

# Understanding Fluorescent Amyloid Biomarkers by Computational Chemistry

Von der Naturwissenschaftlichen Fakultät  
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften (Dr. rer. nat)

genehmigte Dissertation

von

Thi Minh Nghia Nguyen, Máster, UCLV (Kuba), Máster, UVa (Spanien)

2023

Referentin : Prof. Dr. rer. nat. Carolin König  
Korreferent : Prof. Dr. rer. nat. Jörg August Becker  
Tag der Promotion : 22.08.2023

## DEDICATION

---

*Gia đình* means family.  
Family means nobody gets left behind, or forgotten.  
— Lilo & Stitch

To my beloved family, whose voice always guides me.



## ABSTRACT

---

Protein misfolding diseases, including neurodegenerative disorders like Alzheimer’s disease, are characterized by the involvement of amyloid aggregation, which emphasizes the need for molecular biomarkers for effective disease diagnosis. The thesis addresses two aspects of biomarker development: firstly, the computation of vibrationally resolved spectra of small fluorescent dyes to detect amyloid aggregation, and secondly, the binding and unbinding processes of a novel ligand to the target protein. In relation to the first aspect, a hybrid model for vibrational line shapes of optical spectra, called VCI-in-IMDHO, is introduced. This model enables the treatment of selected modes using highly accurate and anharmonic vibrational wave function methods while treating the remaining modes using the approximate IMDHO model. This model reduces the computational cost and allows for the calculation of emission line shapes of organic dyes with anharmonicity in both involved electronic states. The interaction between the dyes and their environment is also explored to predict the photophysical properties of the oxazine molecules in the condensed phase. The position and the choice of the solvent molecule have a significant impact on the spectra of the studied systems as they altered the spectral band shape. However, further studies are necessary to confirm the findings.

In addition to neurodegenerative diseases, the systemic amyloidoses represent another group of disorders caused by misfolded or misassembled proteins. In the cardiac domain, the accumulation of amyloid fibrils formed by the transthyretin (TTR) protein leads to cardiac dysfunction and restrictive cardiomyopathy. The investigation of binding and unbinding pathways between the TTR protein and its ligands is crucial for gaining a comprehensive understanding and enabling early detection of systemic amyloidoses and related disorders. Hence, exploring the different binding modes and the dissociation pathways of TTR-ligand complex is the primary objective of the second aspect of this thesis. The experimental study provides evidence of binding and X-ray crystallographic structure data on TTR complex formation with the fluorescent salicylic acid-based pyrene amyloid ligand (Py1SA). However, the electron density from X-ray diffraction did not allow confident placement of Py1SA, possibly due to partial ligand occupancy. Molecular dynamics and umbrella sampling approaches were used to determine the preferred orientation of the Py1SA ligand in the binding pocket, with a distinct preference for the binding modes with the salicylic acid group pointing into the pocket.

**Keywords:** Biomarkers, amyloid, theoretical spectra, multi-level, solvents, molecular dynamics, umbrella sampling, binding energy.



## ZUSAMMENFASSUNG

---

Proteinfehlverfaltungen, einschließlich neurodegenerativer Erkrankungen wie der Alzheimer-Krankheit, zeichnen sich durch die Beteiligung an der Amyloidaggregation aus, was die Notwendigkeit molekularer Biomarker für eine effektive Krankheitsdiagnose betont. Die Dissertation behandelt zwei Aspekte der Biomarker-Entwicklung: erstens die Berechnung vibrationsaufgelöster Spektren kleiner fluoreszierender Farbstoffe zur Erkennung von Amyloidaggregation und zweitens die Bindungs- und Lösungsvorgänge eines neuartigen Liganden an das Zielprotein. Zum einen wird ein hybrides Modell für die vibrationsaufgelösten Linienformen optischer Spektren namens VCI-in-IMDHO vorgestellt. Dieses Modell ermöglicht die Behandlung ausgewählter Moden mit hochgenauen und anharmonischen Schwingungswellenfunktionsmethoden und gleichzeitig die Behandlung der verbleibenden Moden mit dem approximativen IMDHO-Modell. Dieses Modell reduziert den Rechenaufwand und ermöglicht die Berechnung von schwingungsaufgelösten Emissionsspektren organischer Farbstoffe mit Anharmonizität in beiden beteiligten elektronischen Zuständen. Zum anderen wird die Wechselwirkung zwischen den Farbstoffen und ihrer Umgebung untersucht, um die photophysikalischen Eigenschaften der Oxazinmoleküle in kondensierter Phase vorherzusagen. Die Position und die Wahl des Lösungsmittelmoleküls haben einen erheblichen Einfluss auf die Spektren der untersuchten Systeme, da sie die spektrale Bandenform verändern. Weitere Untersuchungen sind jedoch erforderlich, um die Ergebnisse zu bestätigen.

Neben neurodegenerativen Erkrankungen stellen die systemischen Amyloidosen eine weitere Gruppe von Störungen dar, die durch fehlgefaltete oder fehlassemblierte Proteine verursacht werden. Im kardialen Bereich führt die Ablagerung von Amyloidfibrillen, die durch das transthyretin (TTR)-Protein gebildet werden, zu Herzfunktionsstörungen und restriktiver Kardiomyopathie. Die Untersuchung der Bindungs- Dissoziationspfade zwischen dem TTR-Protein und seinen Liganden ist entscheidend, um ein umfassendes Verständnis zu erlangen und eine frühzeitige Erkennung von systemischen Amyloidosen und verwandten Erkrankungen zu ermöglichen. Daher ist die Erforschung der verschiedenen Bindungsmoden und der Dissoziationswege des TTR-Liganden-Komplexes das Hauptziel des zweiten Aspekts dieser Arbeit. Die experimentelle Studie liefert Beweise für die Bindung und röntgenkristallographische Strukturdaten zur TTR-Komplexbildung mit dem fluoreszierenden salicylsäurebasierten Pyren-Amyloidliganden (Py1SA). Die Elektrodendichte aus der Röntgenbeugung erlaubte jedoch keine zuverlässige Platzierung von Py1SA, möglicherweise aufgrund einer partiellen Ligandenbesetzung. Molekulardynamik- und Umbrella-Sampling-Ansätze wurden verwendet, um die bevorzugte Ausrichtung des Py1SA-Liganden in der Bindungstasche zu bestimmen, wobei eine deut-

liche Präferenz für die Bindungsmodi mit der Salicylsäuregruppe in die Tasche zeigte.

**Schlagwörter:** Biomarker, Amyloid, theoretische Spektren, Multi-level, Lösungsmittel, Molekulardynamik, Umbrella-Sampling, Bindungsenergie.



## PUBLICATIONS AND AUTHOR CONTRIBUTIONS

---

### LIST OF PUBLICATIONS INCLUDED IN THIS THESIS AND AUTHOR CONTRIBUTIONS:

1. Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers.

**Nghia Nguyen Thi Minh** and Carolin König.

*Phys. Chem. Chem. Phys.*, 24(24):14825–14835, 2022.

- My contributions: I took charge of setting up and conducting all the calculations. Additionally, I curated and managed the research data, developed software tools for analysis, performed formal analyses and validation to obtain valuable insights. Moreover, I drafted and involved in the correction of the manuscript. I also played a key role in designing the figures.

2. Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands.

Yogesh Tadarwal, Camilla Gustafsson, **Nghia Nguyen Thi Minh**, Ingrid Ertzgaard, Thérèse Klingstedt, Bernardino Ghetti, Ruben Vidal, Carolin König, Mikael Lindgren, K Peter R Nilsson, Mathieu Linares, and Patrick Norman.

*J. Phys. Chem. B*, 125(42):11628–11636, 2021.

- My contributions: I analyzed the vibrational structure and calculated the vibronic spectra for the bTVBT4 ligand in the gas phase. I aided in interpreting the results and worked on the manuscript.

3. Binding of pyrene-based fluorescent amyloid ligand to transthyretin: A combined crystallographic and molecular dynamics study.

**Nghia Nguyen Thi Minh**, Afshan Begum, Jun Zhang, Petter Leira, Yogesh Tadarwal, Mathieu Linares, Patrick Norman, Dean Derbyshire, Eleonore von Castelmur, Mikael Lindgren, Per Hammarström and Carolin König.

*J. Phys. Chem. B*, 127(30):6628–6635, 2023.

- My contributions: I performed all calculations using molecular dynamics simulations and umbrella sampling techniques. I were responsible for the data curation, analyzing the simulation results and validating the findings. Moreover, I actively participated in writing a substantial part of the original draft and designed the figures.

4. The role of micro-solvation on the computed emission spectra: The case of oxazines.

**Nghia Nguyen Thi Minh** and Carolin König.

*(In preparation)*

- My contributions: I assumed the responsibility of setting up and performing all the calculations. Furthermore, I oversaw the curation of the data and conducted an in-depth analysis of the simulation results to validate the findings. Additionally, I actively contributed to a substantial portion of the original draft and played a key role in designing the figures.

#### LIST OF PUBLICATIONS NOT INCLUDED IN THIS THESIS

1. Quantum-derived embedding schemes for local excitations.  
Marina Jansen, **Nghia Nguyen Thi Minh**, Erik D Hedegård, and Carolin König.  
In *Chem. Modell.*, ed. Hilke Bahmann and Jean Christophe Tremblay, Specialist Periodical Reports, Royal Society of Chemistry, Cambridge, UK, 2022, vol.17, pp. 24–60.
2. Tamizaje fitoquímico y evaluación de la actividad sobre el sistema nervioso central del extracto etanólico de *Eugenia clarensis* Britton & P. Wilson.  
**Nghia Nguyen Thi Minh**, Liliana Vicet-Muro, Dany Siverio-Mota<sup>1</sup>, Maria E. Jorge-Rodriguez, Dulce M. González-Mosquera<sup>1</sup>, Idelfonso Castañeda-Noa.  
*J. Pharm. Pharmacogn. Res.* 4(1):39–48, 2016.

#### POSTER PRESENTATIONS

1. Tailored anharmonic-harmonic vibrational profiles for fluorescent biomarker, *57th Symposium on Theoretical Chemistry 2021 (STC 2021)*, online.
2. Tailored anharmonic-harmonic vibrational profiles for fluorescent biomarker, *European Summerschool of Quantum Chemistry 2022 (ESQC 2022)*, near Palermo, Italy.

## ACKNOWLEDGMENTS

---

First and foremost, I would like to extend my deepest appreciation to my supervisor, Prof. Carolin König. Thank you for your exceptional guidance, expertise, and unwavering belief in my abilities. I am immensely grateful to you for not only providing me with a solid foundation of knowledge and guidance but also for offering incredible opportunities that went beyond the academic realm. Your support extended to my overall personal and professional development.

I am also indebted to Prof. Patrick Norman, Prof. Mathieu Linares, Yogesh and other collaborators who graciously shared their knowledge and expertise with me. Their discussions have broadened my understanding and provided me with a solid theoretical framework for my research. To Diana, Erik and Nils, thank you for providing the valuable data for my work. It was a pleasure to work with all of you.

I would also like to thank all members of the doctoral committee for their time and effort dedicated to read this thesis.

Thanks for the financial support by the Deutsche Forschungs-gemeinschaft (DFG) through the Emmy Noether Young Group Leader Programme (project KO 5423/1-1), computing resources provided by the Rechenzentrum (RZ) at Kiel University, the Leibniz University IT Services (LUIS), and the Swedish National Infrastructure for Computing (SNIC).

To Marina and Janine, the “AKK Originals”, thank you both for being always supportive to me through my ups and downs over the last four years. To Marina, although I may still cringe at your taste in “screaming” music (just kidding), thank you for bringing order to the chaos of my “life-folder”, acting as the tech-savvy system administrator, and patiently explaining German phrases that left me utterly perplexed. Thank you, Janine, for being the sweetest girl I’ve ever met. Your hugs, your small acts of kindness (like the “Yogurette”), and your ability to bring a smile to my face have made this journey all the more delightful. To all of my colleagues and friends, Pascal, Erona, Onno, Anna, Florian, Romina, and many others, I want to express my deepest appreciation for the cherished friendship and the countless moments of joy and support we’ve shared.

To my Vietnamese friends, Huong, Giang, Trung, Trang, Hang, and Phuong Anh, I am grateful to have shared so many great times with you. Thanks for all the Vietnamese foods and funny jokes.

To Bac, my love, thank you for coming into my life and being my rock. Thank you for your infinite patience, your understanding, and for loving me even when I’m in the worst mood imaginable.

Finally, my gratitude for my family. To my mom who brought me into this world, thank you for your unconditional love and endless sacrifices. To my family-in-law, thank you for embracing me as your own and for welcoming me into your hearts and homes.

To every one who has inspired and supported me along the way:

*Thank you !*

Nguyễn Thị Minh Nghĩa  
Hannover, June 2023

# CONTENTS

---

Dedication	iii
Abstract	v
Zusammenfassung	vii
Publications	ix
Acknowledgments	xi
1 INTRODUCTION	1
1.1 Investigated systems . . . . .	5
<b>i BACKGROUND AND BASIS</b>	<b>11</b>
2 THEORETICAL BACKGROUND	13
2.1 Molecular mechanics and classical molecular dynamics . . . . .	13
2.2 Potential mean force and umbrella sampling . . . . .	18
2.2.1 The calculation of free energy . . . . .	18
2.2.2 Potential of mean force and umbrella sampling . . . . .	19
2.3 Quantum mechanics . . . . .	25
2.3.1 Electronic structure theory . . . . .	25
2.3.2 Vibrational structure theory . . . . .	34
2.3.3 Computation of molecular optical spectra . . . . .	42
<b>ii RESEARCH</b>	<b>47</b>
3 RESEARCH OVERVIEW	49
4 METHODS AND COMPUTATIONAL DETAILS	51
4.1 Vibrationally resolved UV/Vis spectra calculations . . . . .	52
4.2 Ligand-protein binding interactions . . . . .	55
4.2.1 Molecular dynamic simulations . . . . .	55
4.2.2 Umbrella sampling simulations . . . . .	56
5 TAILORED ANHARMONIC-HARMONIC VIBRATIONAL PROFILES FOR FLUORESCENT BIOMARKERS	59
5.1 VCI-in-IMDHO model . . . . .	60
5.2 Results and discussion . . . . .	63
5.2.1 Oligothiophenes, revisited . . . . .	63
5.2.2 Assessment of the anharmonicity measure . . . . .	66
5.2.3 Organic biomarkers . . . . .	67
6 ENVIRONMENT EFFECT ON THE VIBRATIONALLY RESOLVED EMISSION SPECTRA	73
6.1 Isolated oxazine dyes . . . . .	74
6.2 Effect of individual water molecules . . . . .	74
6.3 Effect of individual betaine molecules . . . . .	78

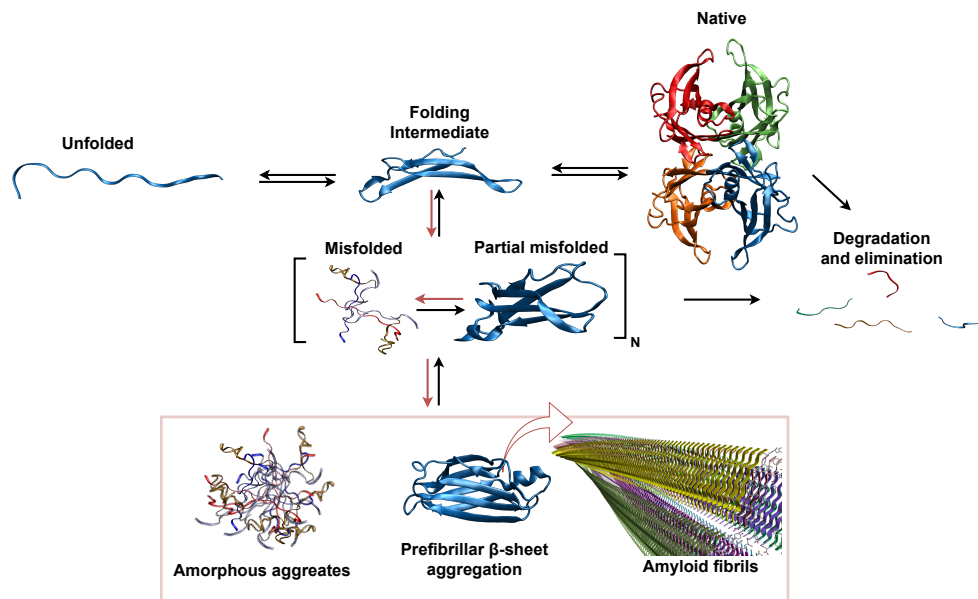
7	MOLECULAR DYNAMICS AND BINDING ENERGY OF PYRENE-BASED FLUORESCENT AMYLOID LIGAND TO TRANSTHYRETIN	81
7.1	Molecular dynamics simulations . . . . .	81
7.2	Umbrella sampling simulations . . . . .	85
8	SUMMARY, CONCLUSIONS, AND OUTLOOK	89
	BIBLIOGRAPHY	97
A	APPENDIX	125
A.1	Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers . . . . .	125
A.2	Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands . . . . .	147
A.3	The role of micro-solvation on computed emission spectra: The case of oxazines . . . . .	198
A.4	Binding of a pyrene-based fluorescent amyloid ligand to transthyretin: A combined crystallographic and molecular dynamics study . . . .	250
	Curriculum Vitae	283

## INTRODUCTION

---

Proteins are crucial biomolecules that participate in nearly all biological processes. To carry out their diverse and essential functions in living organisms, they must undergo a folding process to attain their native structure. [1–3] The folding process shown in Figure 1 is a complex and precise mechanism that enables proteins to adopt their three-dimensional shape, which is critical to their function. [2–4] Proteins coexist within a crowded and diverse cellular environment that can significantly impact their folding and intermolecular interactions. Alterations in the cellular environment such as changes in pH or temperature, or changes in the protein itself e.g. mutations, post-translational modifications, or overexpression, can cause misfolding or partial unfolding of proteins, which can then self-assemble into aggregate structures. [4–6] Consequently, a wide range of human diseases, commonly referred to *protein misfolding* (or *protein conformational*) diseases, result from the failure of specific peptides or proteins to properly adopt or retain their native functional conformation.

Folding abnormalities can be associated with diseases caused by the decreased presence of a specific protein that never achieves a functional folded structure. They can manifest at different life stages and are associated with genetic diseases as cystic fibrosis in the pediatric age or cancer appearing in later life due to somatic mutations in proteins controlling the cell cycle, i.e. p53. [6] The most extensive category of misfolding diseases, however, is associated with the intracellular or extracellular accumulation of insoluble, abnormally aggregated proteins. These structures are generally described as *amyloid fibrils* or *plaques* when they accumulate extracellularly, whereas the term *intracellular inclusions* are used when fibrils morphologically and structurally related to extracellular amyloid form inside the cell. [7] It is worth noting that the diseases induced by protein aggregation, in contrast to diseases caused by protein misfolding and loss of function, are predominantly associated with aging and there is typically a substantial delay before the initial deposition occurs. [6] They can be broadly categorized into three main groups: i) neurodegenerative disorders characterized by aggregation in the brain such as Alzheimer’s disease (AD), Parkinson’s disease (PD), ii) nonneuropathic localized amyloidoses characterized by aggregation in a specific tissue outside the brain like senile systemic amyloidosis (SSA), and iii) nonneuropathic systemic amyloidoses characterized by aggregation occurring in multiple tissues throughout the body such as type-2 diabetes. [4–6, 8–11] These maladies can be sporadic, inherited, or even infectious, and misfolded protein aggregates [8], particularly amyloid fibrils, are often considered biomedical markers of these diseases.



**Figure 1:** Schematic representation of protein folding/misfolding processes. During folding, the polypeptide chain transitions through various intermediate conformations, eventually arriving at its native, functional state. Misfolding can occur when a protein fails to attain its correct conformation, leading to the formation of misfolded intermediates, which may be partially or fully misfolded. Misfolded proteins can form aggregates, which can then accumulate and disrupt normal cellular processes. In some cases, these aggregates can lead to the formation of amyloid fibrils (PDBj ID: 2m5k). Amyloid fibril formation involves the unfolding of native proteins, leading to the formation of partially folded intermediates. These intermediates can associate with each other to form a metastable  $\beta$ -sheet conformations. Over time, they gradually undergo conformational conversion to  $\beta$ -sheet-rich structures, which can act as nucleation seeds, facilitating the formation of additional oligomers or fibrils.



Amyloid fibrils are highly ordered misfolded protein aggregates. They exhibit a distinctive arrangement known as a *cross- $\beta$  sheet array*. This structure arises from the stacking of  $\beta$ -strands within the fibril, forming sheets that are perpendicular to the fibril's central axis and stacking together to form individual filaments. [9] Mature fibrils typically consist of two to six protofilaments that twist around each other to form rope-like fibers (as the amyloid fibril depicted in Figure 1). These fibers have a diameter ranging from five to ten nanometers and can extend up to a few microns in length [9, 12]. The fibrils formed are often unbranched, highly stable, and resistant to degradation by proteases and denaturants. These specific properties contribute to the challenge faced by cells in eliminating the fibrils once they have been formed.

