energetically favorable configuration by calculating the free energy of the system. The umbrella sampling method is employed to sample configurations along a reaction coordinate, followed by PMF calculations to gain insights and uncover the underlying free energy landscape. The PMF profile displays the energetics of molecular processes, such as binding or conformational transitions of a protein–ligand system. Computation of this profile and error associated with it is calculated using the weighted histogram analysis method (WHAM).91

In the specific cases of ligands binding to amyloid protein fibrils, the generation of PMF profiles involved several steps. Starting with the identified strong binding site on a fibril for a ligand, as discussed in the preceding section, we segmented this specific portion from our larger, half-pitch fibril model. This resulted in a reduced fibril–ligand system, which consists of 20 protofilament chains from either Aβ(1–42) or tau fibrils, including the ligand. A long MD simulation on this system was then conducted to ascertain the most effective binding mode. Subsequently, the binding energy for this mode was calculated by first conducting a steered molecular dynamics (SMD) simulation, where a ligand is pulled out from the binding site. In this process, we first choose a reaction coordinate (typically, this is the distance between the center of mass of the ligand and the center of mass of selected amino acid residues at the binding site of the protein fibril) along which the force will be applied to pull a ligand out of the binding site. This external force is modeled as a spring with a certain stiffness and is applied gradually with a constant pulling rate to avoid a non-physical behavior. The pulling rate should be carefully selected, as a lower pulling rate will increase the computational cost, whereas a higher pulling rate might not be able to sample configurations around a barrier effectively. The end of this simulation results in the ligand freely moving in a water environment.

From this pulling trajectory, the reaction coordinate is divided into a number of umbrella windows. These windows should be spaced at a small enough distance to ensure continuous sampling of the system along the reaction coordinate. The subsequent step involves running several independent simulations, each starting from a chosen set of umbrella windows. These simulations must
be run for a sufficient amount of time to ensure proper sampling within each umbrella window. The harmonic bias potential is applied to each umbrella window, which ensures that the reaction coordinate remains near the center of each window.

![Free Energy Profile](image)

**FIGURE 4.4:** The averaged PMF profiles for four ligands in their stable conformers, bound to respective sites on the Aβ(1–42) fibril.

Finally, the WHAM tool from GROMACS is employed to calculate the PMF profile along the reaction coordinate and histograms showing the distribution of configurations within each umbrella window. This is achieved by combining the results from the independent umbrella window simulations. It is important to ensure that each histogram overlap with the adjacent histogram to produce a meaningful profile. In case of non-overlap, a few umbrella window simulations must be added in that region, followed by re-estimating the PMF profile. The resultant free energy profile depicts the ligand unbinding process and also provides an estimation of the binding energy. Figure 4.4 serves as an illustrative example of a PMF profile obtained by applying the methodology described above. It shows the PMF profiles generated by pulling four distinct ligands from their respective binding sites in Aβ(1–42) fibril.

Despite its advantages, the PMF method has several limitations. One of most important is to properly select the reaction coordinate, which if not correctly identified, can distort results. This step is a critical because free energy landscapes are typically high-dimensional, and reduction to a single or few reaction coordinates may lose crucial information. Another drawback is the computation cost, as one needs to choose many umbrella windows and perform umbrella sampling on multiple trajectories to ensure the reliability of calculated PMF profile. Finally, it is also important to note that the PMF profile depends on the quality
of the force field used in the simulations, with inaccurate force fields leading to incorrect free energy estimates.

Limitations

Though MD simulation offers valuable insights into the supramolecular systems at the atomic level, there are certain limitations associated with it. One of the crucial limitation is the time scale of simulations. The current computational capabilities allow for up to millisecond timescale simulations, however, many biological processes such as protein folding, or conformational changes occur over much longer timescales. Simulating for shorter time can lead to incomplete sampling of the conformational space and may affect the results obtained from simulation. Another limitation emerges from the initial conditions of the simulation. The starting structure of a supramolecular system is critical in determining the results of the simulation. Any inaccuracies present in the initial structure can propagate throughout the simulation, potentially yielding erroneous results.
SPECTROSCOPIC PROPERTIES OF MOLECULES

Spectroscopy is the study of the interaction between light and matter, enabling us to peer into the invisible molecular world and observe how atoms and molecules interact. With advancements in theory and computational methods to calculate spectroscopic properties, our understanding of molecular behaviors and their interactions with light has increased. Light in this context spans over various regions of wavelengths, each with its own interaction characteristics (see Figure 5.1). In this thesis, we concentrate on the absorption and circular dichroism (CD) spectroscopic techniques in the visible range of the electromagnetic spectrum. In absorption spectroscopy, photons from a light source interact with molecules, and the resultant amount of absorbed photons is plotted as a function of wavelength. It gives valuable information about the electronic and vibrational energy levels of molecules, which provide insight into their structure and the nature of chemical bonds. On the other hand, CD spectroscopy on a chiral molecule measures the difference in absorption of left and right circularly polarized light as a function of wavelength. It provides information about molecular conformational changes and their chiroptical properties. This can be useful for designing novel materials with tailored properties for specific applications, such as chiral catalysts, sensors, or molecular switches.