Amyloid fibrils typically arise from partially or fully unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. [9] For example, the major component of AD-associated amyloid deposits is the 39- to 42-residue-long amyloid  $\beta$  ( $A\beta$ ) peptide. [8, 9] It is noteworthy that  $A\beta$  1-40 is the major variant in brain, while the  $A\beta$  1-42 fragment is the dominant  $A\beta$  species in the amyloid plaques of AD patients. [8, 9] The primary structure of  $A\beta$  plays a crucial role as variants of different lengths exhibit distinct characteristics in terms of solubility and the formation of amyloid fibrils. For instance,  $A\beta$  1-40 is relatively soluble in aqueous media, whereas the  $A\beta$  1-42 which has two extra hydrophobic amino acids compared to  $A\beta$  1-40, quickly forms insoluble, fibril-like structures. [13] Another example of amyloid fibrils-linked diseases is the systemic amyloidoses which are often associated with accumulation of significant amounts of amyloid deposits (even kilograms in some cases) in the affected tissues and organs. [14] Systemic amyloidosis involves the formation of amyloid fibrils composed of a diverse range of at least twenty different proteins [15]. Particularly, in the heart, accumulating of amyloid fibrils formed by transthyretin (TTR), a homotetrameric plasma protein implicated in the transport of thyroxine (T4) and retinol (vitamin A), lead to progressive ventricular wall thickening and stiffness. As a consequence, diastolic dysfunction develops, eventually progressing to a condition known as restrictive cardiomyopathy. [16, 17] Cardiac amyloidosis can manifest in different forms, including amyloid light chain (AL) amyloidosis associated with underlying plasma cell dyscrasia and amyloid transthyretin amyloidosis (ATTR). The latter can be of either non-mutant, also known as wild-type, TTR occurring in elderly individuals and hereditary leading to the wild-type transthyretin amyloidosis (ATTRwt) or mutant TTR (ATTRm) caused by genetic mutations that render the TTR protein unstable. [16-18] The dissociation of TTR tetramers is a crucial step in the formation of amyloid fibrils. [19]

Developments in experimental techniques [12, 20-22] have resulted in increasingly detailed knowledge of the molecular structures of amyloid fibrils. Yet, the diagnosis and treatment of amyloid-related disease e.g. AD and ATTRwt are still limited [16, 23] and mainly when patients already exhibit early cognitive losses. Thus, it is crucial to detect these diseases in their presymptomatic stage, as this would enable the development of an efficient and fast-acting treatment for these devastating disorders at an early stage. Apart from the direct determina-

tion of the structure by means of X-ray and neutron diffraction experiments, optical spectroscopy is a powerful tool in studying the interaction of light with biomolecules and their environments. [24] The non-invasive nature and ability to visualize biological processes with high precision are one of many reasons for its wide-spread implementation. [25–27] To enhance accuracy and sensitivity, the development of chemical tools that can detect biological species and monitor key physiological processes is critical. These chemical tools are commonly referred to as fluorescent probes, reactive dyes or fluorophores. [11] An ideal fluorescent probe should possess several key characteristics, including high selectivity for the target pathological marker whilst in the presence of competing species, excellent biocompatibility, and long-wavelength excitation (a crucial requirement for *in vivo* imaging). [28] The addition of a fluorescent biomarker to aggregated protein, for instance, provides detection capability arising from alterations in the absorption and fluorescence spectra. These spectral changes manifest as shifts in color, providing a sensitive and non-invasive means to monitor and detect protein aggregation events. However, interpreting the complex indirect information derived from the analysis of experimental spectra is often challenging due to the intricate interplay of various factors, making it difficult to separate and assess their specific contributions. In such a context, computational spectroscopy is undoubtedly a powerful and reliable tool to unravel the different contributions to the spectroscopic signal and understand the underlying physical phenomena. Optical line shapes, in particular, offer abundant insight into various aspects of molecules in excited electronic states. Such line shapes can yield information about a molecule’s lifetime, equilibrium structure, stiffness, vibrations, and interaction with its surroundings. [29, 30] Computing an accurate spectra, performing a comprehensive analysis of spectroscopic signals and extracting information about molecular properties can be a challenging task, especially for medium to large molecules in condensed phases.

Theoretical spectroscopy has made tremendous progress in the past few decades. It employs a synergistic blend of quantum mechanics and molecular mechanics to investigate the intricate details of light–matter interactions and spectroscopic properties. On one hand, by applying quantum mechanical methods, such as (time-dependent) density functional theory or *ab initio* calculations, the electronic structure and energy-related properties of relatively small molecules in gas phase can be accurately determined, allowing for the prediction of absorption, emission, and vibrational spectra. [31–33] Molecular mechanics, on the other hand, provides insights into the molecular conformation, dynamics, and intermolecular interactions [34] that influence the properties of biomarkers and the observed spectra. Despite the remarkable advancements in theoretical chemistry, several challenges persist in studying complex molecular systems and the transitions between electronic states are more complex to handle for several reasons. [35–37] One significant challenge arises from the size of the molecules, which increases the computational demands of accurate quantum mechanical calculations. In practical applications, when dealing with medium-sized systems, it is common to utilize a harmonic approximation to describe the molecular vibrations for calcu-

lating transition energies and intensities. [35, 38] However, in many cases, accurate simulation of anharmonic vibrationally resolved electronic spectra is highly demanding, yet often required to interpret experimental spectra. Incorporating anharmonic corrections further complicates matters and significantly increases computational costs, making it feasible only for small molecules [39, 40] or specific degrees of freedom [32, 41]. Additionally, the influence of the surrounding environment, such as solvents or biological environment, introduces further complexity, as it can modify the electronic structure and affect the spectroscopic properties of the molecule. As a result, accuracy is often sacrificed for interpretability when studying larger and more complex systems. To overcome this challenge, simplified models and approximations are often used, but they may not fully capture the true complexity of the system. Hence, there is a need for more efficient and accurate computational methods that can handle larger and more complex systems while maintaining high accuracy.

Therefore, in this thesis, I focus on developing and implementing theoretical approaches for the reliable calculation and prediction of vibrationally resolved emission spectra of pathological biomarkers. Specifically, I investigate the use of a reduced-space vibrational configuration interaction (VCI) anharmonic wave function treatment within the independent mode, displaced harmonic oscillator (IMDHO) model to estimate the impact of individual modes on the vibronic profile of several fluorescent dyes. Furthermore, the effects of the environment, including solute-solvent interactions in the continuum and micro-solvation model on the vibronic spectra of oxazine dyes have been explored through quantum chemical calculations. This study provides the first insight into how solvent molecules in the first solvation shell affect the spectral properties of these dyes. Such findings are of great importance for understanding the photochemistry and photophysics of these dyes in realistic environments.

In addition, to comprehend the intrinsic properties of such fluorescent probes, understanding their interaction with the target protein is essential for accurately interpreting spectroscopic data and designing efficient ligands. This includes studying the binding and unbinding processes. Thus, the other goal of this thesis is to provide computational evidence of binding on TTR complex formation with a salicylic acid-based pyrene fluorescent ligand (Py1SA) by using molecular dynamic and umbrella sampling models. Further details about the investigated systems can be found in the following subsection.

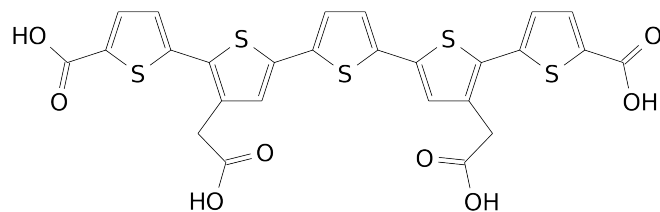
## 1.1 INVESTIGATED SYSTEMS

As mentioned earlier, fluorescent imaging is an extremely powerful tool for a detection of misfolded proteins. This detection can be achieved using small, molecule-sized fluorescent probes that have the ability to specifically bind to the targeted proteins. However, conventional probes such as thioflavin T [42] and congo red [43] are not suitable for clinical studies due to their toxicity and inability to cross the

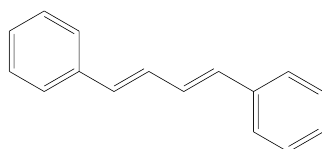
blood–brain barrier (BBB). In this regard, a novel group, namely luminescent conjugated oligothiophenes (LCOs), [44] has been developed. LCOs show aggregate-specific strong fluorescent signals upon binding to a wide range of protein aggregate morphotypes. [44–49]. They have the ability to pass through the BBB [50] and facilitate early-stage detection of the buildup of misfolded protein aggregates [44–46, 49].

The introduction of LCOs as amyloid imaging agents has revealed the advantages of using a flexible backbone compared with the sterically rigid scaffold found in conventional dyes. One of the LCOs examined in this thesis is the oligothiophene-based fluorescent dye HS84 [48] (Figure 2). HS84 is an isomer to p-FTAA, a highly promising compound for the detection of  $\beta$ -sheet containing proteins due to its ability to induce specific conformation-dependent shifts in emission spectra, which vary depending on the protein’s nature. [44] In contrast to p-FTAA, HS84 features acetic side chains of the trimeric building block arranged in a tail-to-tail configuration instead of head-to-head. [48] In solution, HS84 produces a broad fluorescent signal, whereas when mixed with recombinant A $\beta$  1-42 amyloid-like fibrils, it produces a structured fluorescent signal [49]. Additionally, the aggregation of HS84 dye to the amyloid fibril results in a change in fluorescence color, which is particularly noteworthy because it represents a promising avenue for improved *in vivo* detection of amyloid fibrils [50]. The other ligand subject of the present thesis, bTVBT4, belongs to the bi-thiophene-vinylene category of LCOs and features a cationic benzothiazole moiety, as shown in Figure 2. Unlike p-FTAA, bTVBT4 exhibits binding specificity for the tau aggregates, which is one of two major proteinaceous deposits in brain tissue with AD pathology. [51]

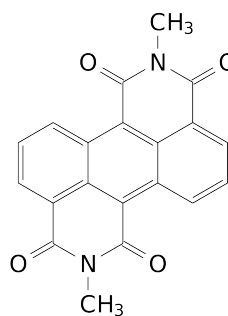
Furthermore, in order to assess the applicability of the VCI-in-IMDHO model to larger systems, my research delves into the properties of two other organic compounds: 1,4-diphenylbutadiene (DPB) and an anthracene diimide (ADI) derivative (Figure 2). DPB is a fluorescent form of linear polyenes, which have been extensively studied as models for photochemical and photobiological systems. [52, 53] Interestingly, the first two excited singlet states of DPB are nearly degenerate with the ordering dependent on the environment [53, 54]. In the absence of solvent, the ground and excited states of DPB are planar. In solution, however, DPB is believed to adopt a non-planar conformation for its ground state, with the two phenyl rings rotated at an angle of 75° relative to the butadiene unit [54] influencing the photophysics and photochemistry of the polyenes. Another extensively studied organic class is that of aromatic imides and diimides. Among those, six-membered tetracarboxylic aromatic diimides, typically naphthalene diimide (NDI) and perylene diimide (PDI), attract enormous attention because of their high electron affinity and mobility. They demonstrated their potential as fluorescent probes in biological systems [55–57], sensors for the detection of chemical species [58, 59], and electron acceptors in organic photovoltaics [60, 61]. As their analogue, ADIs and derivatives are also of great appeal because of their exceptional chemical and photochemical properties. [60–62]



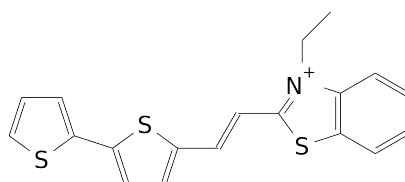
**HS84**



**DPB**

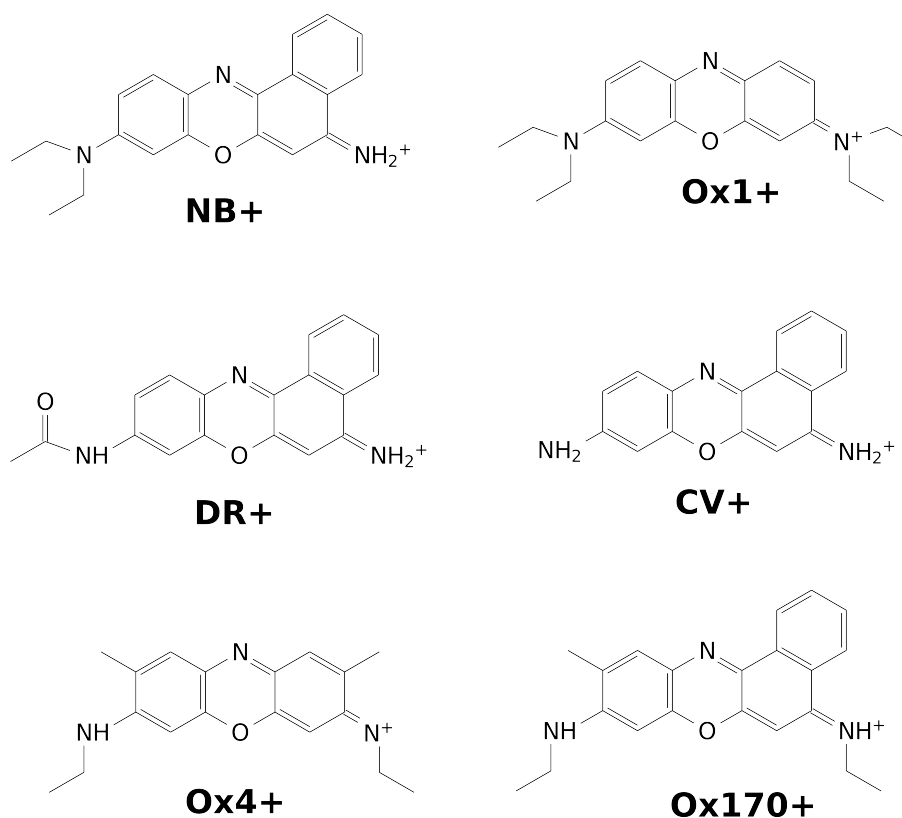


**ADI**



**bTVBT4**

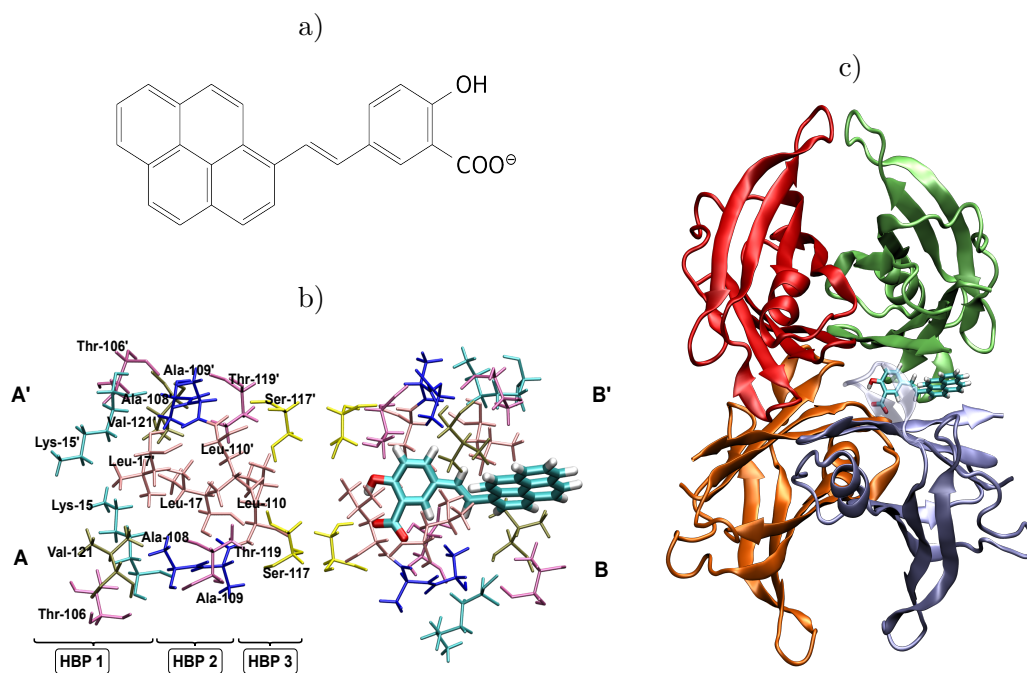
**Figure 2:** Representation of the investigated dyes: pentameric oligothiophene derivative (HS84), 1,4-diphenylbutadiene (DPB) and anthracene diimide (ADI), and bithiophene vinylene benzothiazole 4 (bTVBT4).



**Figure 3:** Chemical structures of investigated oxazine: nile blue (NB+), oxazine 1 (Ox1+), darrow red (DR+), cresyl violet (CV+), oxazine 4 (Ox4+), and oxazine 170 (Ox170+).

The other investigated systems in this thesis involve oxazine molecules. Oxazines are heterocyclic compounds that exhibit unique photochemical and photophysical properties, making them a popular choice for various applications in biology, histology [63–65] and laser technology [66]. Their photophysical properties are strongly dependent on solvent polarity and pH, thereby providing information about their immediate chemical environment. [67] Some of these molecules exist as neutral species while others bear a positive charge, which is particularly significant for certain biotechnological applications. [68] The cationic form of oxazine dyes absorbs in the red spectral region (550–650 nm) with a fluorescence maximum exhibiting a Stokes shift of about 30 nm in ethanol. [66] The structure of oxazine dyes examined in this thesis including cresyl violet (CV+), darrow red (DR+), oxazine 1 (Ox1+), oxazine 4 (Ox4+) and oxazine 170 (Ox170+) are depicted in Figure 3.

Furthermore, beyond investigating the inherent spectroscopic characteristics of the probes in both the gas and condensed phases, it is crucial to examine the process of ligand binding and unbinding to the target protein. Transthyretin (TTR) is a highly conserved protein, mainly synthesized in the liver and choroid plexus [69]. In addition to the well-known T4-carrier role [70], it has been suggested to function as a molecular chaperone to prevent aggregation of proteins associated with neurodegenerative diseases [71]. From the structural point of view, TTR is a tetrameric protein comprising four identical subunits of 127 amino acids with a monomer molecular weight of approximately 14 kDa. [69] TTR related amyloidogenesis lacks an effective therapy, although it has been observed that amyloid fibril formation is prevented by the binding of the small molecule T4. [72] The TTR tetramer structure shows two funnel-shaped binding sites for the T4 hormone, each defined by a dimer-dimer interface (Figure 4c). These binding sites consist of an inner and outer binding cavity. Six halogen binding pockets (HBPs) were identified based on the positions of the halogen atoms of T4 in the TTR:T4 crystal structure. The outer binding site comprises HBP 1 and 1' formed primarily by residues Lys-15, Leu-17, Thr-106, and Val-121 of both subunits. The inner binding cavity is composed of HBP 3 and 3' formed by the side chains of Ser-117, Leu-110, Thr-119 and Ala-108 of both subunits. The interface of the inner and outer binding cavities comprises residues Leu-17, Ala-108, Ala-109, and Leu-110 of both subunits, defining HBP 2 and HBP 2'. Hydrogen bonding to the bound ligand is enabled by buried Ser-117 and Thr-119 at the bottom of the binding site. Lys-15 residues flank the entrance to the binding site from the opposing sides. [73, 74] Typically, the binding of TTR inhibitors and T4 occurs in the so-called *forward* binding mode, where the anionic substituents, such as carboxylate, are positioned in the outer binding pocket and interact electrostatically with the Lys-15  $\epsilon$ -ammonium groups ( $-\text{NH}_3^+$ ). However, the *reverse* binding mode, where the carboxylate is oriented towards the inner binding pocket to hydrogen bond with Ser-117 and Ser-117', is also observed and has been previously reported in TTR complexes with diclofenac (a biarylamine), several diclofenac analogues, and some diflunisal analogs.[74–76] Figure 4c depicts the tetrameric TTR with the novel ligand (Py1SA) and Figure 4b shows the close-up view of the hormone binding pocket. Despite the availability of TTR structures in the Protein Data Bank (PDB) and its discovery in 1978, [77] the binding modes of TTR ligands continue to be an active area of research. Recently, Py1SA, a trans-stilbene ligand with a salicylic acid moiety (Figure 4a), was identified as a potential amyloid fibril probe for multiple amyloid proteins and was found to bind to TTR's T4 binding site. [78–80] Upon binding to preformed fibrils Py1SA showed a broad emission band around 498 nm, attributed to the strong intramolecular charge transfer between pyrene and benzene ring. [80] To determine the potential of Py1SA as a diagnostic tool and kinetic stabilizer for TTR amyloidosis, a thorough understanding of the binding modes between Py1SA and TTR is essential. Theoretical studies play a pivotal role in advancing our understanding of such amyloid biomarkers.



**Figure 4:** a) Chemical structures of Py1SA, b) the halogen binding pockets (HBPs) of the thyroxine ( $T_4$ ) hormone binding site, and c) the orthogonal views of the crystal structures of the TTR–Py1SA complex. The individual molecules of the TTR tetramer (shown as ribbons) are highlighted in colors (orange: A, red: A', lilac: B, green: B'). Py1SA (in sticks, here reverse binding mode) binds in the  $T_4$  binding site, formed at the B:B' interface. The equivalent A:A' site is devoid of ligand. The inner binding pocket, HBP 3, is located between the side chains (in sticks) of Ser-117, Thr-119, Ala-108, and Leu-110, the central HBP 2 pocket is formed by the side chains of Leu-17, Ala-108, Ala-109 and Leu-110 and the outer pocket HBP 1 is located between the side chains of Lys-15, Leu-17, Thr-106 and Val-121.



Part I

BACKGROUND AND BASIS



## THEORETICAL BACKGROUND

---

Due to the complexity outlined in the previous chapter, accurate atomistic modeling of optical properties of isolated fluorescent dyes as well as their interactions with the solvents or protein is extremely challenging: It requires the integration of different computational models, which span from quantum–mechanical methods to classical models. Given that, this chapter emphasizes the essential traits peculiar to simulation techniques in general as well as in the molecular dynamics (MD) approach and computational vibronic spectroscopy in particular. It begins by exploring the theory behind classical MD simulations including the empirical energy function or force field, which describes the interaction between different atoms within the system. Following that, an enhanced MD simulation strategy, namely the umbrella sampling (US) method is introduced and discussed. This method is utilized in this thesis to explore and investigate the complex interactions and binding modes between a ligand and target protein.

In addition to MD simulations, quantum mechanics (QM) plays a vital role in understanding the behavior of molecular systems. While MD provides valuable information about the dynamics and interactions of molecules at the atomic level, QM calculations offer a more accurate description of electronic structure and energy properties. Therefore, the chapter sheds light on the importance of considering the quantum mechanical nature of the fluorescent probe to accurately predict its optical properties. This is achieved by presenting the fundamental principles of electronic and vibrational structure theory, which are essential for a comprehensive understanding of the molecular behavior of fluorescent probes. Different methods for the ground state calculation e.g. Hartree–Fock (HF) theory and density functional theory (DFT) are discussed. The chapter also explains the use of time dependent (TD) DFT with linear response formalism for the calculation of excited state properties. In addition to electronic effects, vibrational effects will be discussed, with a particular focus on the variational self-consistent field (VSCF) and vibrational configuration interaction (VCI) methods. The chapter closes up by emphasizing the calculation of optical properties within the framework of the Franck–Condon (FC) principle, and the independent mode, displaced harmonic oscillator (IMDHO) approximation.

### 2.1 MOLECULAR MECHANICS AND CLASSICAL MOLECULAR DYNAMICS

Molecular dynamics (MD) is a technique to predict how every atom in a many-body system moves over time based on classical models, i.e. using Newtonian

mechanics. [81–83] Since its initial implementation in the late 50s [84], MD simulation has become more popular and visible in recent years [82, 83], particularly in the field of biology and drug discovery [83]. It has been used to address a wide variety of important biomolecular processes such as ligand binding [85–88], protein folding [89–92], conformational change [93–95], mechanisms of large biomolecular networks (e.g. virus simulations [96]), and drug application (e.g. drug for coronavirus disease 2019 (COVID-19) [97]), etc.

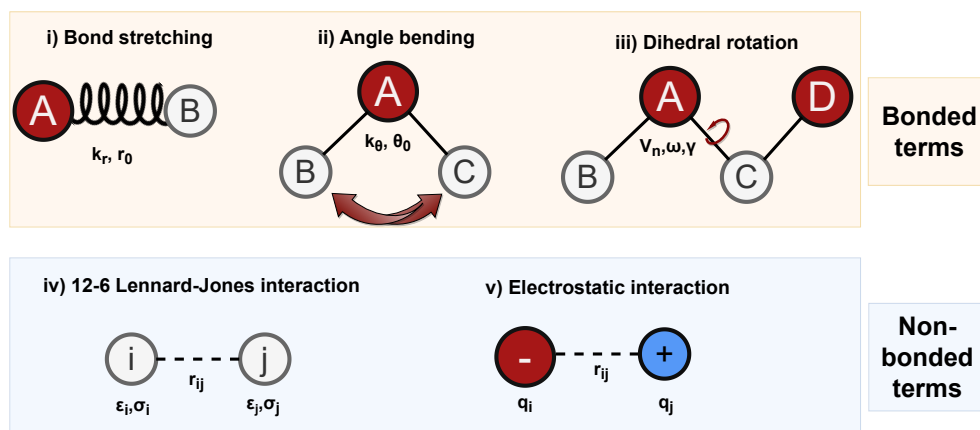
The concept behind MD simulation is based on the validity of several assumptions including the Born–Oppenheimer approximation [98] leading to the separation of electronic and nuclear motions. Unlike QM methods, molecular mechanics (MM) neglect the electronic motions of a system and calculate the energy of a system as a function of nuclear position only [34]. Within the framework of MM, the molecular systems can be represented in terms of a relatively simple model of interactions or a so-called force field (FF).

A FF can be thought of as a map from atom positions to an energy value including intra- and inter-molecular forces within the system [34], for example, a class-I FF can be written as

$$\begin{aligned}
 V = & \underbrace{\sum_{\text{bonds}} \frac{1}{2} k_l (r_l - r_{l,0})^2 + \sum_{\text{angles}} \frac{1}{2} k_m (\theta_m - \theta_{m,0})^2 + \sum_{\text{torsions}} \frac{1}{2} V_n (1 + \cos(n\omega - \gamma))}_{\text{bonded}} \\
 & + \underbrace{\sum_{i < j}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)}_{\text{non-bonded}}
 \end{aligned} \tag{1}$$

where  $V$  denotes the potential energy composed of bonded and non-bonded interaction energies. The different contributions are schematically represented in Figure 5. The first and second terms in equation 1 cover the dynamic of covalent bond stretching and angle bending of the system. In the class-I FF, they are described by simple harmonic motion, i.e. the magnitude of restoring force is assumed to be proportional to the displacement of the actual bond length  $r_l$  or bond angle  $\theta_m$  from the equilibrium position  $r_{l,0}$  and  $\theta_{m,0}$ , respectively.  $k_l$  and  $k_m$  are the stretching and bending force constants that are determined empirically. The third term in equation 1 is a summed over all torsional angle rotations between atoms that are vicinal to each other. Here,  $V_n$  is the barrier to free rotation for the “natural” bond,  $n$  is the periodicity of the rotation,  $\omega$  is the torsion angle, and  $\gamma$  is the phase factor.

Independent atoms and molecules interact through non-bonded forces. The non-bonded interactions are usually considered in two groups, one comprising van der Waals interactions and the other electrostatic interactions. Typically, the van der Waals contribution is built by the Lennard-Jones 12-6 functions [34]. This potential is commonly expressed in terms of the well depth  $\epsilon$  (the measure of the strength of the interaction) and the van der Waals radius  $\sigma$  (the distance at which



**Figure 5:** Schematic representation of a typical force field model. Energy dependencies are related to i) stretching or compressing a bonded pair of atoms, ii) increasing or decreasing the bond angle, iii) dihedral angle rotations, iv) van der Waals interactions (modeled by Lennard–Jones potential) and v) electrostatic interactions (modeled by Coulomb’s law). (i–iii) arise from interactions between atoms that are chemically bonded to one another while (iv–v) arise from interactions between atoms that are not bonded.

the intermolecular potential between the two particles is zero). Electrostatic terms describe the Coulomb interaction between atoms with partial charges  $q_i$  and  $q_j$  at a distance  $r_{ij}$ . Here,  $\epsilon_0$  is an effective dielectric constant which is one in vacuum but higher when there are intermediate atoms or solvent. More sophisticated FFs may have additional terms [34, 99–113] to predict certain properties of the system.

Among a plethora of FFs, the Amber ff14SB [114] is a commonly used FF for MD simulations of proteins and nucleic acids. It improves the treatment of side-chain and backbone torsions and has a revised treatment of hydrogen bonding, leading to better agreement with experimental data for proteins. Another commonly employed FF in MD simulations is the General Amber Force Field (GAFF). GAFF is renowned for its versatility and is widely used for a diverse range of molecules, including small organic compounds and drug-like molecules. [115, 116] The combination of GAFF with other FFs such as the Amber FF for proteins, enables the simulation of complex systems involving both biomolecules and small organic compounds.

As shown in equation 1, in conventional FFs, modelling of electrostatic interactions via pairwise Coulomb potentials is based on assigning fixed point charges (also known as “partial charges”) to atoms. These forces are particularly important in biomolecules since the strengths of hydrogen bonds are largely determined by the electrostatics of these polar interactions. Currently, the electrostatic potential (ESP) charges, derived from fitting a classical Coulomb model to quantum mechanical molecular electrostatic potentials, are commonly employed in simulations of macromolecules. However, it is known that these charges tend to overestimate bond polarities in the gas-phase. [117] To mitigate this overestimation, a restraint

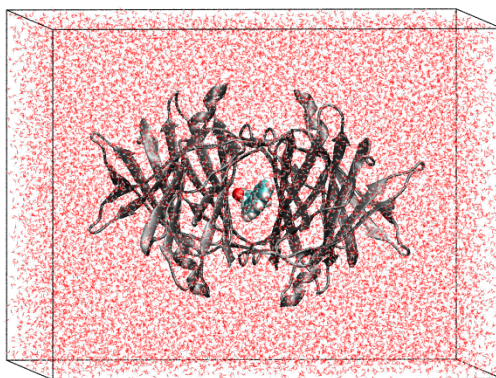
function can be applied during the fitting process to attenuate the magnitudes of the charges, resulting in what is known as RESP-charges. [117, 118] Despite the limitations of fixed-charge models in accurately capturing charge transfer and polarization, simulation methods utilizing RESP-charges often yield reasonably accurate results in aqueous solutions with significant lower computational effort compared to polarizable models. [119]

Once the forces acting on individual atoms are obtained, MD simulations use classical Newton’s law of motion to calculate accelerations ( $\mathbf{a}$ ) and velocities ( $\mathbf{v}$ ) and to update the atom positions ( $\mathbf{q}$ ) generating the trajectory of the molecular system by integrating the equation of motion

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \left( \frac{d^2 \mathbf{q}_i}{dt^2} \right). \quad (2)$$

The force on atom  $i$  is denoted by  $F_i$  and  $t$  is the time. Besides the accuracy of the underlying FFs, the predictive ability of MD simulations relies also on the efficiency of the algorithms used to integrate the equations of motion. The most widely used MD algorithms in MD studies are Verlet [120], Beeman [121], and Leapfrog [122]. All methods used to calculate MD simulations are based on expressing  $\mathbf{q}$ ,  $\mathbf{v}$ , and  $\mathbf{a}$  of the system as a Taylor series. Since they generate identical trajectories [123], the latter is usually the preferred scheme because of its minimal computer memory storage and computational requirements [123]. In all schemes, the integration is performed in small time steps  $\delta t$ , normally between 1 and 10 fs for atomistic simulations [123], which is the major bottleneck of the simulation procedure, particularly in large-scale dynamic simulations. As  $\delta t$  in MD simulations is limited by the highest frequency motions occurring in the system, it can be increased by freezing the internal vibrations, such as bond-length or possibly bond-angle vibrations [123].

Furthermore, most experiments are conveniently undertaken at constant pressure and temperature. Therefore, to mimic experimental conditions, many MD simulations, including those of this thesis, are rather performed under constant temperature and/or pressure, employing thermostats [124] and/or barostats [125, 126] algorithms to constrain thermodynamics variables along the simulations. The Bussi stochastic velocity rescaling thermostat (V-rescale) [127, 128] is a popular algorithm for temperature coupling in MD simulations. It ensures that the system remains at a desired temperature by adding a stochastic term to the velocity rescaling algorithm, which results in a more realistic temperature distribution of the system. The V-rescale thermostat is particularly useful for simulations of systems with long-range electrostatic interactions, where other thermostats may introduce artifacts in the sampling of phase space. In contrast, barostats are used to control the pressure in MD simulations. There are several types of barostats [34], one of which is the Berendsen barostat [129]. In this method, the system pressure is regulated by introducing an external bath, to which the global pressure is weakly linked. The volume is periodically rescaled with the aim of producing the least local disturbance while preserving the required global coupling. The cou-



**Figure 6:** *Simulation box of targeted protein (grey) related to amyloid disease with novel bound ligand Py1SA (cyan) in water (red).*

pling strength can be modified, thereby allowing for easy control and evaluation of the coupling effect.

Furthermore, as most biological reactions occur in water, it is crucial to consider effects of solvent in MD simulations. The evaluation of solvent effects is a major challenge in developing analytical and numerical molecular modeling methods. Equation 1 can incorporate solvation effects in two different ways. The first is an implicit solvation model, which modifies interaction parameters to create a solvated potential of mean force [130, 131]. This is less accurate but requires fewer computer resources. The second method is an explicit inclusion of solvent molecules in the model, treated at the same level as groups within the protein, represented by additional terms in the equation [132]. This more accurate approach, but also more demanding of computer resources, requires an accurate solvent-solvent potential to describe the major properties of the solvent and a way to determine an effective solute-solvent potential based on individual solvent-solvent and solute-solute parameters. Among a vast number of water models [103, 133–136], rigid nonpolarizable water models e.g. TIP3P [103] 3-point models are the most commonly used models in biomolecular studies due to their simplicity and computational efficiency. In this model, water is explicitly represented as a set of point charges at fixed positions relative to the oxygen nucleus.

In addition, for a better description of bulk properties in finite-sized systems, periodic boundary conditions (PBC) are also commonly employed in MD simulations. The calculation of full electrostatics is achieved through the Ewald sum methods, whose efficiency can be improved with the use of spherical cutoffs and an appropriate splitting parameter between short and long-range interactions i.e. particle-mesh Ewald (PME) method [137]. By using these algorithms, MD simulations now enable researchers to observe events in realistic conditions for hundreds of microseconds or more.

The benefits of using MD simulations are numerous, as it provides information for nuclear degrees of freedom in a system, including dynamic, structural, and energetic information for the entire system and its components. It is useful in the

simulation of large molecules like proteins, where quantum mechanics is not a realistic option. Figure 6 illustrates the typical simulation of a solvated protein (TTR) with its novel ligand (Py1SA). The system contains about 75,000 atoms, which is far beyond the current limit for quantum mechanic calculation [138]. Thus, MD simulation gives researchers insight into the behavior of complex systems and the ability to predict their behavior in different conditions. It has greatly impacted various fields and remains an important tool for advancing our understanding of nature.

## 2.2 POTENTIAL MEAN FORCE AND UMBRELLA SAMPLING

### 2.2.1 *The calculation of free energy*

Free-energy calculations are important in many areas of computational science, including chemistry, materials science, and biophysics [139–141]. They are used to predict the outcome of chemical reactions, the stability of biomolecules, and the behavior of materials. [142–144] From statistical mechanics, the free energy (here the Gibbs free energy,  $G$ ) can be calculated by evaluating [34]

$$G = -\frac{1}{\beta} \ln Q, \quad (3)$$

where

$$Q = \iint \exp(-\beta H(\mathbf{q}^N, \mathbf{p}^N)) d^N \mathbf{q} d^N \mathbf{p}. \quad (4)$$

Here,  $\beta = 1/k_B T$  where  $k_B$  is the Boltzmann constant at the absolute temperature  $T$  and  $N$  is the number of degrees of freedom of the system.  $H(\mathbf{q}^N, \mathbf{p}^N)$  gives the total energy of the system in a given configuration, i.e., a given set of momenta  $\mathbf{p}$  and coordinates  $\mathbf{q}$ . If the volume, rather than the pressure, is kept constant, the Helmholtz free energy (usually denoted as  $A$ ) is obtained. Most experiments are conducted under conditions of constant temperature and pressure, where the Gibbs free energy is the appropriate free energy quantity.

The absolute free energy in equation 3 can only be calculated directly for small simple systems governed by a simple Hamiltonian [34, 145, 146]. For larger systems with strong interactions between particles, obtaining an analytical formulation of the partition function is not feasible. We should bear in mind that in many instances, the potential energy  $V(\mathbf{q}^N, \mathbf{p}^N)$  of the system is usually the primary focus rather than the total energy. [34] If the potential energy is independent of the momentum, then the contribution of the momenta to the partition function  $Q$  can be considered as a constant, which can be neglected [147]. Consequently,

$$Q = \int \exp(-\beta V(\mathbf{q})) d^N \mathbf{q}. \quad (5)$$

The other important feature of equation 3 is that configurations with high energy significantly contribute to the integral due to the presence of the exponential



term  $\exp(-\beta H(\mathbf{q}^N, \mathbf{p}^N))$ . Standard MD simulations preferentially sample low-energy regions of the phase space [34]. Such simulations will never adequately sample the important high-energy regions. Hence, calculating the free energy using a conventional simulation will lead to poorly converged and inaccurate values. It is worth noting that, in many cases, one is generally interested in free-energy differences between two states instead of the absolute quantities. [34] These differences are crucial for determining important chemical quantities such as association and dissociation constants, solubilities, binding energies, etc.

Generally speaking, the difference in free energy between two states A and B can be calculated as

$$\Delta G_{AB} = G_B - G_A = -\frac{1}{\beta} \ln \frac{Q_B}{Q_A}. \quad (6)$$

When using molecular simulation techniques to determine free energy differences, the selection of the Hamiltonian (typically potential energy, as mentioned earlier) and the sampling method used to estimate the relative free energy are the primary obstacles that need to be addressed. In practical applications the choice of Hamiltonian is often a compromise between accuracy and efficiency. While a quantum mechanical (QM) description of a system provides explicit modeling of electronic degrees of freedom, a classical approach treats each atom as a particle. The latter is referred to as force fields, which were previously explained (see section 2.1). The free energy difference between two states can be obtained through evolved statistical mechanical procedures [34]. It should be stressed that the free energy difference equals the energy difference at the point of intersection between the energy difference distributions of two states. [34] Thus, in order to obtain a reliable estimation of the free energy difference, it is crucial to sample all significant configurations of the two states to construct energy difference distributions that have converged at the intersection region. [145, 148, 149]. In practice, however, the energy difference distributions between two states do not always exhibit significant overlaps due to the existence of energy barriers or high-energy barriers separating the two states. Hence, enhanced sampling methods are necessary to bridge the gap between two states with rare events and to improve the sampling of the intermediate region. In the following subsection, one of the commonly used methods will be discussed.

### 2.2.2 *Potential of mean force and umbrella sampling*

To analyze and sample sizable systems, the concept of reaction coordinates (RCs), namely collective variables (CVs), is extremely useful. In general, these RCs represent low-dimensional projections of the phase space. The terms CVs and RCs are used interchangeably in this thesis. The RCs can be defined based on geometric factors such as the distance or angle between specific atoms, as well as the orientation of a molecule, or other collective variables that describe the process under investigation. For example, in the calculation of the free energy of binding or unbinding of a ligand to a protein, the RC could be the distance between the

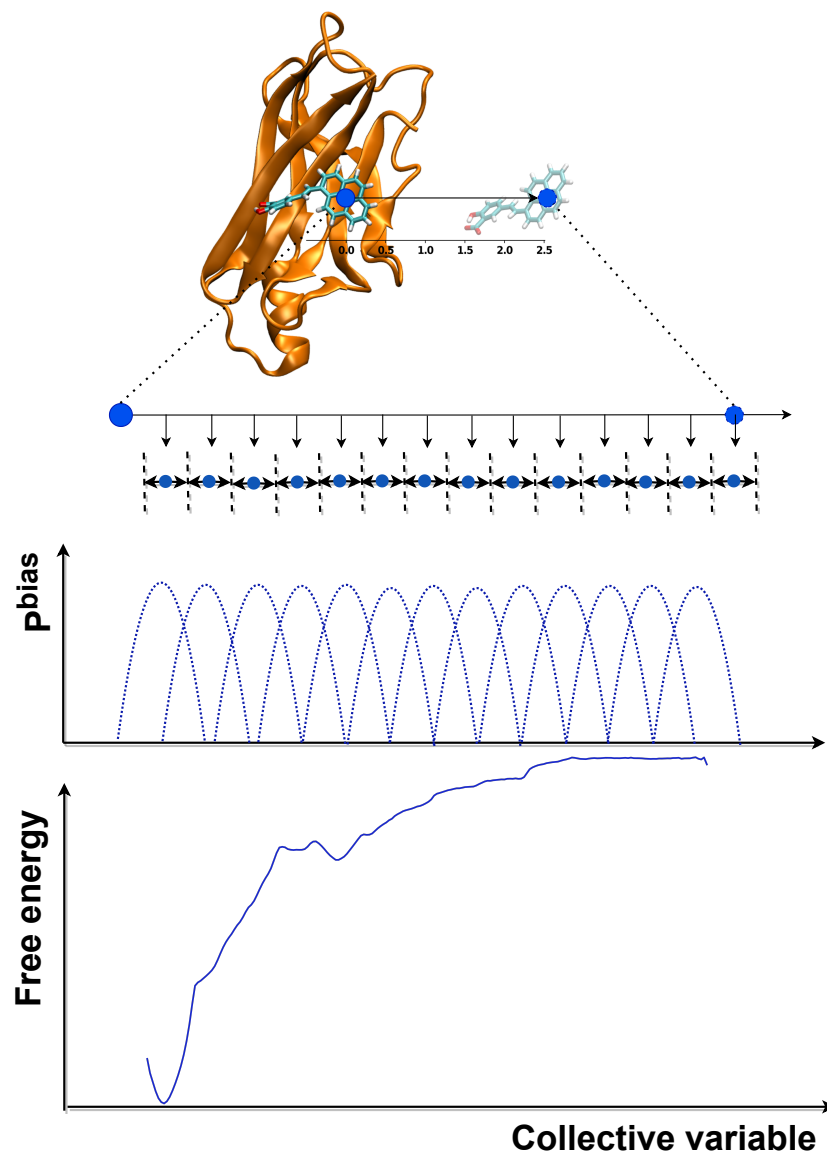
center of mass of the ligand and the protein pocket. The free energy surface along the chosen RCs  $\xi(\mathbf{q})$  is also known as the potential of mean force (PMF) [34, 150]. Here, only reaction coordinates  $\xi(\mathbf{q})$  depending on nuclear coordinates  $\mathbf{q}$  of the system will be considered. The PMF can be used to predict the behavior of a system over time and is an important concept in understanding the thermodynamics of a system [151]. Once  $\xi(\mathbf{q})$  is defined, the probability distribution of the system along  $\xi(\mathbf{q})$  can be determined by integrating out all degrees of freedom except for  $\xi(\mathbf{q})$ . Consequently, the PMF can be evaluated as

$$G(\xi) = \frac{1}{\beta} \ln \frac{\int \delta(\xi(\mathbf{q}) - \xi) \exp(-\beta V(\mathbf{q})) d^N \mathbf{q}}{\int \exp(-\beta V(\mathbf{q})) d^N \mathbf{q}}, \quad (7)$$

where  $\delta$  represents Dirac’s delta function. Calculating direct phase-space integrals as in equation 7 is not feasible in computer simulations due to the limited simulation time and resources. Nevertheless, if the system satisfies ergodicity, the ensemble average  $Q(\xi)$  can be approximated by the time average  $P(\xi)$  after an infinite sampling time. So, in principle, the free energy can be obtained directly from MD simulations by tracking the system’s distribution along the reaction coordinate,  $P(\xi)$ .