This chapter provides a brief overview of theoretical and computational approaches to calculate and interpret the spectroscopic properties of complex molecular systems. First, we will discuss the fundamentals of molecular response theory, which will provide a framework for calculating molecular properties. Next, the polarizable embedding (PE) approach is presented, which allows for accurately capturing the effects of complex environments like amyloid protein fibrils. Finally, we discuss conformational averaging, a method to include the dynamic behavior of molecules.

Response Theory

Response theory in the context of this thesis is used to calculate the spectroscopic properties of a system when it is subjected to an external electromagnetic field. In response to this external stimulus, a molecule undergoes time-varying changes in its electronic and molecular structure. In the first chapter, we focused on the mathematical framework to study a system using a time-independent approach, as the Hamiltonian used to describe the energy of the system was time-independent. However, now upon the inclusion of a time-dependent external potential
into the Hamiltonian, the system will be governed by the time-dependent Schrödinger equation.

\[ i\hbar \frac{\partial}{\partial t} \Psi(r, t) = \hat{H} \Psi(r, t) \]  

(5.1)

In this equation, \( \hbar \) is the reduced Planck constant, and \( \Psi(r, t) \) is the wave function of the perturbed system. To further analyze the response of the system, we expand the wave function \( \Psi(r, t) \) in orders of perturbation as follows:

\[ \Psi(r, t) = \Psi^{(0)} + \Psi^{(1)} + \Psi^{(2)} + \ldots \]  

(5.2)

The \( n \)th-order term of the wave function can be written as:

\[ \Psi^{(n)} = \sum_j c_j^{(n)}(t)e^{-iE_j t/\hbar} |j\rangle \]  

(5.3)

Where \( |j\rangle \) denotes the unperturbed states with \( E_j \) being their corresponding energies and \( c_j^{(n)}(t) \) are the time-dependent coefficients. The total Hamiltonian operator, represented by \( \hat{H} \) in Equation (5.1), is composed of the unperturbed Hamiltonian, \( \hat{H}_0 \), and the time-dependent perturbation potential, \( \hat{V}(t) \), which arises from the external field.

\[ \hat{H} = \hat{H}_0 + \hat{V}(t) \]  

(5.4)

The wavelength of the applied field is typically in the order of a few hundred nanometers, while the size of molecules is just a few nanometers. Given the substantially smaller molecular size compared to the wavelength of the applied field, it is reasonable to assume that the molecule experiences a uniform field. The dipole approximation is thus applied in this scenario, wherein the external potential, \( \hat{V}(t) \), is represented as:

\[ \hat{V}(t) = -\hat{\mu}_\alpha F_\alpha(t) \]  

(5.5)

In this equation, there is an implied summation over the index \( \alpha \), which signifies a Cartesian coordinate \( (x, y, \text{ and } z) \) within the molecular frame. The term \( \hat{\mu}_\alpha \) denotes the component of the dipole moment operator associated with the Cartesian coordinate, while \( F_\alpha(t) \) represents the respective component of the external electric field at time \( t \). Moreover, the component \( F_\alpha(t) \) can be expressed using a Fourier series, which is written as:

\[ F_\alpha(t) = \sum_\omega F_\alpha(\omega)e^{-i\omega t} \]  

(5.6)

This Fourier series decomposition of \( F_\alpha(t) \) allows for a detailed analysis of how different frequency components \( \omega \) of the external field influence the molecular system over time.
To analyze the response of the system to the perturbation, we focus on the first-order correction. This correction will also be used later in this chapter and is given by:

\[ c_m^{(1)}(t) = \frac{i}{\hbar} \sum_{\omega} \int_0^t dt' F_\alpha(\omega)e^{i(\omega_m - \omega)t'} \langle m|\hat{\mu}_\alpha|0 \rangle \] (5.7)

Here, \( c_m^{(1)}(t) \) represents the first-order correction to the coefficient corresponding to the state \(|m\rangle\) at time \( t \). The summation over \( \omega \) represents the range of frequencies of the external electric field. \( F_\alpha(\omega) \) is the Fourier component of the external electric field at frequency \( \omega \) along the \( \alpha \) Cartesian coordinate. \( \omega_{m0} \) is the transition frequency between the state \(|m\rangle\) and the ground state \(|0\rangle\), indicating the energy difference required for this transition. For a detailed description and derivation of the equation, refer to the book "Principles and Practices of Molecular Properties".92

Having established the expression for the first-order correction to the wave function under an external electric field, we next use it to examine induced molecular dipole moments in response to external field. The equation below provides the expression of the induced dipole moment (\( \hat{\mu}(t) \)) in a molecule upon external perturbation (such as an electric field). It is expressed as a power series in terms of the field strength \( (F(\omega)) \), and utilizes Einstein summation notation for its tensor form representation:

\[ \mu(\alpha)(t) = \mu_0(\alpha) + \sum_{\omega} \alpha_{\alpha,\beta}(\omega)F_\beta(\omega)e^{-i\omega t} \]

\[ + \frac{1}{2} \sum_{\omega_1,\omega_2} \beta_{\alpha,\beta,\gamma}(\omega_1,\omega_2)F_\beta(\omega_1)F_\gamma(\omega_2)e^{-i(\omega_1 + \omega_2)t} \]

\[ + \ldots \] (5.8)

The first term in this series, \( \mu_0(\alpha) \), represents the permanent dipole moment of a molecule before interacting with the applied field. The coefficients \( \alpha_{\alpha,\beta}(\omega) \) and \( \beta_{\alpha,\beta,\gamma}(\omega_1,\omega_2) \), representing the linear- and hyperpolarizability tensors, respectively, quantify the response of the molecule to the applied field.