As previously stated, one of the challenges faced in MD simulations is the sampling and analysis of *rare* or *infrequent events*. Rare events are characterized by transitions between metastable states in phase space that require the system to overcome sizeable free energy barriers. The lower panel of Figure 7 represents a typical energy landscape of a complex system with many metastable states separated by high-energy barriers. As a consequence, the system is trapped in an energy well that is difficult to escape in normal MD simulations with an extensive timescale of milliseconds to microseconds. Several advanced sampling methods have been developed to accelerate the dynamic of complex systems in MD simulations allowing exploring these rare events. These methods can be categorized into two groups: constraint and unconstrained enhanced sampling, based on the application of constraints to the CVs [152]. In the first class of methods, to enhance the sampling of desired conformations, a bias potential based on selected CVs can be applied to the simulated system to constrain it into a specific CVs space. However, defining a suitable reaction coordinate for CVs-based methods is often challenging as it requires prior knowledge of the system being studied [153]. To address this issue, unconstrained enhanced sampling methods [154–157] are commonly used. An exhaustive description of all the methods belonging to these classes is beyond the scope of this work. Here I focus on the umbrella sampling (US), introduced by Torrie and co-workers [158], which is considered an important enhanced sampling approach due to its quick convergence and accurate estimation of the PMF [151, 157, 159].

In US simulations, harmonic bias potentials ( $\omega_i(\xi)$ ) [160] are generally used to restrain a system to a specific region of its conformational space  $i$  (commonly known as windows). It forces the ligand to sample the higher energy regions of



**Figure 7:** Schematic representation of umbrella sampling algorithm for computing the unbinding free energy landscape of TTR (in orange)–Py1SA (in cyan) complex along the chosen collective variable (CV). Upper panel: Simulation running (blue dot) for each simulating window (dashed lines) along the CV in umbrella sampling. Middle panel: Overlap of probability density between simulated windows. Lower panel: potential of mean force (PMF) extracted from umbrella sampling by the weighted histogram analysis method (WHAM).

the unbinding pathway that would otherwise not be sampled by an unconstrained simulation. Therefore, the bias potential ( $V^b$ ) can be written as [147]

$$V^b = V^u + \omega_i(\xi) = V^u + \frac{k}{2} (\xi_i - \xi_{0,i})^2, \quad (8)$$

where ( $V^u$ ) indicates a desired unbiased potential. In each window  $i$ , a harmonic bias function of strength  $k$  keeps the system close to the reference point  $\xi_{0,i}$  of the respective window. The free energy  $G^u(\xi)$  can be computed from the probability distribution of the system along  $\xi$

$$G^u(\xi) = -\frac{1}{\beta} \ln P^u(\xi) \quad (9)$$

with  $P^u(\xi)$  being the unbiased probability distribution along the reaction coordinate  $\xi$ . The unbiased distribution  $P^u(\xi)$  can be computed by integrating out all degrees of freedom but  $\xi$

$$P^u(\xi) = \frac{\int \delta(\xi(\mathbf{q}) - \xi) \exp(-\beta V^u(\mathbf{q})) d\mathbf{q}}{\int \exp(-\beta V^u(\mathbf{q})) d\mathbf{q}}. \quad (10)$$

By running the MD simulation for each window, the biased distribution along the reaction coordinate  $P_i^b(\xi)$  can be calculated in the same way as  $P^u(\xi)$  but adding the bias potential to equation 10

$$\begin{aligned} P_i^b(\xi) &= \frac{\int \delta(\xi(\mathbf{q}) - \xi) \exp(-\beta(V^u(\mathbf{q}) + \omega_i(\xi(\mathbf{q})))) d\mathbf{q}}{\int \exp(-\beta(V^u(\mathbf{q}) + \omega_i(\xi(\mathbf{q})))) d\mathbf{q}} \\ &= \exp(-\beta\omega_i(\xi)) \frac{\int \delta(\xi(\mathbf{q}) - \xi) \exp(-\beta V^u(\mathbf{q})) d\mathbf{q}}{\int \exp(-\beta(V^u(\mathbf{q}) + \omega_i(\xi(\mathbf{q})))) d\mathbf{q}} \end{aligned} \quad (11)$$

since the bias potential depends only on the reaction coordinate  $\xi$ . Using equation 10 and equation 11 to compute  $P_i^u(\xi)/P_i^b(\xi)$ , the  $P_i^u(\xi)$  can be related to the  $P_i^b(\xi)$  as

$$P_i^u(\xi) = P_i^b(\xi) \exp(\beta\omega_i(\xi)) \frac{\int \exp(-\beta\omega_i(\xi(\mathbf{q})) \exp(-\beta V^u(\mathbf{q})) d\mathbf{q}}{\int \exp(-\beta V^u(\mathbf{q})) d\mathbf{q}}. \quad (12)$$

The fraction is nothing else but the ensemble average of  $\exp(-\beta\omega_i(\xi(\mathbf{q})))$  denoted as  $\langle \exp(-\beta\omega_i(\xi(\mathbf{q}))) \rangle$ . Thus,

$$P_i^u(\xi) = P_i^b(\xi) \exp(\beta\omega_i(\xi)) \langle \exp(-\beta\omega_i(\xi)) \rangle. \quad (13)$$

Inserting equation 13 in equation 9 the unbiased free energy  $G_i^u(\xi)$  for each window can be evaluated as

$$G_i^u(\xi) = -\frac{1}{\beta} \ln(P_i^b(\xi)) - \omega_i(\xi) + F_i \quad (14)$$

with  $F_i = 1/\beta \ln \langle \exp(-\beta\omega_i(\xi)) \rangle$ . Unfortunately, the  $F_i$  cannot directly be obtained from sampling. Numerous methods have been proposed for the estimation of  $F_i$  [161–163]. Among those, the most popular postprocessing method is the

weighted histogram analysis method (WHAM) [164, 165]. In umbrella sampling, the histogram is a plot of the distribution of the sampled data for each value of the reaction coordinate. Typically, the histogram is constructed by counting the number of samples that fall within a certain range of the reaction coordinate and then normalizing the counts by the total number of samples and the width of the bin. The resulting normalized histogram can be used to obtain the probability density function of the reaction coordinate. The idea of WHAM is to estimate the statistical error of the global unbiased distribution  $P^u(\xi)$  given the umbrella histograms, and subsequently to compute the PMF that corresponds to the smallest uncertainty. The global distribution  $P^u(\xi)$  is obtained by combining the distributions of each window using weighted averaging, as follows,

$$P^u(\xi) = \sum_i^{N_w} \rho_i(\xi) P_i^u(\xi) \quad (15)$$

with  $N_w$  being the total number of windows, and  $\rho_i$  being the weights, which are chosen in order to minimize the statistical error of  $P^u(\xi)$ . This leads to

$$\rho_i(\xi) = \frac{a_i(\xi)}{\sum_j^{N_w} a_j} \quad (16)$$

with

$$a_i(\xi) = N_i \exp[-\beta\omega_i(\xi) + \beta F_i], \quad (17)$$

where  $N_i$  is the total number of steps sampled for each window  $i$ . From the definition of  $F_i$

$$\beta F_i = \ln\langle \exp(-\beta\omega_i(\xi)) \rangle, \quad (18)$$

thus

$$\begin{aligned} \exp(\beta F_i) &= \langle \exp[-\beta\omega_i(\xi)] \rangle \\ &= \int P^u(\xi) \exp[-\beta\omega_i(\xi)] d\xi. \end{aligned} \quad (19)$$

Hence,  $P^u$  and  $F_i$  have to be computed iteratively until convergence. The convergence of the WHAM equations depends on the number of histograms and the height of the barriers in the PMF, and can range from tens to tens of thousands of iterations [165].

The computation of free energy using enhanced sampling methods, particularly US, depends on the identification of an appropriate set of collective variables. In this method, a potential bias is applied along one or a few predefined degrees of freedom (CVs), thus they should appropriately describe the process under investigation. Not surprisingly, much effort has been devoted to the identification and improvement of useful CVs. [166–170] In simulations of ligand unbinding in complex systems, a few CVs are used, and PMF reconstruction becomes slower as the number of CVs increases. Therefore, other enhanced sampling methods like metadynamics and adaptive biasing force (ABF) [171] are used for exploring the free energy space using multiple CVs, returning a multidimensional PMF. Furthermore,

although the concept of US is relatively straightforward, its effective implementation requires a reliable method to combine the samples obtained from different umbrellas into a cohesive set. Particularly, in the WHAM method, it is related directly to the overlap of the probability of distribution along the chosen CVs. Thus, the choice of window force constant and distance between simulations in the CV space are crucial to ensure adequate overlap between neighboring windows and uniform sampling [147, 151]. A large force constant  $k$  is ideal to drive the system over high-energy barriers, but it can cause narrow distributions that require additional windows for efficient sampling. Besides the large force constant, the large time step can overestimate the configurations with high energies [172]. Despite these limitations, US has achieved success due to its quick convergence and ease of expansion, particularly for computing the PMF of noncovalent ligand–receptor association/dissociation processes. It has been utilized to determine the binding strength of different receptor–ligand systems, which vary from small chemical molecular systems [173] to big biological systems [174–177].

## 2.3 QUANTUM MECHANICS

In the previous section, we have seen that in MD, the behavior of molecules is modeled using classical mechanics, which is based on Newton’s laws of motion. However, this approach ignores the quantum–mechanical nature of electrons, which play a crucial role in chemical processes like photochemical reactions. Quantum mechanics (QM) can accurately model the behavior of electrons and their interactions with nuclei, allowing for a more accurate simulation of chemical processes at the microscopic scale of atoms and molecules. Within this section, I will explore the foundations of quantum mechanical description of a single molecular system in vacuum. Subsequently, its interaction with external perturbations such as light will be revised.

### 2.3.1 *Electronic structure theory*

#### 2.3.1.1 *Basic quantum chemistry*

The starting point for a non-relativistic quantum mechanical description of a system is the time dependent (TD) Schrödinger equation (SE) [178]. It reads, in atomic units,

$$\hat{H}\Psi = i\frac{\partial}{\partial t}\Psi, \quad (20)$$

where the TD wave function  $\Psi$  contains all information of the system, and the Hamilton operator  $\hat{H}$  represents its total energy. For describing a stationary state, where the Hamiltonian is time-independent (TI), the SE is also TI, which takes the form

$$\hat{H}(\mathbf{r},\mathbf{R})\Psi(\mathbf{r},\mathbf{R}) = E^t\Psi(\mathbf{r},\mathbf{R}). \quad (21)$$

The eigenvalue  $E^t$  represents the total energy of the stationary state corresponding to the wave function  $\Psi(\mathbf{r},\mathbf{R})$  of the system at the positions of the electrons ( $\mathbf{r}$ ) and nuclei ( $\mathbf{R}$ ).

The Hamilton operator is the sum of kinetic ( $\hat{T}$ ) and potential ( $\hat{V}$ ) energy operator of the system. They can be further expanded into the kinetic energy operators of the nuclei ( $\hat{T}_N(\mathbf{R})$ ) and electrons ( $\hat{T}_e(\mathbf{r})$ ), and the potential energy operators of the nuclear–nuclear ( $\hat{V}_{NN}(\mathbf{R})$ ) and electron–electron repulsion ( $\hat{V}_{ee}(\mathbf{r})$ ) and nuclear–electron attraction ( $\hat{V}_{eN}(\mathbf{r},\mathbf{R})$ ), respectively. So that the Hamiltonian becomes

$$\begin{aligned} \hat{H}(\mathbf{r},\mathbf{R}) &= \hat{T}_N(\mathbf{R}) + \hat{T}_e(\mathbf{r}) + \hat{V}_{ee}(\mathbf{r}) + \hat{V}_{NN}(\mathbf{R}) + \hat{V}_{eN}(\mathbf{r},\mathbf{R}) \\ &= \hat{T}_N(\mathbf{R}) + \hat{T}_e(\mathbf{r}) + \hat{V}(\mathbf{r},\mathbf{R}). \end{aligned} \quad (22)$$

While the SE can be solved analytically for simple systems such as an isolated hydrogen atom, it, however, is not possible to find an exact solution for more complex systems [34, 179]. It is worth noting that the  $\hat{V}_{eN}(\mathbf{r},\mathbf{R})$  term represents the coupling between the electronic and nuclear degree of freedom. Thus,

simply neglecting this term does not appear promising. Therefore, further approximations and/or numerical solutions need to be considered. One of the most frequently applied approximations is the so-called Born–Oppenheimer approximation (BOA) [98] which disentangles the nuclear degrees of freedom from the electronic ones.

Due to the large difference in mass between nuclei and electrons, the nuclei are nearly fixed with respect to electronic motion leading to  $\hat{\mathbf{T}}_{\mathbf{N}}(\mathbf{R}) = 0$ . Thus, for a fixed nuclear configuration, a “clamped-nuclei” SE is obtained

$$(\hat{\mathbf{T}}_e(\mathbf{r}) + \hat{\mathbf{V}}(\mathbf{r}, \mathbf{R}))\Psi_e^I(\mathbf{r}; \mathbf{R}) = E_e^I(\mathbf{R})\Psi_e^I(\mathbf{r}; \mathbf{R}), \quad (23)$$

where  $I$  is a label denoting the electronic states. Here, the electronic wave function  $\Psi_e^I(\mathbf{r}; \mathbf{R})$  depends explicitly on the electronic coordinates and parametrically on the nuclear coordinates (denoted by the  $;$  symbol) since the nuclear coordinates are external parameters to the electronic problem.  $E_e(\mathbf{R})$  is the electronic energy for the nuclear coordinates  $\mathbf{R}$ . Since  $\hat{\mathbf{V}}_{\mathbf{N}\mathbf{N}}(\mathbf{R})$  is constant, it is often neglected in the above equation.

Expanding the total wave functions in an orthonormal set of electronic wave functions for each nuclear configuration, the exact solution of the TI SE (equation 21) can be written as

$$\Psi(\mathbf{r}, \mathbf{R}) = \sum_{\mathbf{I}} \Psi_{\mathbf{N}}^{\mathbf{I}}(\mathbf{R})\Psi_e^{\mathbf{I}}(\mathbf{r}; \mathbf{R}). \quad (24)$$

Here, the nuclear wave functions  $\Psi_{\mathbf{N}}^{\mathbf{I}}(\mathbf{R})$  are considered to be expansion coefficients. Inserting equation 24 to the TI SE (equation 21), and multiplying to the left with the electronic wave functions of state  $J$  then integrating over the electronic coordinates gives

$$(\hat{\mathbf{T}}_{\mathbf{N}}(\mathbf{R}) + E_e^J(\mathbf{R}) + \langle \Psi_e^J(\mathbf{r}; \mathbf{R}) | \hat{\mathbf{T}}_{\mathbf{N}}(\mathbf{R}) | \Psi_e^J(\mathbf{r}; \mathbf{R}) \rangle)\Psi_{\mathbf{N}}^J(\mathbf{R}) = E^t \Psi_{\mathbf{N}}^J(\mathbf{R}). \quad (25)$$

The electronic energy  $E_e^J(\mathbf{R})$  is a function of the nuclear geometry. To simplify the description, the coupling  $\langle \Psi_e^J(\mathbf{r}; \mathbf{R}) | \hat{\mathbf{T}}_{\mathbf{N}}(\mathbf{R}) | \Psi_e^J(\mathbf{r}; \mathbf{R}) \rangle$  term can be neglected due to the significant difference in mass between the electrons and the nuclei

$$(\hat{\mathbf{T}}_{\mathbf{N}}(\mathbf{R}) + E_e^J(\mathbf{R}))\Psi_{\mathbf{N}}^J(\mathbf{R}) = E^t \Psi_{\mathbf{N}}^J(\mathbf{R}). \quad (26)$$

In the BOA, the total wave function can be expressed as a product of an electronic wave function satisfying equation 23 and a nuclear wave function satisfying equation 26

$$\Psi^{\text{BOA}}(\mathbf{r}, \mathbf{R}) = \Psi_{\mathbf{N}}^J(\mathbf{R})\Psi_e^J(\mathbf{r}; \mathbf{R}). \quad (27)$$

$E_e^J(\mathbf{R})$  is the energy that is usually called the potential energy surface (PES) providing a map of the energy of the system as a function of the positions of the nuclei. To simplify the notation, the electronic Hamiltonian, wave functions and energies are hereafter denoted as  $\hat{\mathbf{H}}$ ,  $\Psi$ , and  $E$ , respectively.

BOA approximation greatly simplifies the problem and allows for the use of a number of efficient computational methods for electronic structure calculations.



In general the BOA is a good approximation for describing phenomena occurring where the electronic states that are very well separated. However, it is important to note that this approximation is not always valid, especially for systems where different electronic states are close and their coupling is appreciable [180]. Through the separation of the electronic and nuclear degrees of freedom, the BOA establishes the distinct subdisciplines within theoretical chemistry.

### 2.3.1.2 *Computation of the ground state*

Built upon the mean-field theory, Hartree–Fock (HF) theory is one of the simplest methods to find an approximate solution of the electronic SE. The HF theory [181–183] simplifies the description of the interactions of  $N$  electrons by considering just one electron interacting with the mean field of the  $N - 1$  remaining electrons. As a result, the many-body (many-electron) problem reduces to an effective one-body (one-electron) problem. The wave function in HF method is represented as a linear combination of a set of basis functions (so-called linear combination of atomic orbitals (LCAO)). With the variational principle, under the constraint of orthonormality, the orbitals can be optimized till the entire system reaches its energy minimum. The coefficients of the linear combination are obtained by solving the HF equations self-consistently. In practice, the HF method can capture the vast majority of the total energy (with a large basis set) [184]. However, it is incapable of providing a robust description of reactive chemical events in which electron correlation has a crucial role. To overcome this limitation, more sophisticated *ab initio* methods such as configuration interaction (CI) [185, 186], coupled cluster (CC) [187], or perturbation-based methods (e.g., Møller–Plesset perturbation theory [188]) have been developed, which are able to provide much more accurate solutions for molecular systems. Each of these correlated wave function-based approaches has a different computational scaling with the number of electrons and can be the method of choice for different types of problem. [184, 189]

Unfortunately, standard implementations of correlated wave function methods are not practical for calculations on large molecular systems consisting of hundreds of atoms. The basic problem here is the enormous complexity of the many-electron wave function ( $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ , where  $\mathbf{r}_i$  specifies the location of electron  $i$  using for example three Cartesian coordinates or spherical polar coordinates). Since a molecular orbital (MO) ( $\phi_i(\mathbf{r})$ ) is often expressed as a LCAO or basis functions, the cost of a quantum chemical calculation is determined by the number of basis functions rather than by the number of atoms. The larger the number of basis functions, the more accurate the description of the electronic structure of the molecule, but also the higher the computational cost of the calculation. In addition, the accuracy of the results in MO calculations is also determined by the type of basis functions used, in addition to the size of the basis set. Two types of functions are commonly used to describe atomic orbitals: Slater type orbitals (STOs) [190] and Gaussian type orbitals (GTOs) [191]. Different basis

sets are available with varying accuracy and size, and the choice of the appropriate basis set for a particular calculation depends on the size and complexity of the system, as well as the desired level of accuracy. The simplest basis sets are termed “minimal basis sets”, where only sufficient functions are used to contain all the electrons (core through valence) of the neutral atoms. One method of improving on a minimal basis set is to double, triple, quadruple, and so forth, the number of functions. Such basis sets are known as double- $\zeta$  (DZ), triple- $\zeta$  (TZ), quadruple- $\zeta$  (QZ), respectively. In practice, based on the fact that valence electrons are more chemically important, it is sensible to treat core and valence basis functions differently, where core electrons are still described by a minimal basis set while the valence electrons are treated with a larger basis set called split-valence. Two of the most commonly used split-valence basis set families are the Pople (e.g. 6-31G [192]) and Karlsruhe (e.g. def2-TZV [193]) basis set. To enhance the quality of a basis set beyond increasing the  $\zeta$ , additional higher angular momentum functions can be added to the basis set. For instance, polarization is a particularly important effect in optical spectra calculations. This phenomenon occurs when atoms come close to each other to form chemical bonds or when a molecule is placed in an external electric field, which results in a distortion of the shape of the atomic orbitals and a change in the charge distribution. To account for the required flexibility of atomic orbitals to shift to one side or the other during chemical bonding, polarization functions can be added to the Pople and Karlsruhe basis sets leading to e.g. 6-31G(d,p) [194] and def2-TZVP [195], respectively. For systems that contain loosely bound electrons, such as anionic systems or excited states, high accuracy in property calculations necessitates the inclusion of diffuse functions. For the Pople basis sets, diffuse functions are added on non-hydrogen atoms to improve their accuracy (indicated by + or ++). It is worth noting that the def2-TZ basis sets are slightly more diffuse than unaugmented Pople sets [196], which makes them robust and efficient in large-scale HF and DFT calculations, with an excellent cost-to-performance ratio [197]. This feature has led many computational studies to use them as their preferred method.