Now the expectation value of induced dipole moment (\( \hat{\mu} \)) operator along the \( \alpha \) coordinate becomes

\[ \langle \Psi(r, t)|\hat{\mu}_\alpha|\Psi(r, t) \rangle = \langle \hat{\mu}_\alpha \rangle^{(0)}(t) + \langle \hat{\mu}_\alpha \rangle^{(1)}(t) + \langle \hat{\mu}_\alpha \rangle^{(2)}(t) + \ldots \] (5.9)

Each term in the series expansion, \( \langle \hat{\mu}_\alpha \rangle^{(n)}(t) \), is also a sum of terms, where each term involves the dipole moment operator \( \hat{\mu} \) acting on two different states \(|\Psi^{(m)}\rangle\) and \(|\Psi^{(n-m)}\rangle\) as shown below.
We specifically look at the first-order expectation value of the dipole moment, \( \langle \hat{\mu}_\alpha \rangle^{(1)}(t) \). It is given by the sum of two terms. One term involves the dipole moment operator acting on the unperturbed state and the first-order perturbed state, while the other term involves the operator acting in the opposite order.

\[
\langle \hat{\mu}_\alpha \rangle^{(1)}(t) = \langle \Psi(0) | \hat{\mu}_\alpha | \Psi^{(1)} \rangle + \langle \Psi^{(1)} | \hat{\mu}_\alpha | \Psi(0) \rangle
\]  
(5.11)

Using equations (5.7) and (5.3), along with the initial condition that the system is in the ground state, \( |0\rangle \), before applying the field, we obtain the final expression.

\[
\langle \hat{\mu}_\alpha \rangle^{(1)}(t) = \sum_\omega \frac{1}{\hbar} \sum_{m \neq 0} \left[ \frac{\langle 0 | \hat{\mu}_\alpha | m \rangle \langle m | \hat{\mu}_\beta | 0 \rangle}{\omega_{m0} - \omega} + \frac{\langle 0 | \hat{\mu}_\beta | m \rangle \langle m | \hat{\mu}_\alpha | 0 \rangle}{\omega_{m0} + \omega} \right] F_\beta(\omega) e^{-i\omega t}
\]  
(5.12)

Comparing this equation with the equation (5.8), we get the expression for linear polarizability

\[
\alpha_{\alpha\beta}(\omega) = \frac{1}{\hbar} \sum_{m \neq 0} \left[ \frac{\langle 0 | \hat{\mu}_\alpha | m \rangle \langle m | \hat{\mu}_\beta | 0 \rangle}{\omega_{m0} - \omega} + \frac{\langle 0 | \hat{\mu}_\beta | m \rangle \langle m | \hat{\mu}_\alpha | 0 \rangle}{\omega_{m0} + \omega} \right]
\]  
(5.13)

In the above equation, when state \( |m\rangle \) is equal to \( |0\rangle \), both terms inside the square brackets cancel each other for all frequencies, and therefore it is not included in the summation. When the transition frequency (\( \omega_{m0} \)) becomes equal to the frequency of the applied field (\( \omega \)), we get a pole. The residue of this pole is used to calculate the probability of this transition, also known as the oscillatory strength, and is given by the following expression:

\[
f_{f \leftarrow 0} = \frac{2m_e}{3\epsilon^2} \Delta \omega_{f0} \sum_{x,y,z} |\langle 0 | \hat{\mu}_\alpha | f \rangle|^2
\]  
(5.14)

Here, \( \Delta \omega_{f0} \) is the excitation energy for the transition from the ground state \( |0\rangle \) to the excited state \( |f\rangle \), and \( m_e \) and \( e \) represent the electron mass and elementary charge, respectively.

While response theory provides a theoretical framework for calculating spectroscopic properties, its practical applications can be constrained by computational costs. For large systems, the need to determine numerous excited states and account for their couplings to accurately compute a spectrum makes this approach practically unfeasible due to these costs. In the next section, we will examine a method for practically calculating the spectroscopic properties of large systems.
Polarizable Embedding

Building on the foundation laid in the previous section, this section introduces a method for calculating spectroscopic properties of a molecule in a surrounding environment. Pure quantum mechanical methods become infeasible for such large systems due to their computational demands. An alternative method, known as the polarizable embedding (PE) approach, offers a viable and efficient way to account for the environmental effects on the absorption spectra of a molecule. This method is particularly used in current studies of the following systems: the bTVBT4 ligand in water and tau protein environments (Paper I), and helicenes in non-polar solvents (Paper VIII).

In the PE model, the system is partitioned into two distinct regions: the core region, which contains a ligand for which spectra are calculated, and the environment. The core region is treated using a quantum mechanical method, while the environment is represented by multipole moments representing the permanent charge distribution and dipole–dipole polarizabilities, located at the coordinates of environmental atoms. The polarizability tensor facilitates the capture of mutual polarization occurring between the environment and the core region. The environmental influence is incorporated when calculating photophysical properties by introducing a polarizable embedding potential into the Hamiltonian of the core region while solving the time-dependent Schrödinger equation.

For the case of bTVBT4 in an aqueous solvent, water molecules which are within 20 Å from the ligand were considered, and test calculations were performed to calculate the absorption spectra using the PE model. Among the various test calculations, it was found that a hybrid model in which the environmental region was further divided into polarizable and non-polarizable regions, converged more smoothly compared to a fully polarizable model. The polarizable region is represented by Ahlström charges and isotropic polarizabilities. In contrast, the non-polarizable (NP) region is defined solely by TIP3P charges, where polarization is implicitly accounted for. Figure 5.3 illustrates a sample calculation of the transition wavelength as a function of shell thickness. It compares the polarizable region (depicted in red), non-polarizable region (in blue), and a hybrid of both regions (in green). Test calculations suggest that a shell thickness of 15 Å for the polarizable region and 5 Å for the non-polarizable region are sufficient to account for the effect of water on bTVBT4. The differences between the two hybrid models, which use different water models (Ahlström and TIP3P) for the non-polarizable region, are not very pronounced for the selected shell thicknesses. Similar results were observed in a study conducted on another anionic ligand, pFTAA, from the class of LCO ligands in water.
FIGURE 5.3: The transition wavelength of bTVBT4 as a function of shell thickness is presented for the polarizable (red), non-polarizable (blue), and polarizable with full NP (green) regions. These results correspond to trans snapshot, which was randomly extracted from the simulation trajectory. The bold lines denote when the non-polarizable region is represented by Ahlström charges, and the dashed lines indicate when TIP3P charges are used. The absorption spectra calculations were performed using the CAM-B3LYP/aug-cc-pVDZ method.

For the ligand in a protein environment, long protein chains exist which are continuous and large, which makes deriving charges and polarizability using quantum mechanical methods computationally expensive. Hence, the protein environment is split into smaller fragments, and then charges and polarizabilities are determined using the molecular fragment conjugated caps (MFCC) method on each fragment separately. In the MFCC method, the peptide bonds are cut and capped with appropriate groups. The task of splitting the protein into fragments is performed with the PyFraMe module. Next, charges and polarizabilities are calculated using the LoProp approach as implemented in the Dalton or VeloxChem program. When using VeloxChem, this process can be easily executed by providing the molecular coordinates in the file "structure.xyz", as demonstrated in the following Python code snippet:

```python
import veloxchem as vlx

molecule = vlx.Molecule.read_xyz("structure.xyz")
basis = vlx.MolecularBasis.read(molecule, "6-31G")

# SCF optimization
scf_drv = vlx.ScfRestrictedDriver()
```
scf_results = scf_drv.compute(molecule, basis)

# LoProp
loprop_drv = vlx.LoPropDriver()
loprop_out = loprop_drv.compute(molecule, basis, scf_results)

loprop_charges = loprop_out["localized_charges"]
loprop_polarizabilities = loprop_out["localized_polarizabilities"]

In the final step, the potential file is created, containing charges and polarizabilities of only the protein atoms while removing the charges of the caps appropriately.

FIGURE 5.4: Illustration of the MFCC amino acid fragmentation scheme

The application of this technique can be illustrated by a specific study from “Paper I” concerning the absorption spectra of bTVBT4 in tau protein. For spectra calculations on snapshots extracted from MD, only a polarizable region of 20 Å around the chromophore was considered. The rationale for this decision was based on few test calculations and benchmark studies conducted by Beerepoot et al. These studies highlights the influence of polarization effects from protein environment on the excitation energies of a chromophore. To ensure convergence of excitation energies with respect to size of environment shell, they suggested that it is necessary to include at least 20 Å of the protein environment in the calculations.

Moreover, typically, amyloid fibrils remain highly stable during the simulation timeframe when a ligand is bound. Previous studies conducted by Gustafsson et al. found that the charges and isotropic polarizability derived from the LoProp approach are stable across snapshots extracted from MD simulations. We adopted the same methodology, but in addition, we analyzed the maximum standard deviation of charges and isotropic polarizability for all atoms in the amino acid residues ILE 360, THR 361, and HIS 362 (the binding site for bTVBT4), which are present in 10 chains of tau protein surrounding bTVBT4 within the snapshot. The maximum standard deviations of charges and isotropic polarizability were 0.01 and 0.10 a.u., respectively. Given the stability of charges and polarizability tensors in 10 chains, we decided to use the average charge and average isotropic polarizability values for all the chains, both within and across all snapshots.

FIGURE 5.5: The bTVTB4 ligand in a tau protein environment, highlighting a 20 Å polarizable region around the ligand. This region is described by point charges and dipole–dipole polarizability and was considered in the calculation of absorption spectra.
Conformational Averaging

In the last section, we looked at how we can use the PE approach to include the environmental effect on a ligand while calculating spectra. When we want to accurately capture the dynamical nature of a molecule, the conformational averaging technique becomes indispensable.

From MD simulations, a series of snapshots representing the dynamical behavior of the ligand is extracted. Each snapshot essentially captures a specific configuration at a particular moment in time. Next, spectra are computed for each of these snapshots. The fundamental concept here is that each individual snapshot can contribute to the overall absorption spectra of the system. Thus, by calculating the spectrum for each snapshot and summing these individual spectra, a more comprehensive and averaged spectrum is obtained.