Various approximate methods have been developed that simplify the wave function while still retaining its essential features. One such method is density functional theory (DFT), which expresses the energy of the system in terms of the electronic density rather than the full wave function. By using the electron density ( $\rho(\mathbf{r})$ ), the problem has changed from a three-N-dimensional equation to N separate three-dimensional ones. The basis of modern DFT relies on two fundamental theorems formulated by Hohenberg and Kohn in 1964 [198]. According to the first theorem, the ground-state (GS) electron density determines the electronic wave function, which in turn determines all GS properties of an electronic system. In particular, the exact energy can then be written as a *functional* of the exact GS density

$$E_e[\rho] = F_{\text{HK}}[\rho] + \int \rho(\mathbf{r})v_{\text{ext}}(\mathbf{r})d\mathbf{r}, \quad (28)$$

where  $v_{\text{ext}}(\mathbf{r})$  is the external potential uniquely defined by the electronic density.  $F_{\text{HK}}[\rho]$  is the *universal functional* of Hohenberg–Kohn (HK) given by the sum of

the kinetic ( $\hat{T}_e[\rho]$ ) and electron-electron repulsion energies ( $\hat{V}_{ee}[\rho]$ ) of an  $N$  electron system. Unfortunately, they do not specify any analytical form to calculate the properties from the electron density. In the second theorem, they established a variational principle for the electron density. It states that for an exact functional any trial density,  $\tilde{\rho}$ , yields an higher energy than the true electronic energy

$$\langle \Psi | \hat{H} | \Psi \rangle = F_{\text{HK}}[\tilde{\rho}] + \int \tilde{\rho}(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} = E_e[\tilde{\rho}] \geq E_e[\rho]. \quad (29)$$

Similar as in wave function based methods, the variational principle leads to a criterion for the search of a good approximation to the electron density of the GS. It should be noted that the Hamiltonian operator and all properties of the GS are determined by the associated electron density. This implies that the density possesses *pure-state  $v$ -representability*. An electron density is considered pure-state  $v$ -representable when it corresponds to a unique antisymmetric wave function of the Hamiltonian,  $\hat{H}[N, v]$ . The general conditions for  $v$ -representability are still unknown leading to a difficult side task. This can be circumvented using the Levy–Lieb (LL) constrained-search formalism [199, 200],  $F_{\text{LL}}[\rho] = \langle \tilde{\Psi} | \hat{T}_e + \hat{V}_{ee} | \tilde{\Psi} \rangle$ , which states to search over all wave functions (or ensemble density matrices) that produce the trial density. In this regard, the HK functional is not unknown, but it is too complicated to be practical and so approximations becomes necessary. The classic historical example for such approximations is Thomas–Fermi (TF) theory. [201, 202] In this theory, the kinetic energy,  $T_e[\rho]$ , is given by a local functional of the electron density, which is exact for a uniform gas of non-interacting electrons. While TF functional reasonably atomic energies, it is unable to describe important chemical phenomena such as covalent bonds. [203] The primary drawback of this theory lies in its treatment of electron dynamics within the functional, leading to a substantial error in the estimation of kinetic energy.

A practical solution for this problem was introduced by Kohn and Sham [204]. The main concept behind Kohn–Sham (KS) theory is to identify a *non-interacting* reference system that can be described by a single Slater determinant, which yields the same electron density as the interacting system. This non-interacting system would have the same GS energy and properties as the interacting system. The key advantage of the KS approach is that the total kinetic energy of the non-interacting reference system (labeled by the index  $s$ ) can be partitioned into a known part  $T_s[\rho]$  and an unknown correction term  $T_e[\rho] - T_s[\rho]$ . The accuracy of the single-determinant approximation can be improved by minimizing the correction term, which becomes smaller as the system is better described. Similarly, the electron–electron interaction term ( $V_{ee}[\rho]$ ) is divided into a precisely known Coulombic component ( $J[\rho]$ ), which can be formulated in relation to the electron density, and a correction term that involves the exchange and correlation contributions ( $V_{ee}[\rho] - J[\rho]$ ). Thus, in the KS approach, the universal functional ( $F_{\text{KS}}[\rho]$ ) is defined as

$$F_{\text{KS}}[\rho] = T_s[\rho] + J[\rho] + E_{\text{xc}}[\rho]. \quad (30)$$

The only unknown term is the exchange–correlation (XC) functional ( $E_{xc}[\rho]$ ), which reads

$$E_{xc}[\rho] = T_{el}[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho]. \quad (31)$$

Since the analytic form of the XC functional remains unknown, it has to be approximated. It should be stressed that the main advantage of KS theory compared to HK and LL schemes is that the energy contribution to be approximated is rather small. Therefore, the overall energy can often be reasonably well described with available approximations for the XC functional. In addition, it might be worthwhile to note that the exchange energy is typically about one order of magnitude larger than the correlation term. Hence, getting a reasonable approximation for the exchange term is often more important than for the latter part.

Over the years, numerous XC functionals have been developed based on theoretical reasoning and/or by adjusting parameters to fit experimental or high-level ab initio data. They can be classified into different categories based on the components used in constructing the XC functionals. These categories are known as the local density approximation (LDA), local generalized gradient approximation (GGA), and meta-GGA, hybrid functionals, and range-separated hybrid functionals. The basic idea is that by gradually incorporating additional information concerning the local density, reduced gradient, local kinetic energy, and other relevant quantities, users gain increased flexibility to construct more accurate functionals. It must be emphasized that, although the inclusion of additional variables into the functional allows for more flexibility, the systematic improvement of accuracy is not guaranteed. [205] Therefore, the selection of the XC functional for a specific problem is somewhat based on experience and benchmarking of the DFT.

### 2.3.1.3 *Computation of excited states*

So far, the methods discussed have primarily focused on the electronic GS. However, many physicochemical processes, such as photochemistry, fluorescence, etc., are initiated by or require excited-state (ES) species. In spectroscopy measurements, a sample is exposed to an electronic field, which interacts with the molecules. [206] This interaction can lead to the absorption of photons by molecules in their initial state, which in turn can excite electrons to higher electronic states. In the absence of an electromagnetic field, the system decays from excited state by emitting an photon. Theoretical treatments of ESs have been proposed for various physicochemical processes, but accurately computing excitation energies and optical spectra can be a challenging task. [207] Firstly, most ESs are open-shell states, particularly those involving excitation to valence-like orbitals, which inherently require a multiconfigurational treatment. Secondly, achieving convergence of a self-consistent field (SCF) solution for an ES can be difficult due to the complex nature of ES. The energy difference between the ES and the GS SCF energy is often approximated as  $\Delta\text{SCF}$ . [207] To address these challenges, e.g. within the wave function theory (WFT), the multiconfiguration self-consistent field (MC-

SCF) [208]-based methods have proven to be effective. These methods incorporate CI steps that directly extend to excited states.

Another prominent approach for ES calculations is time-dependent density functional theory (TD-DFT), which is an extension of GS KS DFT to describe electronic excitations. The formal foundation for TD-DFT is provided by the Runge-Gross theorem [209], which is an analog of the HK [198] theorem for the TD version of the SE (equation 20). They demonstrated the one-to-one mapping between TD potentials and TD electronic density functionals [209, 210]. In practical terms, this means, that if one can calculate the TD density of a system, then the TD potential and all of its associated properties can be obtained using TD-DFT. Similarly to the case of TI DFT, by replacing the interacting system with an auxiliary non-interacting system (KS system) [211] the TD formalism reduces the many-electron problem to a set of self-consistent single particle equations [212, 213]

$$i\frac{\partial}{\partial t}\psi_i(\mathbf{r}, t) = \left( \frac{-1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r}, t) \right) \psi_i(\mathbf{r}, t), \quad (32)$$

where the TD effective potential  $v_{\text{eff}}(\mathbf{r}, t)$  can be written as

$$v_{\text{eff}}(\mathbf{r}, t) = v_{\text{SCF}}[\rho](\mathbf{r}, t) + v(t). \quad (33)$$

Here,  $v(t)$  refers to a perturbation of an applied field that is gradually turned on in the distant past. The SCF  $v_{\text{SCF}}[\rho](\mathbf{r}, t)$  is defined as

$$v_{\text{SCF}}[\rho](\mathbf{r}, t) = \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}[\rho](\mathbf{r}, t), \quad (34)$$

where the XC potential is given as the functional derivative of the XC action  $A_{\text{xc}}$  and is approximated by  $E_{\text{xc}}$ , the XC functional of TI KS theory, which is a function of space at a fixed time  $t$

$$v_{\text{xc}}[\rho](\mathbf{r}, t) = \frac{\partial A_{\text{xc}}[\rho]}{\partial \rho(\mathbf{r}, t)} \approx \frac{\partial E_{\text{xc}}[\rho_t]}{\partial \rho_t(\mathbf{r})} = v_{\text{xc}}[\rho_t](\mathbf{r}). \quad (35)$$

Similar to the XC potential in GS DFT, the XC kernel in TD-DFT is also unknown thus the choice of XC functionals can significantly impact the results obtained with TD-DFT. Despite the improvements made for the XC functionals in TD-DFT, solving the full TD KS equation for even moderately sized systems can still be computationally expensive. Therefore, several approximations are made to simplify the calculations. Noteworthy, TD-DFT is more complicated than in GS DFT as it is formally influenced by the entire past density of the system, leading to a memory effect, and must be solved self-consistently in the time domain. Thus, one of the most common approximations is the adiabatic approximation, which neglects the time-memory effect on the XC kernel.

Molecular properties are essential quantities underlying the macroscopic behavior of molecular systems. The determination of these properties constitutes one of the most fruitful areas of interplay between experiment and theory. From a theoretical point of view, the determination of molecular properties can be achieved

by calculating response functions that describe how a molecule responds to weak perturbing fields like internal magnetic moments or external electric and magnetic fields. [214] This means in circumstances where the external TD potential is small, it may not be necessary to solve the full TD KS equations. Instead, perturbation theory can be employed to determine the behavior of the system and the linear response (LR) of the system can be taken as an approximation to the full solution. The resulting theory is known as LR TD-DFT([215–217]), and often simply denoted as TD-DFT. In this approach, when a system is initially in the GS, the impact of a perturbation on the KS Hamiltonian due to the application of a field is considered to be of first order as

$$\delta v_{\text{eff}}(\mathbf{r}, t) = \delta v_{\text{SCF}}[\rho](\mathbf{r}, t) + \delta v(t), \quad (36)$$

where  $\delta v_{\text{SCF}}[\rho](\mathbf{r}, t)$  is the LR of the SCF due to the change in the charge density given by

$$\delta \rho(\mathbf{r}, \omega) = \sum_{\mathbf{a}\mathbf{i}} \delta P_{\mathbf{a}\mathbf{i}}(\omega) \psi_{\mathbf{a}}(\mathbf{r}) \psi_{\mathbf{i}}^*(\mathbf{r}) + \sum_{\mathbf{i}\mathbf{a}} \delta P_{\mathbf{i}\mathbf{a}}(\omega) \psi_{\mathbf{i}}(\mathbf{r}) \psi_{\mathbf{a}}^*(\mathbf{r}), \quad (37)$$

transforming to the frequency representation. In this relation, following the convention of labeling MOs as  $\mathbf{i}, \mathbf{j}$  for occupied,  $\mathbf{a}, \mathbf{b}$  for virtual, and  $\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v}$  for general orbitals,  $\delta P_{\mathbf{s}\mathbf{t}}(\omega)$  is the LR of the KS density matrix in the basis of the unperturbed MOs

$$\delta P_{\mathbf{s}\mathbf{t}}(\omega) = \frac{\Delta n_{\mathbf{s}\mathbf{t}}}{(\epsilon_{\mathbf{s}} - \epsilon_{\mathbf{t}}) - \omega} \delta v_{\mathbf{s}\mathbf{t}}^{\text{eff}}(\omega), \quad (38)$$

where  $\Delta n_{\mathbf{s}\mathbf{t}}$  is the difference in occupation numbers and has a value of 1 for  $\mathbf{s}\mathbf{t} = \mathbf{a}\mathbf{i}$  and -1 for  $\mathbf{s}\mathbf{t} = \mathbf{i}\mathbf{a}$ . Equation 38 is, however, complicated by the fact that  $\delta v_{\text{SCF}}[\rho]$  depends on the response of the density matrix

$$\delta v_{\mathbf{s}\mathbf{t}}^{\text{SCF}}[\rho](\omega) = \sum_{\mathbf{u}\mathbf{v}} K_{\mathbf{s}\mathbf{t},\mathbf{u}\mathbf{v}}(\omega) P_{\mathbf{u}\mathbf{v}}(\omega), \quad (39)$$

where the coupling matrix  $K$  describes the response of the SCF to changes in the charge density

$$\begin{aligned} K_{\mathbf{s}\mathbf{t}\sigma,\mathbf{u}\mathbf{v}\tau} &= (\psi_{\mathbf{s}\sigma}^*(\mathbf{r}) \psi_{\mathbf{t}\sigma}(\mathbf{r}) | \psi_{\mathbf{v}\tau}^*(\mathbf{r}') \psi_{\mathbf{u}\tau}(\mathbf{r}')) \\ &+ \int \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}' \psi_{\mathbf{s}\sigma}^*(\mathbf{r}) \psi_{\mathbf{t}\sigma}(\mathbf{r}) \frac{\delta^2 E_{\text{xc}}[\rho]}{\delta \rho_{\sigma}(\mathbf{r}) \delta \rho_{\tau}(\mathbf{r}')} \times \psi_{\mathbf{v}\tau}^*(\mathbf{r}') \psi_{\mathbf{u}\tau}(\mathbf{r}') \end{aligned} \quad (40)$$

with  $\sigma$  and  $\tau$  being spin indices. Within the adiabatic approximation, this matrix is time and frequency-independent.

It is important to note, that the determination of excitation energies in response theory involves identifying the poles of the response functions. They can be determined as a solution to the non-Hermitian eigenvalue problem

$$\begin{pmatrix} A & B \\ B & A \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = w \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (41)$$

where  $X_{ai} = \delta P_{ai}(\omega)$  and  $Y_{ai} = \delta P_{ia}(\omega)$ . Here, the matrices  $A$  and  $B$  are defined as

$$A_{ai,bj} = \delta_{ab}\delta_{ij}(\epsilon_a - \epsilon_i) + K_{ai,bj}, \quad (42)$$

$$B_{ai,bj} = K_{ai,jb}. \quad (43)$$

The particular XC functional employed determines the specific structures of  $A$  and  $B$  [218].

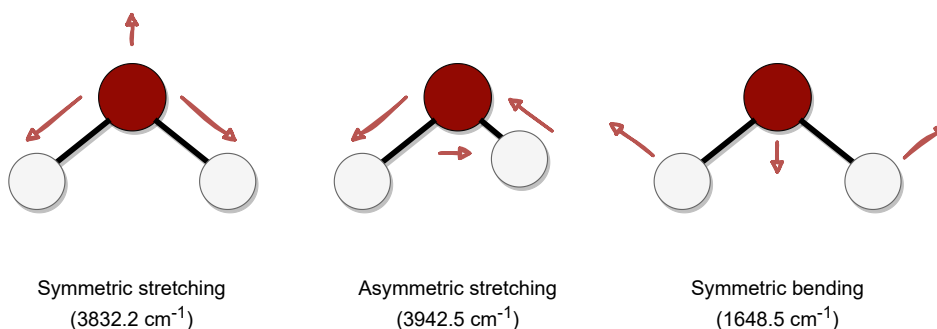
Due to the favorable cost-performance ratio, TD-DFT in its LR form has become the *de facto* workhorse in the calculation of ES-related properties like electronic transition energies, absorption and emission spectra in a wide range of fields. [177, 210, 219] However, TD-DFT has its own limitations, particularly in dealing with near degeneracies and double excitation [220, 221]. Similar to the DFT methods, the accuracy of TD-DFT results is sensitive to the choice of XC functional. It is worth noting that besides the lack of a systematical way of improving the method, one of the widely prominent limitations of DFT (and also TD-DFT) is the self-interaction error (SIE) [222–224]. In WFT methods, the self-interaction term included by the Coulomb potential is exactly cancelled out by a similar term in the exchange potential. This is, however, not the case for the common XC functionals since the exchange term is treated only in an approximate way. The resulting SIE leads to an over-delocation of the electron density compared to the true density. Therefore, the conventional functionals often significantly underestimate charge-transfer [220, 225], Rydberg [226, 227] and core excitation energy [228]. A common way to mitigate this issue is the application of hybrid functionals, which use a certain amount of exact-exchange (EXX) instead of DFT exchange. One should be aware that the optimal fraction of EXX may differ significantly for different properties and systems. In recent years, various strategies have been developed to improve upon the excitation energies. The currently most popular class is the range-separated hybrid functionals. [223] The concept of range separation is based on dividing the electron-electron interaction into short and long-range components, allowing for the use of different fractions of HF-like exchange at varying inter-electronic distances. To date, various range-separated hybrid functionals have been developed. Among those, the coulomb attenuating methods (CAMs) like CAM-B3LYP [229] couple global hybrids at short interelectronic distances (19%) and increase the HF-like exchange at longer distances (65%) resulting in an accurate description of electronic excitations and charge-transfer transitions [230].

For more comprehensive information, one can refer to textbooks on theoretical and computational chemistry [179, 184, 231], as well as a series of reviews [210, 216, 232, 233] treating both mathematical aspects and applications of DFT/TD-DFT.

## 2.3.2 Vibrational structure theory

### 2.3.2.1 Molecular vibrations and representation of potential energy surfaces

Molecular vibrations refer to the motion of atoms within a molecule, where the atoms oscillate back and forth around their equilibrium positions. In molecular physics, each independent mode of motion is referred to as a degree of freedom (DOF), which represents a unique way in which the energy of the molecule can be distributed. Denoting the set of coordinates describing the system as  $\mathbf{q} = (q_1, q_2, \dots, q_M)$  where  $M$  is the number of DOFs in a system, a molecule containing  $N$  atoms has  $3N$  DOFs to describe the positions of the atomic nuclei. Typically, three DOFs are assigned for translation and three (or two for linear molecules) for rotation, resulting in  $3N - 6(5)$  DOFs describing the vibrational motions of the molecule. For example, water molecules display three fundamental vibrational modes: symmetric stretching, asymmetric stretching, and bending (as illustrated in Figure 8). Each mode is linked to a different fundamental frequency of oscillation, which can be experimentally observed through techniques such as infrared spectroscopy.



**Figure 8:** Illustrations of the three vibrational modes of the water molecule: symmetric stretching, asymmetric stretching, and bending with their corresponding frequency [234]. The atoms move in the directions indicated by arrows.

As discussed previously, in the framework of BOA, the system's wave function can be expressed as the product of an electronic and nuclear component. The PES represents a potential, where each point on the surface corresponds to the electronic energy of the system at a specific arrangement of the nuclei. The vibrational states can then be determined by solving the nuclear Schrödinger equation with the Hamiltonian for nuclear motion

$$\hat{H}_N = \hat{T}_N + \hat{V}_N. \quad (44)$$

The specific form of the kinetic energy operator  $\hat{T}$  and potential energy surface  $\hat{V}$  are dependent on the chosen coordinates and the approach pursuit to obtain  $V(\mathbf{q})$  [235]. The most standard choice of vibrational coordinates for medium-sized



systems is normal coordinates defined from the eigenvectors of the mass-weighted second derivative matrix (Hessian matrix) at the minimum on the PES.

On the basis of the mass-scaled normal coordinates, the nuclear Hamiltonian of a non-linear molecule in equation 44 can be written as

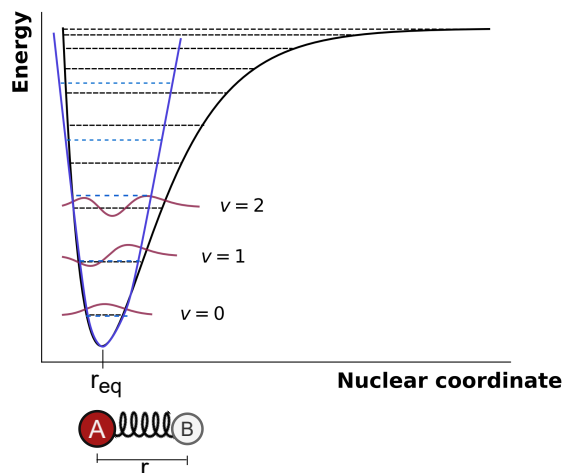
$$\hat{H}_N = -\frac{1}{2} \sum_{m=1}^M \frac{\partial^2}{\partial q_m^2} + \frac{1}{2} \sum_{\alpha\beta} (\hat{J}_\alpha - \hat{\pi}_\alpha) \mu_{\alpha\beta} (\hat{J}_\beta - \hat{\pi}_\beta) - \frac{1}{8} \sum_{\alpha} \mu_{\alpha\alpha} + \hat{V}(\mathbf{q}). \quad (45)$$

In this Hamiltonian, or so-called Watson operator [236],  $\hat{J}_\alpha$  and  $\hat{\pi}_\alpha$  is the total and the vibrational angular momenta operator (respectively),  $\mu_{\alpha\beta}$  is the inverse of the effective moment of inertia, and  $V(\mathbf{q})$  is the usual Born–Oppenheimer PES. When dealing with systems that have more than a few DOFs, approximations must be made for both, the operator and the wave function, in practical calculations. Considering the Watson for a non-rotating molecule and neglecting the terms involving the inverse effective moment of inertia, the simpler Hamiltonian is obtained as

$$\hat{H}_N = -\frac{1}{2} \sum_{m=1}^M \frac{\partial^2}{\partial q_m^2} + \hat{V}(\mathbf{q}). \quad (46)$$

In addition, the proper representation of the PES,  $V(\mathbf{q})$ , is critical for accurate vibrational structure calculations. However, constructing the PES poses a significant challenge for many-body systems due to its high dimensionality and the lack of an analytical form. In practice, the vibrational motions of large molecules are often analyzed using normal modes assuming a simplified quadratic PES with respect to atomic displacements [237, 238]. In this model, the potential is represented as a harmonic oscillator (HO), which assumes that the PES is parabolic and that the motion of the atoms is harmonic around the equilibrium position (blue curve in Figure 9) resulting in the identification of characteristic normal modes of vibration.