While the process of conformational averaging is computationally expensive, it is a preferred strategy as it takes into account the inherent dynamical behavior of the molecule, a feature that is absent in a spectrum based on a single snapshot representing an average attribute distribution of the ligand. This is achieved by line broadening of the electronic transitions in the spectrum, which is a more realistic representation of spectral features as it accounts for the intrinsic variability in the molecular configurations. Conformationally averaged spectra also serve as a powerful tool for comparing theoretical spectra with experimental spectra, providing an essential step in validating the results of our MD calculations, like the bTVBT4 binding site in tau fibrils.

The number of snapshots required for a reliable averaged spectrum is determined by gradually increasing the numbers of extracted snapshots and including the contribution of all the spectra to the final spectrum. As the spectrum profile starts to converge and does not undergo significant changes with the addition of more snapshots, we know that the final spectrum is representative of the ligand dynamics. This analysis allows us to achieve a balance between computational cost and accuracy.

The technique can be illustrated considering the simple MD simulation of bTVBT4 in vacuum. In order to include the dynamical behavior, we need to consider its conformational changes over simulation time. The analysis shows that the dihedral SCCS (defined in the supplementary information of Paper I) is predominantly in the \textit{trans} conformation (70\%) and less frequently in the \textit{cis} conformation (30\%). Next, we calculate the conformationally averaged spectra considering different numbers of snapshots (100, 200, and 300) extracted from the MD simulation. These snapshots maintained the proportion of the \textit{trans} and \textit{cis} conformations at 70\% and 30\% respectively. By comparing the spectra generated from 100, 200, and 300 snapshots, we are able to deter-
FIGURE 5.6: Convergence analysis of conformationally-averaged absorption spectra for bTVBT4 Ligand in Vacuum: A comparison of calculated spectra using 100, 200, and 300 snapshot sets extracted from MD simulation. Each set preserves a 70% trans and 30% cis conformational distribution as determined by the dihedral angle between the two thiophene rings. Calculations were performed using the CAM-B3LYP/aug-cc-pVDZ

mine the number of snapshots necessary for reliably representing the dynamics of the ligand. As shown in the corresponding figure, the spectral profile starts to converge and no longer undergoes significant changes after 200 snapshots. Therefore, it is concluded that 200 snapshots are sufficient for a reliable averaged spectrum of the bTVBT4 ligand.
OVERVIEW OF ARTICLE FINDINGS

Paper I

In this article, we focused on examining the binding of the bi-thiophene-vinylene-benzothiazole (bTVBT4) ligand with tau fibrils associated with Alzheimer’s disease (AD) progression. The primary objective was to pinpoint specific binding sites and binding modes, and understand the interactions between the ligand and tau fibrils. This was achieved using a combined approach of theoretical methods and experimental spectroscopy.

We begin with creating a periodic model of the tau fibril based on its cryo-EM structure. Following this, we performed unbiased MD simulations to explore the interactions between the fibril and bTVBT4. These simulated trajectories led to the identification of multiple binding sites and modes, which were thoroughly analyzed. Density plot showing ligand distribution around the reference tau protofilament revealed that the key binding site for bTVBT4 in the tau fibril is a hydrophobic pocket, comprised of residues ILE360, THR361, and HIS362. The estimated binding energy for the strongest binding mode within this site is 33 kJ/mol. The congruence of the theoretical and experimental redshifts observed in the absorption spectra of bTVBT4, comparing its transition wavelength in water to that in tau protein environments, validates the findings.

Furthermore, the proposed binding site was corroborated with observations drawn from fluorescence imaging of brain tissue samples of patients suffering from AD and PiD stained with bTVBT4. These findings indicate that bTVBT4 binds exclusively to the tau from AD and not to the tau fold found in PiD. Interestingly, the specific proposed binding site is not accessible in the tau fold associated with PiD, elucidating the reason why bTVBT4 shows the specificity towards the tau fold from AD.

These findings enhance our understanding of the molecular interactions associated with tau physiopathology and suggest possible way for controlling disease-specific ligand binding via chemical design.

Paper II

This paper aimed to conduct a comparative analysis of the binding properties of various ligands (pFTAA, qFTAA-CN, HS-276, and bTVBT4) to amyloid fibrils (Aβ(1–42) and tau). Atomistic molecular dynamics (MD) simulations were performed for two types of systems: one involving all ligands interacting with Aβ(1–42) fibrils, and the other with tau fibrils. The objective was to identify binding sites, compute free energy profiles, and compare the binding properties of different ligands to amyloid fibrils.
The simulation results, analyzed using a specific spatial distribution function and visualized via density plots, revealed the binding sites of ligands on Aβ(1–42) and tau fibrils. The binding of anionic ligands pFTAA and qFTAA-CN, with charges of −4 and −3 respectively, to both fibrils is driven by Coulomb interactions. In contrast, the binding of neutral HS-276 and cationic bTVBT4 with a +1 charge, is driven by LJ interactions. In Aβ(1–42) fibrils, both pFTAA and qFTAA-CN are found to bind to a common site. Conversely, HS-276 and bTVBT4 have different binding sites. Similarly, in tau fibrils, pFTAA and qFTAA-CN share a common binding site. However, HS-276 and bTVBT4 bind to another common site, which is distinct from the one where anionic ligands bind. For Aβ(1–42) fibrils, pFTAA and qFTAA-CN were found to bind to a common site, on the other hand, HS-276 and bTVBT4 were found to have different binding sites. Similarly, in tau fibrils, pFTAA and qFTAA-CN share a common binding site, however HS-276 and bTVBT4 bind to a common site, which is distinct from the site where anionic ligands bind.