In many situations, such as those involving hydrogen bonds, the assumption of harmonic behavior is insufficient, and harmonic analysis yields unsatisfactory results when compared to experimental observations. To overcome the limitations of HO approximation, the anharmonic oscillator model has been developed, wherein the PES can be generated by calculating the electronic energy of the molecule for various nuclear displacements (for example the Morse anharmonic model in Figure 9). These models provide a more accurate description of the vibrational behavior of polyatomic molecules, but can be computationally more demanding to solve than the simple HO model. In fact, the construction of the full global potential for molecules with more than a few atoms becomes computationally demanding because the number of points scales exponentially with the number of DOFs. Therefore, achieving greater efficiency in vibrational calculations can be done through two approaches: reducing the dimensionality of the problem and decreasing the number of single point calculations required for constructing the potential.



**Figure 9:** Schematic representation of potential energy curve of the Morse anharmonic (solid black curve) and harmonic oscillator (solid blue curve) with different vibrational states (dashed black line and dashed blue line, respectively) and corresponding wave functions (red curves) of molecule A-B. In the harmonic oscillator model, the corresponding energy levels of the vibrational states are equally spaced by  $\hbar\omega$ , and the equilibrium bond length ( $r_{\text{eq}}$ ) is constant for all energy levels. In the anharmonic oscillator model, the potential energy surface is no longer parabolic, and the frequency of vibration can depend on the amplitude of the motion. The anharmonic potential is constructed from calculations of the molecular electronic energy for different internuclear distances.

To tackle the high-dimensionality issue, the full-dimensional PES can be approximated by using potential energy functions (PEFs) of lower dimensionality and high-order mode couplings in a hierarchical manner. This approximation is known as the *n-mode representation* or *n-mode expansion* of the potential in the vibrational context [239–241]. To construct the PES expansion, a series of approximate PESs are utilized

$$V^{(1)}, V^{(2)}, \dots, V^{(M)}, \quad (47)$$

where  $V^{(1)}$  corresponds to the sum of uncoupled anharmonic oscillators while  $V^{(2)}$  includes all couplings among pairs of modes, etc. For example, summing over terms containing only one and two-mode couplings a two-mode representation of the PES can be written as

$$V^{(2)} = \sum_{m_1} \bar{V}^{m_1}(q_{m_1}) + \sum_{m_1 < m_2} \bar{V}^{m_1 m_2}(q_{m_1} q_{m_2}) \quad (48)$$

where  $\bar{V}^m$  denotes the PEFs, which are also referred to as “cuts-functions”. These sub-potentials are functions of selected set of modes, which is also referred to as a mode combination (MC). For the two-mode case example, the sub-potential of a mode  $m_1$  and  $m_2$  reads

$$\bar{V}^{m_1 m_2}(q_{m_1} q_{m_2}) = V^{m_1 m_2}(q_{m_1} q_{m_2}) - V^{m_1}(q_{m_1}) - V^{m_2}(q_{m_2}). \quad (49)$$

In this framework, the full-dimensional PES can be approximated as a sum over many low-dimensional sub-potentials as

$$V(q_1, q_2, \dots, q_M) \approx V^{\text{MCR}} = \sum_{m \in \text{MCR}[V]} \bar{V}^{m_n}, \quad (50)$$

where mode combination range (MCR) denotes the range of included MCs of various sizes. Depending on the desired accuracy and efficiency, we can choose to include or exclude certain MCs in the potential. For instance, in this thesis, a selected set of modes was used for generating the PES and conducting vibronic calculations, up to the second order of the n-mode expansion. When all mode-coupling levels are included the expansion becomes exact.

It should be noted that the required number of single points for the n-mode expansion increases with  $M^n$ . Thus, an obvious way to improve efficiency in vibrational calculation, after the dimensionality problem has been dealt with, is to decrease the number of single point calculations needed for constructing the potential. Equidistant grid points are one method of selecting grid points; however, it may be necessary to use a higher density of grid points for certain regions, particularly near the minimum geometry. To achieve this, an adaptive approach can be taken to enhance the PES’s accuracy in critical regions. In the MidasCpp [242] program package, to avoid using pre-defined static grids of single points, the adaptive density guided approach (ADGA) [243] is used to automatically construct the grid of single points. In the ADGA scheme, the generation of the evaluation grids is guided dynamically by the densities of the vibrational wave functions, taking into account both the size of the grid and the mesh of evaluation points. This method also allows the incorporation of multilevel or multiresolution capabilities [244].

### 2.3.2.2 *Vibrational wave function*

As discussed previously, the inclusion of anharmonic terms in the potential energy surface becomes necessary for more accurate descriptions of molecules. One should bear in mind that developing a suitable potential cannot be viewed as independent of the task of determining the wave function and corresponding energies.

Numerous methods have been developed to incorporate the correlation between vibration motions through mode-mode couplings, which can be considered analogous to the hierarchy of wave function-based approaches used in electronic structure theory [245–247]. The main focus of this subsection is thus on wave function methods using the vibrational self-consistent field (VSCF) as starting point and vibrational configuration interaction (VCI) as an approach used for including the correlations between the different degrees of freedom.

At variance with electrons, vibrational modes are treated indistinguishably so that the  $M$ -mode wave function of a given vibrational configuration  $m$  does not

need to be antisymmetrized and can be written as a Hartree product of one-mode functions  $\phi_{s^m}^m(\mathbf{q}_m)$  as

$$\Phi_s(\mathbf{q}) = \prod_{m=1}^M \phi_{s^m}^m(\mathbf{q}_m) = \phi_{s^1}^1(\mathbf{q}_1)\phi_{s^2}^2(\mathbf{q}_2)\dots\phi_{s^M}^M(\mathbf{q}_M) \quad (51)$$

with  $s$  being an  $M$ -dimensional index vector indicating the excitation level for each mode. A general wave function in the  $M$ -mode space can be written as a linear combination of the direct product functions

$$\Psi_{\mathbf{c}}(\mathbf{q}) = \sum_{\mathbf{s}} c_{\mathbf{s}} \Phi_{\mathbf{s}}(\mathbf{q}) = \sum_{s^1}^{N^1} \sum_{s^2}^{N^2} \dots \sum_{s^M}^{N^M} c_{s^1 s^2 \dots s^M} \Phi_{s^1 s^2 \dots s^M}(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_M). \quad (52)$$

From this equation, it is noticed that the dimension of the problem is equal to the (usually very high) number of  $\mathbf{c}$  coefficients. This extreme increase in complexity with system size prevents “exact” quantum dynamics for more than a few DOFs.

One of the fundamental techniques employed in vibrational structure theory is the VSCF method [246–248], which is a vibrational equivalent of the HF method in the electronic structure theory. It is a mean-field approach in which each vibrational DOFs interacts with an averaged potential of the other modes. By selecting a reference (GS) index vector  $\mathbf{i}$ , the Hartree product is written as

$$\Phi_{\mathbf{i}}(\mathbf{q}) = \prod_{m=1}^M \phi_{i^m}^m(\mathbf{q}_m). \quad (53)$$

The reference index  $\mathbf{i}$  vector defines which level for each mode that is considered to be “occupied”. The energy

$$E_{\mathbf{i}} = \langle \Phi_{\mathbf{i}} | \hat{H} | \Phi_{\mathbf{i}} \rangle \quad (54)$$

is optimized under variation of orthonormal one-mode functions  $\phi_{s^m}^m(\mathbf{q}_m)$ . Applying the variational principle, we lead to the SCF equations for one-mode functions

$$F^{m,i} \phi_{i^m}^m(\mathbf{q}_m) = \epsilon_{i^m}^m \phi_{i^m}^m(\mathbf{q}_m) \quad (55)$$

where an effective operator (mean-field Hamiltonian) is introduced for each mode

$$F^{m,i} = \left\langle \prod_{m'=1, \neq m}^M \phi_{i^{m'}}^{m'}(\mathbf{q}_{m'}) \middle| \hat{H} \middle| \prod_{m''=1, \neq m}^M \phi_{i^{m''}}^{m''}(\mathbf{q}_{m''}) \right\rangle.$$

It is obtained by integrating the Hamiltonian over all other modes. The VSCF eigenvalues can be written in terms of the Hamiltonian and the eigenfunctions as

$$\epsilon_{i^m}^m = \langle \phi_{i^m}^m(\mathbf{q}_m) | F^{m,i} | \phi_{i^m}^m(\mathbf{q}_m) \rangle = \langle \Phi_{\mathbf{i}} | \hat{H} | \Phi_{\mathbf{i}} \rangle = E_{\mathbf{i}}. \quad (56)$$

The  $\phi_{i^m}^m(\mathbf{q}_m)$  functions can be expressed using a primitive basis. Choosing an appropriate primitive basis for each vibrational motion is primordial, as they define the building blocks of the anharmonic vibrational wave function and will

therefore determine its quality (flexibility and compactness). The primitive basis could consist of HO functions, but local basis functions like B-spline functions have proven to be more effective for ADGA calculations. [249]

It is noteworthy that terms incorporating anharmonic treatment in the wave function can be categorized into two distinct components: (i) anharmonic terms in a single mode, and (ii) mode-coupling terms that relate to two or more modes simultaneously. In the absence of mode-coupling terms, the one-mode terms are treated exactly by VSCF. Within the VSCF, the mode-mode correction is only partially addressed through the mean-field description: each mode experiences the effects of other modes only through the averaged field of the coupling terms across these modes. This is analogous to the problem of electronic correlation in electronic SCF. Despite these limitations, the VSCF method provides a good starting point for more accurate methods [245, 246, 248, 250, 251]. One of the widely used approach for dealing with mode-mode correlation is vibrational configuration interaction (VCI) [245, 246, 248, 250] using the ansatz shown in equation 52.

The underlying formulation of this approach is often based on the second quantization (SQ) [251]. It is an important concept to enable the treatment of many-body systems in a compact and efficient way. Analogous to orbitals in molecular electronic structure theory, a set of one-mode wave functions known as “modals” are represented in an occupation number vector (ONV)

$$|\mathbf{k}\rangle = | \{k_0^1, k_1^1, \dots, k_{N^1-1}^1\}, \dots, \{k_0^m, k_1^m, \dots, k_{N^m-1}^m\}, \dots, \{k_0^M, k_1^M, \dots, k_{N^M-1}^M\} \rangle \quad (57)$$

with  $k_p^m$  being an integer denoting the occupancies of the one-mode level corresponding to mode  $m$  with modal  $p$ ,  $N^m$  is the size of the modal basis for mode  $m$ . The integers specify how many times the modal enters the Hartree product. Figure 10 illustrates an example of modals that are occupied and a comparison between the first quantization (FQ) notation and the SQ notation for the ONV. In the context of molecular vibrations, each mode is described as having a single occupied modal which leads to the equation

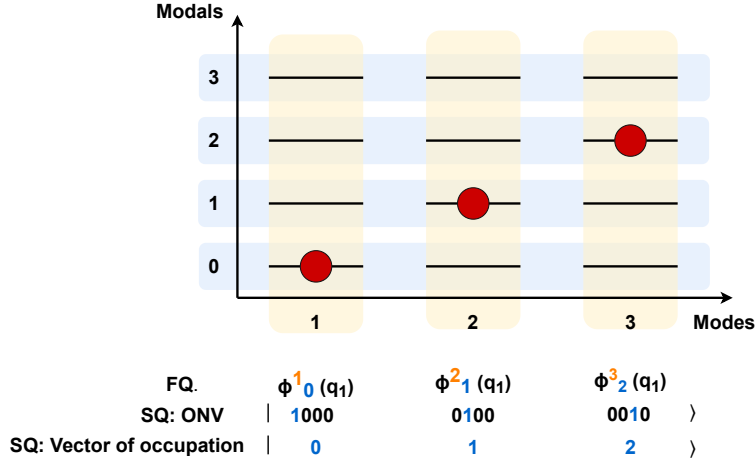
$$\sum_{p=1}^{N^m} k_p^m = 1. \quad (58)$$

The ONV representing the state with no occupied modes, where all  $k$  values are zero, is referred to as the vacuum state  $|\text{vac}\rangle$ . To fully define the space, the inner product between two ONVs is defined as

$$\langle \mathbf{k} | \mathbf{l} \rangle = \prod_{m=1}^M \prod_{p=1}^{N^m} \delta_{k_p^m l_p^m}, \quad (59)$$

where  $\delta$  is the Kronecker delta function arising from the overlap integrals of the orthonormal modals. To manipulate ONVs, creation and annihilation operators are introduced

$$a_p^{m\dagger} |\dots, k_p^m, \dots\rangle = \sqrt{k_p^m + 1} |\dots, k_p^m + 1, \dots\rangle, \quad (60)$$



**Figure 10:** An example with three vibrational modes represented by yellow shadows, where each mode consists of four modals represented by blue shadows. The red dots indicate the occupation of a specific modal, and the resulting product of one-mode functions and occupation number vectors for first quantization (FQ) and second quantization (SQ), respectively, as well as the notation for the vector of occupations, are shown below the modals.

$$a_p^m |\dots, k_p^m, \dots\rangle = \sqrt{k_p^m} |\dots, k_p^m - 1, \dots\rangle. \quad (61)$$

These operators denoted  $a_p^{m\dagger}$  and  $a_p^m$ , respectively, create and annihilate an occupied state for the mode  $m$  in the  $p$ -th modal. They satisfy the following commutation relations

$$[a_p^{m\dagger}, a_q^{m'\dagger}] = [a_p^m, a_q^{m'}] = 0 \quad (62)$$

$$[a_p^m, a_q^{m'\dagger}] = \delta_{mm'} \delta_{pq}. \quad (63)$$

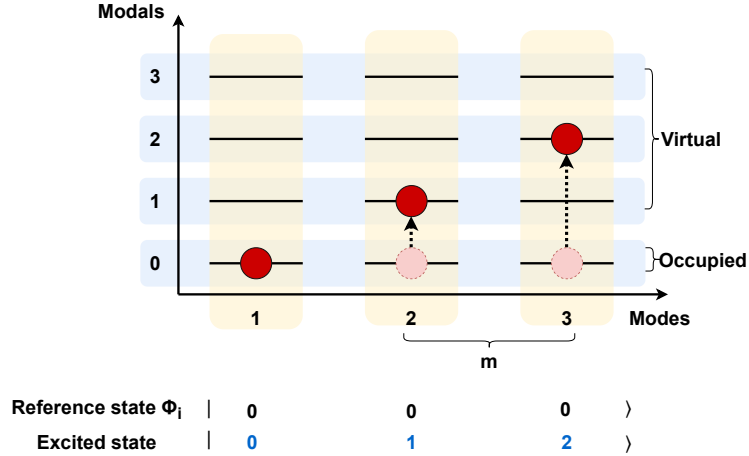
In SQ, the indices  $i, j, k$ , etc., refer to the modals occupied in the reference state, while  $a, b, c$ , etc., correspond to virtual modals, and  $p, q, r$ , etc., refer to arbitrary modals. It is important to note that this occupation vector should not be confused with the ONV (see Figure 10). The VSCF wave function, which is usually used as the reference Hartree-product for excited wave function e.g. VCI can be written in terms of a string of creation operators as

$$|\text{VSCF}\rangle = |\phi_i\rangle = \prod_{m=1}^M a_{i_m}^{m\dagger} |\text{vac}\rangle. \quad (64)$$

Excited configurations can be obtained by introducing excitation operators of the form

$$\tau_{\mu^m} = \prod_{m \in \mathbf{m}} a_{a^m}^{m\dagger} a_{i^m}^m. \quad (65)$$

An example of excitation from the reference state to another state in SQ is illustrated in Figure 11. Here, “excitation” refers to the promotion of one or more



**Figure 11:** Schematic illustration of the vibrational state space of vibrational self-consistent field as reference state and correlated vibrational wave functions in its excited state. MC  $\mathbf{m}$  is the set of modes that are excited from the reference state (dashed dot) to the excited state (solid dot) by the excitation operator.

modes from the occupied levels to the unoccupied levels relative to the reference ket.  $\mathbf{m}$  refers the set of modes that are excited by the operator, while  $\mu^{\mathbf{m}}$  is an index that specifies the modals excited to. The Hermitian adjoint to  $\tau_{\mu^{\mathbf{m}}}$  correspond to a de-excitation.

Given a particular choice of reference state, the linear VCI wave function ansatz can be written as

$$|\text{VCI}\rangle = |\Phi_i\rangle + \sum_{\mu^{\mathbf{m}}} C_{\mu^{\mathbf{m}}} \tau_{\mu^{\mathbf{m}}} |\Phi_i\rangle \quad (66)$$

using intermediate normalization  $1 = \langle \Phi_i | \text{VCI} \rangle = C_i$ . The  $C_{\mu^{\mathbf{m}}}$  are the variational VCI parameters determined from VCI eigenvalue equations.

The exact wave function for a given Hamiltonian and one-mode basis set can be obtained by including the full sum over all excitations, resulting in the full VCI wave function. Due to the large number of possible excitations, the full VCI expansion becomes computationally intractable for larger systems. Therefore, truncated VCI methods are often used, which include only a selected set of excitations. By limiting the number of excitations to only  $n$ -mode excitations, we obtain approximated wave functions that have significantly fewer parameters. This approach leads to a hierarchy of VCI models, denoted VCI[1], VCI[2], VCI[3], VCI[4], and so on, which are guaranteed to converge to the full VCI results for that specific set of modals within basis set limit. For instance, in this thesis, a truncated VCI method with up to 3 or 4 orders of excitations is employed.

### 2.3.3 Computation of molecular optical spectra

The following subsection examines the theory behind calculating spectral intensity and line shapes, as well as the widely used independent mode, displaced harmonic oscillator IMDHO approximation.

#### 2.3.3.1 Spectral intensity and Franck-Condon factors

One of the main tasks in the computation of optical spectra is the determination of the band intensities or transition probabilities. The intensity of the absorption or emission peak can be quantified by the oscillator strength. This value is dependent on the dipole moment change between the initial and final states. Accurately simulating electronic spectra requires accounting for the complex interplay between electronic and nuclear motion, which can be challenging. Vibrational and rotational transitions accompany electronic transitions, leading to the broadening of spectral lines [252]. In theoretical calculations of electronic spectra, it is common to neglect the explicit treatment of rotational structure. This simplification is mainly due to the experimental challenges in resolving rotational transitions, particularly for large molecules. [253] Instead, the focus is typically on the vibrational structure, which can be accounted for using the Franck-Condon (FC) approximation [254, 255]. This approximation assumes that the vibrational motion is much faster than the electronic motion, allowing us to separate the electronic and vibrational contributions to the transition.

In most cases, the intensity  $I_{n \leftarrow m}$  of a transition between the two electronic states  $\Psi_m$  and  $\Psi_n$  is proportional to the square of the transition dipole moment, which has the form

$$\boldsymbol{\mu}_{n,m} = \langle \Psi_n | \hat{\boldsymbol{\mu}} | \Psi_m \rangle, \quad (67)$$

where  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_e + \hat{\boldsymbol{\mu}}_N$  is the transition dipole operator with  $\hat{\boldsymbol{\mu}}_e$  and  $\hat{\boldsymbol{\mu}}_N$  being the electronic and nuclear dipole operators, respectively. Assuming a product wave function and considering only vibrational modes of nuclear motion, we have  $|\Psi_m\rangle = |\Psi(\mathbf{r}, \mathbf{R})_e \Psi(\mathbf{R})_v\rangle$  and  $|\Psi_n\rangle = |\Psi(\mathbf{r}', \mathbf{R}')_{e'} \Psi(\mathbf{R}')_{v'}\rangle$  with  $\mathbf{r}$ ,  $\mathbf{R}$  being electron and vibrational coordinates, respectively. The variables  $v$  and  $e$  represent the vibrational and electronic components of the initial  $m$  state, while  $v'$  and  $e'$  represent the corresponding components in the final  $n$  state. By considering the vibronic transition  $ev \rightarrow e'v'$ , the transition dipole moment can be re-written as

$$\begin{aligned} \boldsymbol{\mu}_{n,m} &= \langle \Psi_{e'} \Psi_{v'} | \hat{\boldsymbol{\mu}}_e + \hat{\boldsymbol{\mu}}_N | \Psi_e \Psi_v \rangle \\ &= \langle \Psi_{e'} | \Psi_e \rangle \langle \Psi_{v'} | \hat{\boldsymbol{\mu}}_N | \Psi_v \rangle + \langle \Psi_{e'} \Psi_{v'} | \hat{\boldsymbol{\mu}}_e | \Psi_e \Psi_v \rangle. \end{aligned} \quad (68)$$

The  $\Psi_e$  and  $\Psi_{e'}$  wave functions are orthogonal, consequently the first term on the right-hand side of equation 68 vanishes. The electronic transition dipole moment can be defined as  $\boldsymbol{\mu}_{e',e} = \langle \Psi_{e'} | \hat{\boldsymbol{\mu}}^{e1} | \Psi_e \rangle$ . Thus, the transition dipole moment can be written as [252]

$$\boldsymbol{\mu}_{e'v',ev} = \langle \Psi_{v'} | \boldsymbol{\mu}_{e',e} | \Psi_v \rangle. \quad (69)$$



The electronic part of the transition dipole moment  $\boldsymbol{\mu}_{e',e}$  is dependent on the nuclear coordinates [252]. Thus, it is convenient to expand the  $\boldsymbol{\mu}_{e',e}$  into a Taylor series in terms of displacements along the normal modes ( $Q_k$ ) around the equilibrium position ( $Q_0$ ) of the initial state[253]

$$\boldsymbol{\mu}_{e',e}(Q) \approx \boldsymbol{\mu}_{e',e}(Q_0) + \sum_{k=1}^N \frac{\partial \boldsymbol{\mu}_{e',e}}{\partial Q_k} (Q_k - Q_{k,0}) + \dots, \quad (70)$$

where  $N$  is the number of vibration modes. By substituting equation 69 into equation 70, the following expression is obtained

$$\boldsymbol{\mu}_{e'v',ev} \approx \boldsymbol{\mu}_{e',e}(Q_0) \langle \Psi_{v'} | \Psi_v \rangle + \sum_{k=1}^N \frac{\partial \boldsymbol{\mu}_{e',e}}{\partial Q_k} \langle \Psi_{v'} | Q_k - Q_{k,0} | \Psi_v \rangle + \dots, \quad (71)$$

In the Condon approximation, the electronic transition dipole moment is assumed to be independent of the nuclear coordinates. As a result, only the zero-th order term or the so-called FC factor [254, 256–260] of the expansion provided in equation 71 is considered. Although there are other contributions such as Herzberg-Teller factors [261] (first-order terms), they are typically (but not in all cases) insignificant [40] and are often neglected. Thus, the transition dipole moment for a vibronic excitation ( $e \neq e'$ ) within the FC model can be expressed as

$$\boldsymbol{\mu}_{e'v',ev} = \boldsymbol{\mu}_{e',e}(Q_0) \langle \Psi_{v'} | \Psi_v \rangle. \quad (72)$$

The calculation of FC factors often approximates the PESs as harmonic potentials [259, 262, 263]. However, as mentioned earlier, for more accurate descriptions of vibrational motions, the inclusion of anharmonicities and mode-mode couplings e.g. employing VCI wave functions are crucial.