The binding energies obtained from PMF show that pFTAA, qFTAA-CN, and HS-276 bind more strongly to Aβ(1–42) compared to bTVBT4. While in the case of tau fibril, pFTAA, qFTAA-CN, and bTVBT4 bind more strongly than HS-276. However, the difference between the binding energies of HS-276 and bTVBT4 to tau falls within the error range of the method; thus, no concrete conclusions can be drawn. All estimated energies are supported by experimental observations from fluorescence spectroscopy, which show an absence of emission when the weakest ligand binds to fibrils in either case. In addition to their binding characteristics, the ligands exhibit interesting structural properties. They undergo conformational changes and display varying dihedral angle distributions within the fibril environment compared to a water solvent. The extent of these changes in dihedral angle distributions may offer insights into the factors contributing to the observed spectral shift.

In summary, our findings indicate that the affinity of ligands for fibrils can be modulated by adjusting their charge. The insights from this study could serve as a foundation for designing ligands in targeted fibril detection involved in neurodegenerative diseases like Alzheimer’s.

**Paper III**

Similar to Paper II, this paper also focuses on the development of ligands specifically designed to selectively bind to amyloid fibrils associated with AD. This study was conducted in collaboration with experimentalists. On the experimental side, our collaborators synthesized a series of ligands based on the thiophene-vinylbenzothiazolium (TVBT) scaffold, combined with different hete-
rocyclic moieties. The selected ligands (HS-332, HS-334, and HS-335) were then stained to differentiate between aggregated Aβ(1–42) and tau pathologies in human Alzheimer’s disease brain tissue. The unique optical properties of the ligands resulted in bright blue and red fluorescence in Aβ(1–42) and tau aggregates, respectively, allowing for precise pathology differentiation.

On the computational side, we performed MD simulations to investigate the interactions between the three ligands (HS-332, HS-334, and HS-335) and amyloid protein fibrils, as well as their binding modes. The HS-332 ligand, which features a TVBT scaffold, exhibited a specific binding mode to tau fibrils. Moreover, due to the similar structure and charge of bTVBT4 and HS-332, their binding sites were found to be identical. HS-334 and HS-335, ligands based on the TV scaffold, demonstrated various binding modes with Aβ(1–42) fibrils, interacting with residues GLY25, SER26, PHE20, VAL18, and LYS16.

The goal of these computational studies was to provide microscopic details of the selected ligands that demonstrated unique optical properties. Overall, this research contributes to our understanding of ligand–protein interactions and assists in the development of ligands for the selective detection of protein aggregates in neurodegenerative diseases. This has implications in pathological studies investigating disease-specific protein aggregates.

**Paper IV**

The focus of this paper is on understanding the binding modes of a small molecule ligand (Py1SA) to transthyretin (TTR), a protein associated with amyloid diseases. This was achieved through a combination of experimental and computational methods such as spectroscopy, X-ray crystallography, and MD simulations.

The results show that Py1SA binds specifically to TTR. Spectroscopic measurements reveal a red shift in excitation and a blue shift in emission upon ligand binding. Structural data for the TTR–Py1SA complex were derived from X-ray crystallography experiments. However, the interpretation of the binding mode between TTR and Py1SA was not clear due to ambiguous results in the X-ray crystallography experiments, particularly in the characterization of positioning of the Py1SA. To overcome this limitation, MD simulations were employed to determine the preferred orientation of Py1SA in the TTR binding pocket.

The results from simulations show that Py1SA predominantly adopts a reverse binding mode, i.e., when the salicylic acid group points into the binding pocket, and the pyrene moiety is oriented outward. This binding mode is energetically stabilized by hydrogen bonding interactions between Py1SA and key residues within the TTR binding site. Subsequently, the binding free energy was estimated from the PMF calculations, confirming the
a greater affinity for the reverse binding mode over the forward mode.

The findings of this study enhance our understanding of small molecule recognition and binding specificity, especially in the context of TTR amyloid diseases. The ability of the fluorescent Py1SA ligand to bind to TTR is crucial for developing diagnostic tools to detect amyloidosis diseases. Moreover, the integration of experimental and computational techniques proves effective in determining the structures of TTR-ligand complexes. This combined approach addresses and overcomes challenges related to partial ligand occupancy in crystallographic studies.

**Paper V**

This research paper investigates the effect of varying L- or D-tyrosine substitutions on the thiophene backbone of oligothiophenes, and how these changes influence their photophysical properties and self-assembly behavior. The optical characteristics and structural organization of these oligothiophenes were examined using spectroscopic, microscopic, and computational methods.

The results showed that the position of the enantiomeric side-chain significantly influenced the photophysical properties and the architecture of the self-assembled supramolecular structures. Under acidic conditions, the oligothiophenes formed optically active self-assembled aggregates exhibiting $\pi-\pi$ interactions. The absorption, emission, and circular dichroism (CD) spectra revealed distinct optical behavior depending on the positioning of L- or D-tyrosine moieties in the thiophene backbone. The emission spectra exhibited red-shifts and decreased intensity in such acidic condition, suggesting aggregation of the oligothiophene molecules. Both fluorescence anisotropy decay and dynamic light scattering experiments confirmed that aggregation and larger structures form under acidic conditions than in alkaline conditions.

Computational analysis, using MD simulations and quantum-mechanical spectrum simulations, supported the experimental findings. The conformational variations of p-HTA-D-Tyr were analyzed, revealing that MM (Minus–Minus) conformers, characterized by consecutive negative dihedral angles between the thiophene groups, are more probable than PP (Plus–Plus) conformers, which have consecutive positive dihedral angles. This offers insights into induced circular dichroism (ICD) of single molecule p-HTA-D-Tyr. Next, computational spectral analysis further differentiated between various conformers, identifying the strongest signal responses in the enantiomeric classes XMMX and XPPX. Moreover, additional MD simulations were conducted to study the effects of intermolecular interactions in aggregated chromophore systems. To reduce computational
costs, subsequent spectroscopic calculations were performed using the semi-empirical ZINDO approach. This revealed an increase in ICD intensity and a small red-shift in aggregated structures compared to single molecules. The computational results align well with experimental data, demonstrating that ICD signals are mainly influenced by MM conformers and are amplified by intermolecular interactions.

The findings of the study contribute to our understanding of how subtle changes in the position of enantiomeric side-chain functionalities can greatly affect the optical properties and self-assembly behavior of oligothiophenes. Moreover, computational results, corroborated by experiments, showed that MM conformers primarily influence ICD signals, which are amplified in aggregated states. These insights have implications for the design and development of chiral oligothiophenes with tailored optical properties and could find applications in optoelectronics and biosensing.

**Paper VI**

This article presents the synthesis and analysis of [4]cyclonaphthodithiophene diimides ([4]C-NDTIs). This ideal pillar-shaped molecule (nanopillar) was synthesized by concatenating NDTI units in the alpha-positions of thiophene rings through an electrochemical oxidation process. The study highlights the appealing electronic and optical properties of [4]C-NDTIs. These properties enable nanopillar to exhibit near-infrared (NIR) light absorption, strong excitonic coupling, and tight encapsulation of C60.

The results revealed that the inherent chirality emerges from the stable orientation of NDTI units in nanopillars, allowing comprehensive CD studies on isomeric structures. In [4]C-NDTIs, the strong $\pi-\pi$ stacking capabilities were retained, which allowed the formation of unique two-dimensional (2D) lattice arrays on a molecular level. This 2D configuration, along with the ability to encapsulate C60, holds promise for mimicking natural photosynthetic systems and propelling the development of multifunctional organic materials.

The computational studies provided additional insights into the [4]C-NDTI isomers. Theoretical CD spectra were computed and compared to the experimental results, facilitating the identification of isomers based on their CD measurements. Moreover, the analysis of Gibbs free energies indicated that the A4/B4 enantiomeric isomers were the most stable among the various isomers.

Overall, the article sheds light on the synthesis, structures, and properties of [4]C-NDTIs, emphasizing their potential in the development of multifunctional organic materials. It also demonstrates the utility of electrochemical reduction as a syn-
thetic route for constructing \(\pi\)-conjugated macrocycles and the significance of 2D stacking in supramolecular chemistry.

**Paper VII**

This article introduces a novel class of chiral sulfonimidamides that can form low-molecular-weight gels in nonpolar solvents through supramolecular polymerization. This gelation process results from the self-assembly of fibrous structures arranged in a helical fashion. The gelation behavior is explored using scanning electron microscopy (SEM), circular dichroism (CD) spectroscopy, and computational modeling techniques.

The results show that sulfonimidamides form gels at low concentrations in nonpolar solvents, which is a reversible process dependent on heating and cooling. SEM analysis reveals a fibrous network structure in the gels, exhibiting distinct characteristics based on the solvent and a helical arrangement of the fibers. CD spectroscopy provides insights into the supramolecular structure of the gels with strong signals indicating a helical arrangement.

To correlate the supramolecular structure with the observed CD spectrum of the gel, we built a model structure and calculated the theoretical CD spectra based on a two-layer unit composed of eight (R)-SIA. The resulting CD spectrum was found to be in agreement with the experimental spectra, enabling us to get insights from the model structure. The model demonstrates that a stable supramolecular structure is formed due to hydrogen bonding within and between the monomer units.

Overall, this work introduces a novel scaffold of chiral sulfonimidamides as supramolecular organogelators. The study provides insights into their gelation behavior and their fibrous supramolecular structure arrangement. These findings contribute to the development of functionalized materials for various applications in areas such as photoelectronics, drug delivery, and tissue engineering.

**Paper VIII**

Helicenes are important due to their optical properties, which find applications in molecular electronics and chiral optoelectronic devices. However, synthesizing new helicenes and understanding their structural aspects remains challenging. This research paper focuses on the computational evaluation of the optical properties of four azaoxahelicenes. The density functional theory method is employed to investigate the UV-vis absorption, electronic circular dichroism (ECD), emission, and circularly polarized luminescence (CPL) spectra of these helicenes. The spectroscopic properties are studied both in a vacuum and in a solution environment using a combined molecular dynamics and polarizable embedding framework.
The calculated computational spectra are in good agreement with experimental data. Redshifts in the absorption bands are observed with increasing helicene length, whereas the ECD spectra exhibit similar intensities across the studied systems. The absorption bands were found to be influenced by dichloromethane (DCM) due to the solvation effect causing broadening and redshifts in the bands. The fluctuation of helical structures during MD simulations is identified as a crucial factor in determining the dichroism of helicenes. Furthermore, vibronic calculations showed varied emission intensities among helicenes, with MeO[13]OMe exhibiting the highest emission intensity and CPL signal, implying superior optical properties.