### 2.3.3.2 Line shape of the spectra

The spectral profile of one-photon transition is given by

$$\sigma(\omega) \approx \sigma^{\text{FC}}(\omega) = \frac{4\pi^2\omega}{3c} \sum_{e'} \sum_{v'} |\boldsymbol{\mu}_{e',e}(Q_0)|^2 |\langle \Psi_{v'} | \Psi_v \rangle|^2 \delta(E_{v'} - E_v \pm \omega), \quad (73)$$

where  $E_i$  is the (absolute) energy of vibrational level  $i$ ,  $\omega$  represents the frequency of the incident or emitted light,  $c$  is the speed of light and  $\delta$  is the lineshape function. The choice of lineshape function is crucial in accurately representing the broadening of spectral lines. The most commonly encountered line shapes are the Gaussian and the Lorentzian. While a Gaussian line shape is often taken as evidence of static inhomogeneous broadening, a Lorentzian line shape is evidence of “lifetime broadening” according to the Heisenberg uncertainty principle . [264]

In addition, environmental effects, such as solvent or temperature, can significantly influence the line shape of vibronic spectra. As most of the interesting

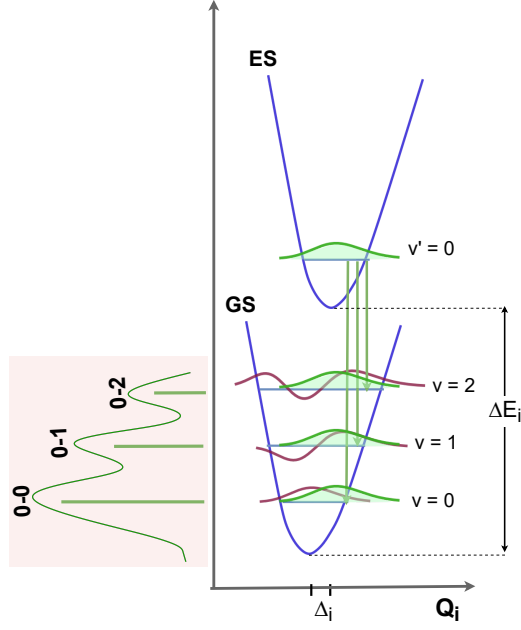
photophysical and photochemical phenomena (i.e., UV-vis absorption and fluorescence in biological systems) occur in solution, in particular in aqueous environment, many efforts have been devoted to modeling the optical spectra of solutes in aqueous solution. The polarizable continuum model (PCM) [265–267] is a commonly employed method that utilizes the bulk properties of the solvent to induce polarization in a solute, making it computationally advantageous. It, however, usually does not sample solute–solvent configurations, which determine excited state properties. [29] Additionally, continuum methods face substantial difficulties in accurately simulating spectral lineshapes, which encompass both the high-energy tail resulting from vibronic transitions and the inhomogeneous broadening arising from solute–solvent interactions. [268, 269]. Treating the solvent explicitly at the QM level allows for a more precise representation of these interactions. However, the substantial computational requirements associated with this approach often limit its usage to studying the microsolvation effect. Despite the limitations in the number of solvent molecules, such solvation effects are particularly important for gaining insights into the local solute-solvent interactions and their impact on various properties and processes, including solute stability, [270] reactivity, [271] spectroscopic properties [272], and intermolecular interactions [273], etc.

### 2.3.3.3 *Independent mode, displaced harmonic oscillator model for vibronic spectra*

As mentioned earlier, calculating FC integrals for polyatomic molecules is a challenging task as it involves determining the PESs of the relevant electronic states and performing a complete frequency analysis. The independent mode, displaced harmonic oscillator (IMDHO) model, which is also known as the multidimensional separable harmonic approximation, [258, 260] is a straightforward but feasible approach for FC calculation, and a widely used approach for calculating the vibrational intensity distribution in the electronic spectra of sizable systems [32, 274–276]. This model makes several assumptions, including:

- i) The molecule’s vibrational modes are independent of one another and can be treated as separate harmonic oscillators.
- ii) The potential energy surface of the final electronic state can be approximated as a displaced initial electronic state. In other words, the vibrational frequencies remain unchanged.
- iii) The final state exhibits the same normal modes as the initial state. In other words, Duschinsky rotations [277] are neglected.

Accordingly, the PESs of the two electronic states involved in a transition are harmonic and differ only by the position of their minimum i.e., shifts in energy ( $\Delta E_i$ ) and position along the normal mode ( $\Delta_i$ ), see Figure 12. The vibrational displacements of the ES are approximated from the ES gradients at the equilibrium structure of the GS [278]. By neglecting any changes in frequency in the ES or



**Figure 12:** Simplified Jablonski diagram of the emission between two singlet states, the electronically ground-state (GS) and excited-state (ES). The intensity distribution of the emission (left) is approximated by the Independent Mode Displaced Harmonic Oscillator model (right). This model neglects changes in wave number and Duschinsky rotations.

Duschinsky mixing effects, the PES of the final state can be obtained from its initial state as

$$V_{e'}(Q_k) = V_e(Q_k - \Delta_i) \mp \Delta E_i = \frac{1}{2}k(Q_k - \Delta_i)^2 \mp \Delta E_i. \quad (74)$$

The vibrational states in the two relevant electronic states can be expressed as a multiplication of  $N_{\text{modes}}$  one-dimensional nuclear wave functions that are mutually orthonormal

$$|\Psi_v\rangle = \prod_{i=1}^{N_{\text{modes}}} |\Psi_{e,i}\rangle \quad (75)$$

leading to the FC integral

$$\langle \Psi_{v'} | \Psi_v \rangle = \prod_{i,i'=1}^{N_{\text{modes}}} \langle \Psi_{e',i'} | \Psi_{e,i} \rangle \approx \prod_{i=1}^{N_{\text{modes}}} \langle \Psi_{e',i} | \Psi_{e,i} \rangle. \quad (76)$$

As a result, the transition cross-section for transitions in equation 73 reads

$$\sigma(\omega) \approx \frac{4\pi^2\omega}{3c} \sum_{e'}^{N_{es}} |\mu_{e',e}|^2 \times \sum_{l_1} \sum_{l_2} \dots \sum_{l_{N_{\text{modes}}}} \prod_{i=1}^{N_{\text{modes}}} |\langle \Psi_{e',i} | \Psi_{e,i} \rangle|^2 \delta(E_{v'} - E_v \pm \omega), \quad (77)$$

where  $N_{\text{modes}}$  are the number of considered normal modes. By multiple summing over the final vibrational quantum numbers  $\Psi_{e',i}$  and  $\Psi_{e,i}$  for the normal mode  $i$

in the electronic states  $e'$  and  $e$ , respectively, all possible combinations of quantum numbers and modes in the different states are taken into account.

The IMDHO model allows the vibrational intensity distribution to be calculated by considering the overlap between the vibrational wave functions of the initial (ES) and final (GS) states, which is determined by the geometrical changes associated with the electronic transition, see Figure 12. In the limit of no Duschinsky mixing, the formulas for the FC factors from the vibrational ground state ( $0$ th) of the initial electronic state to the  $n$ th vibrational state of the final electronic state within the IMDHO approximation can be expressed in atomic units as [257, 258]

$$f_{n \leftarrow 0} = |\langle n|0\rangle|^2 = \frac{1}{n!} \left(\frac{\Delta_i^2}{2}\right)^n \exp\left(-\frac{\Delta_i^2}{2}\right) \quad (78)$$

where  $\Delta_i$  is the dimensionless normal-mode displacement of the two involved electronic states along the considered normal mode  $i$  (c.f. Figure 12). The vibronic profile of, e.g. an emission spectra, within the IMDHO model can directly be calculated via its auto-correlation function as [258, 274, 279]

$$I_{n \leftarrow m}(\omega) = \frac{4\pi\omega}{3c} |\boldsymbol{\mu}^{e1}|^2 \operatorname{Re} \int_0^\infty \exp(i[\omega - (E_{m,0} - E_{n,0})]t) \cdot \exp(-\gamma t) \times \prod_{j=1}^{N_{\text{modes}}} \exp\left[-\frac{\Delta_j^2}{2} (1 - \exp(i\omega_j t))\right] dt, \quad (79)$$

where  $c$  is the speed of light,  $\boldsymbol{\mu}^{e1}$  is the electronic transition dipole moment, which is assumed to be constant in the present work.  $\omega_j$  is the angular vibrational frequency of normal mode  $j$  in the initial state and  $\Delta_j$  the harmonic displacement of normal mode  $j$  in the same state. Both,  $\omega_j$  and  $\Delta_j$ , are assumed to be identical in both involved electronic states.  $E_{n,0}$  is the energy of vibrational level  $0$  in electronic state  $n$ . In the IMDHO model, the difference  $E_{m,0} - E_{n,0}$  is equal to the adiabatic excitation energy between the electronic states  $m$  and  $n$ .  $\gamma$  is the half-width at half-maximum (HWHM) for the underlying Lorentz line-shape function of the individual vibronic peaks. Hence,  $\gamma$  determines the broadening of the vibronic peaks. It is chosen prior to the computation of the IMDHO line shape. This time-dependent approach is particularly beneficial for large systems, as the accumulated intensity can be calculated for every frequency of interest rather than calculating a large number of FC factors.[274, 280]

In summary, achieving accuracy in computational spectroscopy is essential for supporting and complementing experimental results, as well as for predicting molecular and spectroscopic properties for new systems. However, this accuracy typically comes at a high computational cost, making some calculations unfeasible. Thus, there is a need for continued development of new and improved methods that balance accuracy with computational efficiency, as well as for the optimization of existing methods.

Part II

RESEARCH



RESEARCH OVERVIEW

---

Since the early days of protein study, chemists and biochemists have been aware that protein functions largely depend on the acquisition of specific three-dimensional structures by folding processes[281–283]. Under stress conditions, they may aggregate, often in an irreversible manner [2, 284–286] causing protein misfolding diseases. For early diagnosis of such diseases, new approaches including optical spectroscopy are being developed to detect these pathological hallmarks. Fluorescence techniques have emerged as a valuable tool in this field, offering high sensitivity and the ability to probe various structural aspects, such as aggregation states, conformational changes, and binding interactions. [287–291] However, the interpretation of experimental spectra is often a challenging task due to broad line shapes and the intrinsic heterogeneity of complex systems. To disentangle these complex signatures and disclose the underlying molecular properties, theoretical spectroscopy has been applied for decades as a powerful complementary technique [184].

Theoretical studies can provide a valuable understanding at an atomic level of the structure of a probes-protein complex, surpassing the resolution capabilities of experimental methods. A direct *vis-à-vis* comparison between experimental and computed spectroscopic data of complex systems is, however, still far from being standard due to the unfavorable scaling with the number of nuclear degrees of freedom (DOFs). An interesting question is, with the current software and hardware perspectives, how we can reconcile the accuracy and interpretability of theoretical study, particularly when anharmonic effects and/or environment need to be accounted for? This is the main research question of my first sub-project.

In this sub-project, the development and implementation of a set of approaches for the reliable calculation and prediction of vibrationally resolved emission spectra of pathological biomarkers is investigated. The present work mainly builds upon the previous findings that the vibrational progression of organic dyes, e.g. oligothiophenes is dominated by only a small number of modes [32, 292] and the number of required modes scales sublinear [32]. It enables a new scenario to keep the balance between accuracy and computational efficiency. To estimate the impact of individual modes on the vibronic profile, the harmonic displacement between the minima of the two involved electronic states within the independent mode, displaced harmonic oscillator (IMDHO) model is applied. As mentioned in chapter 2, this approximation is relatively simple, yet it enables the precise calculation of vibrational line shapes based on the harmonic frequencies of the initial state and the atomic gradients of the final state, considering only the equilibrium structure of the initial state. In this project, a so-called VCI-in-IMDHO

model, which embeds a reduced-space vibrational configuration interaction (VCI) anharmonic wave function treatment in the IMDHO model for mode selection and vibrational broadenings, is proposed. In contrast to the reduced-space VCI approach developed by Madsen and co-workers [32], this new approach includes the vibronic effect of the vibrational DOFs by focusing the effort on modes of interest and treating the rest harmonically. In the same study, an environmental effects on vibronic spectra for these relatively large molecular systems are also revealed. It suggests that the calculation of vibronic profiles in the condensed phase is also required for accurate simulations. In addition, as most biomarkers are used in the condensed phase, the effects of the environment, e.g., solute–solvent interactions, must be properly taken into account. Although implicit models of the environment can adequately capture the average behavior in solution, they are unable to account for specific environmental effects like atomistic electronic polarization and hydrogen bonding. To address this, spectral calculations incorporating the explicit environment are necessary. Hence, in the first project, I also explore the micro-solvation effects of water and betaine on the harmonic vibrationally resolved emission spectra of the target molecules. By considering anharmonic effects and the influence of the environment, the project contributes to the development of more accurate and efficient computational methods for studying complex systems.

For a rational design and development of new fluorescent biomarkers, a thorough comprehension of the binding interactions between the dye and protein is indispensable. In this regard, the second sub-project focuses on providing spectroscopic evidence of binding and X-ray crystallographic structure data on TTR complex formation with the fluorescent salicylic acid-based pyrene amyloid ligand (Py1SA). Overall, this project aims to bridge the gap between computational predictions and experimental observations, providing a comprehensive understanding of the TTR–Py1SA complex formation. To do so, molecular dynamics and umbrella sampling simulations are used to investigate the binding and unbinding mechanisms of Py1SA at an atomic resolution where a clear preference for one of the binding modes with the salicylic acid group pointing into the pocket, which would not be resolved in experiment, was found. The outcomes of this project have significant implications as they not only enhance a comprehension of the interactions between TTR and ligands but also offer a prospective fluorescent probe for investigating TTR aggregation and misfolding in future research.

Further elaboration on each project within the thesis will be provided in the upcoming chapters, where specific details and results of the research will be thoroughly discussed and analyzed. These chapters will delve into the methodologies employed, computational procedures undertaken, data obtained, and the corresponding interpretations, contributing to a comprehensive understanding of the individual projects and their respective contributions to the overall thesis.



METHODS AND COMPUTATIONAL DETAILS

---

*This chapter includes sections originating from the publications and manuscripts:*

**TMN. Nguyen** and C. König, “*Tailored anharmonic-harmonic vibrational profiles for fluorescent biomarkers*”, *Phys. Chem. Chem. Phys.*, 2022, 24, 14825-14835.

Y. Todarwal, C. Gustafsson, **TMN. Nguyen**, I. Ertzgaard, T. Klingstedt, B. Ghetti, R. Vidal, C. König, M. Lindgren, K P. R Nilsson, M. Linares and P. Norman, “*Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands.*” *J. Phys. Chem. B*, 2021, 125(42):11628–11636.

**TMN. Nguyen** and C. König, “*The role of micro-solvation on the computed emission spectra: The case of oxazines*”  
(*In preparation*)

**TMN. Nguyen**, A. Begum, J. Zhang, P. Leira, Y. Todarwal, M. Linares, P. Norman, D. Derbyshire, E. von Castelmur, M. Lindgren, P. Hammarström and C. König. “*Binding of pyrene-based fluorescent amyloid ligand to transthyretin: A combined crystallographic and molecular dynamics study.*”  
(*Under revision*)

*with some modifications.*

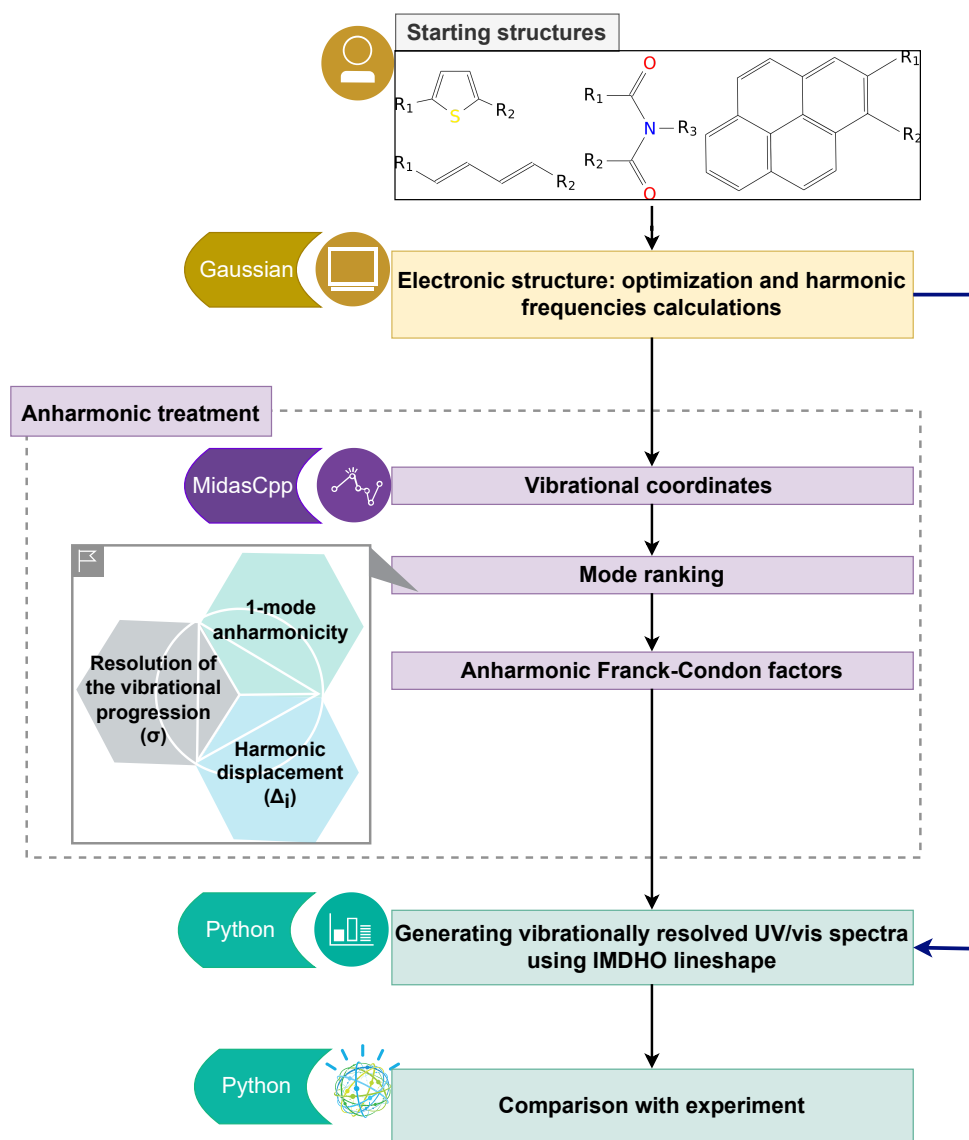
In previous chapters, the theory behind vibronic spectroscopies as well as the molecular and quantum mechanical model for the calculation of vibronic spectra and ligand–protein interaction has been discussed. In this chapter, I will present how such calculations are actually performed within the framework of software applications. Since they are applied to different studies, very different setups have been used. However, they can be divided into two main procedures: the first being the calculation of the vibronic spectra of the ligand with/without the inclusion of solvent effects, and the second being the calculation of the ligand–protein interaction. The first procedure involves setting up the electronic structure of the ligand, including the calculation of the electronic–structure, potential energy surface (PES) generation, and anharmonic vibrational wave function calculations. The procedure is organized in the order of the workflow displayed in Figure 13. The second procedure involves simulating the ligand–protein complex, including the calculation of the binding/unbinding energy of the ligand as shown in Fig-

ure 15. In the following sections, I will explain in more detail the specific methods and techniques used for these calculations, as well as the different software applications that have been employed.