In summary, the study demonstrates the successful computational modeling of the optical properties of helicenes using DFT methods and explicit solvation. The research contributes to understanding optical properties of helicenes and their potential applications in optoelectronic devices and chiral systems.
BIBLIOGRAPHY


Abhinay D. Joshi, Andrew Siderowf, Mark A. Mintun, and F.-A.
V.-A. Investigators. Relationships between flortaucipir PET tau
binding and amyloid burden, clinical diagnosis, age and cognition. 
*Brain*, 140(3):748–763, March 2017. ISSN 1460-2156. doi:
10.1093/brain/aww334.


J. M. H. Olsen, P. Reinholdt, and contributors. PyFraME: Python framework for Fragment-based Multiscale Embedding (version 0.4.0), 2021. See https://gitlab.com/FraME-projects/PyFraME.

Kestutis Aidas, Celestino Angeli, Keld L. Bak, Vebjørn Bakken, Radovan Bast, Linus Boman, Ove Christiansen, Renzo Cimiraglia, Sonia Coriani, Pål Dahle, Erik K. Dalskov, Ulf Ekström, Thomas Enevoldsen, Janus J. Eriksen, Patrick Ettenhuber, Berta Fernández, Lara Ferrighi, Heike Fliegl, Luca Frediani, Kasper Hald, Asger Halkier, Christof Hättig, Hanne Heiberg, Trygve Helgaker, Alf Christian Hennum, Hinne Hettema, Eirik Hjertenæs, Stinne...


LIST OF PAPERS AND MY CONTRIBUTIONS

Paper I Tau Protein Binding Modes in Alzheimer’s Disease for Cationic Luminescent Ligands
*J. Phys. Chem. B* 2021, 125, 42, 11628–11636

Paper II Comparative Analysis of Luminescent Cationic, Anionic, and Neutral Ligands Binding to Amyloid Fibrils in Alzheimer’s Disease: A Computational Insight
Todarwal, Y.; Linares, M.; Norman, P.
*Manuscript*

Paper III Distinct Heterocyclic Moieties Govern the Selectivity of Thiophene-Vinylene-based Ligands Towards Aβ or Tau Pathology in Alzheimer’s Disease

Paper IV Binding of a Pyrene-Based Fluorescent Amyloid Ligand to Transthyretin: A Combined Crystallographic and Molecular Dynamics Study

Paper V Tyrosine Side-Chain Functionalities at Distinct Positions Determine the Chirooptical Properties and Supramolecular Structures of Pentameric Oligothiophenes
*ChemistryOpen* 2020, 9, 1100–1108

Paper VI Naphthodithiophene Diimide Based Chiral π-Conjugated Nanopillar Molecules
Zhang, L.; Zhang, G.; Qu, H.; Todarwal, Y.; Wang, Y.; Norman, P.; Linares, M.; Surin, M.; Zhang, H.; Lin, J.; Jiang, Y.-B.
*Angew. Chem. Int. Ed.* 2021, 60, 24543–24548
Paper VII Ultralight aerogels via supramolecular polymerization of a new chiral perfluoropyridin-based sulfonimidamide organogelator
Proietti, G.; Axelsson, A.; Capezza, A.; Todarwal, Y.; Linares, M.; Norman, P.; Lendel, C.; Olsson, R.; Dinér, P.
Manuscript Submitted to Nanoscale

Paper VIII Modeling Absorption and Emission Spectroscopies of Symmetric and Asymmetric Azaoxahelicenes in Vacuum and Solution
Erbs Hillers-Bendtsen, A.; Todarwal, Y.; Pittelkow, M.; Norman, P.; Mikkelsen, K. V.

PAPERS NOT INCLUDED IN THIS THESIS

Dynamical Effects of Solvation on Norbornadiene/Quadricyclane Systems
Erbs Hillers-Bendtsen, A.; Todarwal, Y.; Pittelkow, M.; Norman, P.; Mikkelsen, K. V.
Manuscript Submitted to J. Phys. Chem.

CONTRIBUTIONS TO PAPERS

In Paper I, Paper II, Paper III, and Paper VII, I participated in conceptualizing the underlying frameworks and performed all computational calculations. In Paper IV, Paper VI, and Paper VIII, my contribution to the computational section was limited to the FF parametrization of the Py1SA ligand, the spectroscopic calculation of [4]C-NDTI isomers, and all aspects related to the molecular dynamics of azaoxahelicenes, respectively. I wrote the complete manuscript for Paper II and the first draft of the theoretical aspects of the manuscript for Paper VII. In Paper V, my contribution was the analysis and interpretation of spectroscopic data. For all papers, except Paper IV, I was involved in the data analysis and interpretation of the final computational results. For all the papers, I reviewed each manuscript and added the details of the calculations that I performed to the supporting information.