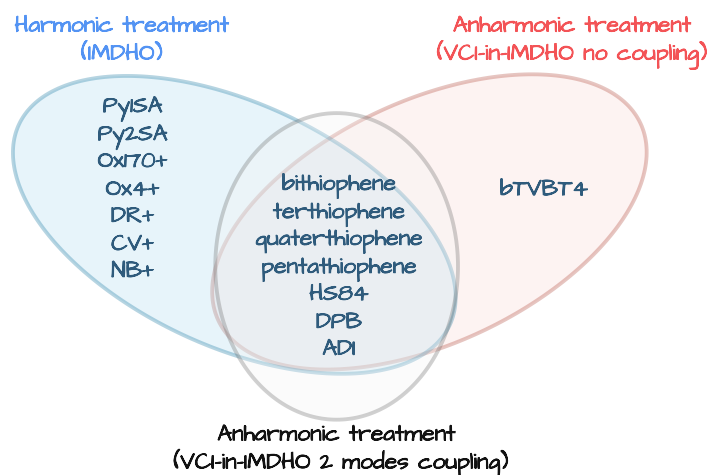
#### 4.1 VIBRATIONALLY RESOLVED UV/VIS SPECTRA CALCULATIONS

All electronic structure calculations were performed with the Gaussian16 program package [293] using time-dependent density functional theory (TD-DFT) with the CAM-B3LYP [229] functional. For computational efficiency and comparison purpose, different basis sets were used. The aug-cc-pVDZ [294] basis set was used for bTVBT4 while a def2-TZVP basis set [193, 195] was employed for organic dyes including pentameric oligothiophene derivative (HS84), 1,4-diphenylbutadiene (DPB), and anthracene diimide (ADI). The latter was also applied for pyrene-based dyes i.e. Py1SA and Py2SA; and oxazine dyes (nile blue (NB+), oxazine 1 (Ox1+), darrow red (DR+), cresyl violet (CV+), oxazine 4 (Ox4+), and oxazine 170 (Ox170+)). The calculations of harmonic absorption spectra for pyrene-based molecules were performed by Erik Rohloff while the microsolvation effects of betaine on the absorption spectra of DR+ and CV+ were calculated by Nils Weber in the internship projects under my supervision.

Different harmonic and anharmonic vibrational calculations were performed on the investigated probes, as shown in figure 14. For the PES generation and anharmonic vibrational wave function calculations, a locally modified version of the Molecular Interactions Dynamics And Simulation C++ package (MidasCpp) [242] version 2019.04.0 and 2019.10.0 was employed. For oligothiophenes, the PESs were taken from reference [32]. To account for new combinations of anharmonically treated coordinates, not previously considered in reference [32], I removed the corresponding regions from the PESs generated in that reference. For all vibrational profile calculations of the organic dyes and bTVBT4, the normal modes and vibrational frequencies of the first excited state were utilized. The vibrational coordinates of the excited state were also employed for the reduced-space anharmonic PESs for both the ground and excited state. The ground- and excited-state PESs for the fluorescent dyes were obtained with the multistate extension [295, 296] to the adaptive density guided approach (ADGA) [243] obtaining reliable PESs for at least the lowest six vibrational states in both electronic states. Additionally, mode-mode coupling up to second order in  $n$ -mode expansion was considered in the PES generation within a selected set of modes of organic dyes. The relative ADGA convergence criterion threshold was set to  $1.0 \times 10^{-2}$  while the absolute ADGA convergence criterion threshold was set to  $1.0 \times 10^{-6}$ . Both were increased by a factor of 10 for the two-mode coupling part of the PES. The PESs cuts were fitted to polynomials with the maximal polynomial order of eight. All vibrational self-consistent field (VSCF) calculations were performed with B-spline basis sets [249] with a basis set density of 0.8.



**Figure 13:** General workflow for the vibronic spectra modeling of fluorescent dyes. The starting structures of the dyes are optimized at the desired level for both ground and first excited states. Harmonic vibrational calculations are performed for every structure. After the mode ranking process using the proposed criteria, the anharmonic Franck–Condon factors are calculated for selected vibrational modes. The resulting stick spectra are then broadened with a line shape obtained from the independent mode, displaced harmonic oscillator (IMDHO) autocorrelation function and compared to experimental data.



**Figure 14:** *Vibrational treatment for investigated molecules*

The anharmonic Franck–Condon (FC) factors were obtained from vibrational configuration interaction (VCI) calculations. For all cases with four or more modes in the VCI treatment up to quadruply substituted configurations are included in the VCI treatment (VCI[4]) with the exception of bTVBT4 with a maximum of three simultaneously excited modes (VCI[3]). For smaller vibrational subspaces, full VCI calculations were conducted. The VCI FC factors [40, 297–299] were obtained as overlaps of the anharmonic VCI wave functions of the different vibrational states within the final electronic state with the vibrational ground state of the initial electronic state. The sum of FCs factors was required to be more than 0.98 to ensure that most of the intensity was captured in the calculated spectra. Hence, different numbers modals were included in the configurational space for different molecules.

The calculations of the FC factors were performed using the respective implementation [32, 296] in MidasCpp [242]. The different measures as well as the IMDHO line shapes were implemented in a python framework. Therefore, *scipy.integrate.quad* with techniques from the Fortran library QUADPACK [300] was used for the integration parts.

All the calculations were performed for the dyes in vacuum, if not specified otherwise. To get a rough estimate of the impact of the solvent environment on the vibrational progression, I performed the pure IMDHO computations additionally each dye using the integral equation formalism of the polarizable–continuum model (IEF-PCM) [301, 302] or its variation, a solvation model based on density (SMD) [303], with the default dielectric constants in Gaussian16. I chose for each dye the solvent of the respective experimental spectrum compared to, that is, water for HS84 (note that the experiment was done in a phosphate buffered saline), hexane for DPB, dichloromethane for ADI, and dimethyl sulfoxide for the Py1SA, Py2SA.

Additionally, the microsolvation effect of water and betaine on different oxazine dyes was also investigated. A preliminary conformational search for isolated oxazine molecules was performed taking into account initial molecular dynamic calculations by Tinker7 molecular modeling package [304]. The starting conformations obtained were also further optimized by the CAM-B3LYP functional in conjunction with the def2-TZVP basis set through Gaussian 16 package. Oxazine–water and oxazine–betaine complex geometries were then optimized at the same level starting from several test geometries, involving placing the solvent molecule at various positions around the chromophore core (see section 6 for more details). The harmonic spectra for each oxazine–solvent complex were generated as described above.

For a consistent theory–experiment comparison, the experimental spectra measured in wavelength scale are transformed in line shapes by applying an intensity correction proportional to  $\omega^2$ . [305, 306] I further normalized the area under the peak to 1 in all spectral comparisons and shifted the maxima of the experiment to the same position as the maxima of the respective most advanced calculation.

## 4.2 LIGAND–PROTEIN BINDING INTERACTIONS

The general workflow of molecular dynamics (MD) simulations as well as the umbrella sampling (US) for studying TTR–Py1SA complex systems is illustrated in Figure 15.

### 4.2.1 *Molecular dynamic simulations*

All molecular dynamics (MD) simulations were performed using the Gromacs version 2019.3 [307] with the Amber ff14SB force field [114] for the TTR protein and re-parametrized General Amber Force Field (GAFF) [308, 309] for Py1SA provided by Yogesh Todarwal. The starting structures of the TTR protein with Py1SA ligand existing in two configurations with respect to protein were based on the X–ray crystallography data. The X–ray analysis resulted in two structures for each conformer due to the two symmetry–equivalent binding sites as depicted in Figure 15. MD simulations were performed for all initial structures and refer to them as forward–B or forward–B' and reverse–B or reverse–B', respectively.

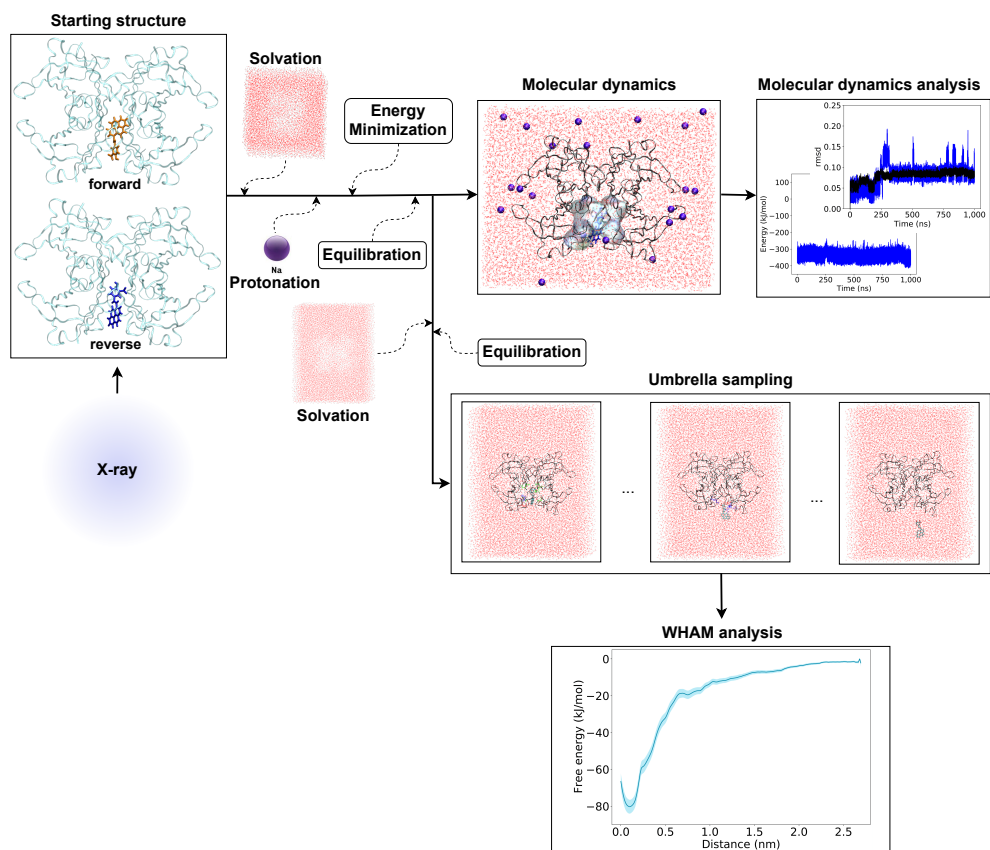
Gromacs Tools were used to solvate the protein–ligand system in a TIP3P [103] water box of size  $7 \times 7 \times 9 \text{ nm}^3$  with a total of approximately 12,000 water molecules. 21  $\text{Na}^+$  ions were then added to obtain system charge neutrality. The long–range electrostatic interactions were calculated using the particle mesh Ewald (PME) method [137] with a long–range cutoff of 1 nm which also is the cutoff of the short–range van der Waals interactions. Default settings were used for Fourier spacing (0.12 nm) and PME order (4). Periodic boundary conditions were applied to all three directions of the simulation box. The energy minimization was

then performed using the steepest decent algorithm with a maximum step size of 0.001 nm and a maximum force of 10 kJ mol<sup>-1</sup>nm<sup>-1</sup>. All atoms except the ones that are within 4 Å from the ligands (4 Å pocket) were under the harmonic constraints with the strength of 1,000 kJ mol<sup>-1</sup>nm<sup>-2</sup>. In addition, the linear constraint solver (LINCS) algorithm [310] was used to constrain all bonds. A short equilibration was propagated using velocity rescaling in the NVT ensemble at 300 K and Berendsen barostat in the NPT ensemble at 1 atm with a simulation course of 100 ps. Subsequently, the MD simulations were performed in the NVT ensemble for 1 μs simulation time. Processing of the different raw data from MD production was done using Gromacs utilities. Molecular graphics were prepared using the Visual Molecular Dynamics (VMD) program [311].

#### 4.2.2 Umbrella sampling simulations

The potential of mean force (PMF) surfaces were computed using umbrella sampling (US) [34, 147, 158]. MD simulations were employed to extract the initial coordinates for binding free energy calculations. The same experimental settings as in the molecular dynamic simulations section using Gromacs version 2021.3 [307] were applied, except the box size was set to 7×11×9 nm<sup>3</sup> and the constraints, which were applied to the backbone in the US simulations. Starting from the last snapshot of each mode from the equilibration MD phase, Py1SA was pulled away from its binding site about 2.5 nm using a force of 5,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the first (B) and 8,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the second structure (B') of reverse mode, 7,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the first (B) and 8,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the other structure (B') of forward mode. The Py1SA was pulled at the rate of 0.005 nm/ps. The spring constant as well as the pulling rate in each simulation were carefully selected based on our empirical study so that large overlaps of the histograms were obtained. From these pulling trajectories, snapshots with the equivalent distance of 0.025 nm were set as a starting configuration for each umbrella sampling simulation, which was independently simulated by performing an NPT equilibration for 100 ps followed by a 1 ns NVT trajectory. The force constant of the umbrella potential is set to 4,000 kJ mol<sup>-1</sup> nm<sup>2</sup> for all trajectories. Each simulation consists of a total of 101 umbrella windows. Additionally, 17 extra windows were added between 0.2 nm and 0.4 nm for each trajectory of these modes. Thus in total, for these modes, each simulation consists of 118 windows.

Finally, the weighted histogram analysis method (WHAM) [164, 165] was used to combine several windows into a PMF curve to estimate the binding free energy. For each initial structure, two pulling simulations followed by umbrella sampling were conducted to get the average PMF curve.



**Figure 15:** General workflow for the molecular dynamics analysis and the calculations of potential of mean force of TTR–Py1SA complex. The simulations of the complex start from experimentally determined structures from the X-ray method. There are two different binding modes of Py1SA (in sticks) to the protein (in ribbons): forward, reverse. Each results in B or B' binding modes due to the averaging of electron density along the AA' BB' symmetry axis. The starting structures are then solvated and protonated to simulate an environment that mimics the *in vivo* conditions. After relaxing, the systems are further equilibrated to the desired temperature and pressure. Molecular dynamics simulations are then run for 1 ms for data collection. The last conformer from the equilibrated system of each mode is then the starting structure for umbrella sampling. The weighted histogram analysis method is used to combine several windows into a potential of mean force curve to estimate the binding free energy.





TAILORED ANHARMONIC-HARMONIC VIBRATIONAL  
PROFILES FOR FLUORESCENT BIOMARKERS

---

*This chapter includes sections originating from the publications:*

**T.M.N. Nguyen** and C. König, “*Tailored anharmonic-harmonic vibrational profiles for fluorescent biomarkers*”, Phys. Chem. Chem. Phys., 2022, 24, 14825-14835.

Y. Tadarwal, C. Gustafsson, **T.M.N. Nguyen**, I. Ertzgaard, T. Klingstedt, B. Ghetti, R. Vidal, C. König, M. Lindgren, K P. R Nilsson, M. Linares and P. Norman, “*Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands.*” J. Phys. Chem. B, 2021, 125(42):11628–11636.

*with some modifications.*

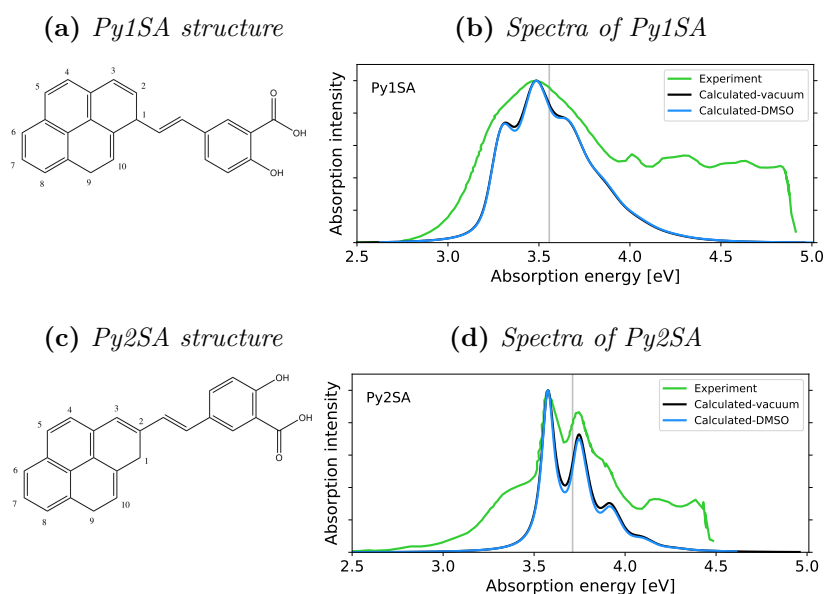
Computation of vibrational profiles of small molecules e.g. fluorescent biomarkers in the gas phase has been essential for understanding fundamental aspects of the structure and dynamics of molecules. As a matter of fact, with recent advancements in computational spectroscopy, many previously confronting problems are now within reach. Yet, the accurate simulation of these fluorescent dyes remains a significant challenge for theoreticians, notably due to the complex interplay of molecular vibrations and electronic transitions. The molecular vibrations, in particular, affect the spectral shape and intensity of the molecular excitation and emission spectra, making accurate predictions difficult. As a result, direct comparison of theoretical to experimental spectra [312–315] and determining the color of a dye [316, 317] can be challenging for medium to large-sized molecules. For example, the line shape of fluorescent spectra of oligothiophenes [318] can be well-described by theoretical models for vibrational broadenings. Given that, still, resolving the spectra line shape is a challenging field to theorists mainly due to the high computational cost of accounting for anharmonic effects.

In this chapter, a new computational framework is proposed for reliable theoretical vibrationally resolved emission spectra accounting for anharmonic effects for different chromophores such as oligothiophenes, 1,4-diphenylbutadiene (DPB), and an anthracene diimide (ADI) derivate. The method relies on the fact that the vibrational progression of a molecule e.g. oligothiophenes is dominated by only a small number of modes [32, 292] which can be sorted by using the displacement [32] from the independent-mode displaced harmonic oscillator (IMDHO) model [257, 260].

In this chapter, the VCI-in-IMDHO model and the measures for the division of the vibrational spaces are presented in the section 5.1 as well as the results in comparison to the experiment in the section 5.2. The computational details can be found in chapter 4.

## 5.1 VCI-IN-IMDHO MODEL

To assess the accuracy of reproducing the experimental spectra using the IMDHO line shape, the harmonic absorption spectra of two pyrene-based dyes are computed (Figure 16). These dyes have a salicylic acid (SA) group connected to either the first position (Py1SA) or the second position (Py2SA). These pyrene derivatives have a remarkable amyloid fibril binding affinity and display intriguing photophysical properties, such as differences in fluorescence brightness and lifetime in absorption spectra. These emissions are also found to change when binding to defective A $\beta$  1-42 fibrils, which are thought to be crucial in the progression of Alzheimer's disease [80].



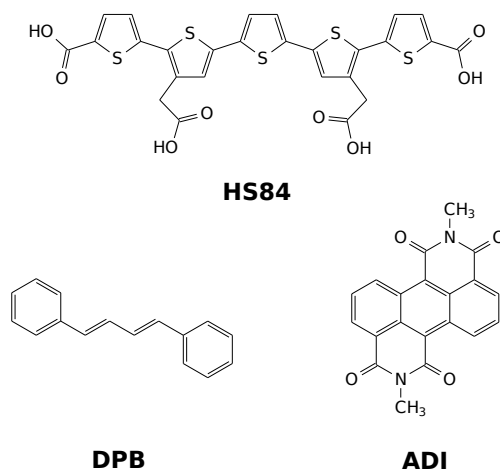
**Figure 16:** Structure and calculated absorption spectra of Py1SA and Py2SA in vacuum (black) and in dimethyl sulfoxide using SMD model (blue) broadened with a HWHM of 0.04 eV compared to an experimental spectrum in dimethyl sulfoxide [80] (green). Vertical electronic transition energy in gas phase is presented with grey lines. The experimental spectra of Py1SA and Py2SA are shifted by -0.019 eV and -0.461 eV, respectively to get the same position of maximum as calculated spectra in vacuum. The spectra in the condensed phase are also shifted by 0.146 eV and -0.347 eV, respectively.<sup>1</sup>

<sup>1</sup> These calculations were performed by Erik Rohloff in a research internship under my supervision, Leibniz Universität Hannover, 2022.

As shown in Figure 16, the crude IMDHO broadening model effectively replicates the difference in absorption spectra between Py1SA and Py2SA, which are two structurally similar molecules. This is important as it indicates that even small structural changes can have a significant impact on the absorption spectra of these types of molecules, which can be crucial in developing new fluorescence probes with specific properties. Additionally, the solvation effects are also accounted for in this study using the polarizable continuum model-SMD showing insignificant differences compared to the vacuum spectra. The ability to mimic the difference in absorption spectra between Py1SA and Py2SA of the IMDHO line shape highlights its utility in predicting the spectra of other structurally similar molecules, making it an important tool for further research in this field.

In the VCI-in-IMDHO, all important anharmonic modes are treated with an accurate VCI wave function, which accounts for the one-mode anharmonicity as well as for mode–mode couplings and correlations, usually restricted to second order. For all other modes, the IMDHO model is employed, which is purely harmonic and neglects frequency differences in the involved electronic states as well as Duschinsky rotations. Note that, in the VCI-in-IMDHO model, mode-mode coupling within the modes treated by the IMDHO model, as well as between the two sets of modes, is not taken into account. In contrast to other hybrid schemes for vibrational line shapes,[41, 267, 319–323] the VCI-in-IMDHO is not a mixed quantum–classical approach, but rather a mixed quantum–quantum scheme. So far, this model is restricted to FC factors, which can be assumed to be a good approximation for the bright transitions considered in this work.

The validation was first conducted on oligothiophenes containing two to five thiophene rings. The validated method is then applied to the fluorescent dyes depicted in Figure 17.



**Figure 17:** Representation of the investigated dyes: pentameric oligothiophene derivative (HS84), 1,4-diphenylbutadiene (DPB) and anthracene diimide (ADI) [324].

In the VCI-in-IMDHO hybrid scheme, a small subset of modes is first selectively chosen based on their importance for the calculation of FC factors, which are calculated as the overlap between the VCI wave functions of the vibrational states participating in the transition. The resulting stick spectra are then broadened with a line shape obtained from the IMDHO autocorrelation function [equation 80].

$$I_{n \leftarrow m}(\omega) = \frac{4\pi\omega}{3c} |\boldsymbol{\mu}^{el}|^2 \sum_s^{N_{\text{states}}} \text{Re} \int_0^\infty \exp\left(i[\omega - (\mathbb{E}_{m,0}^{\text{VCI}} - \mathbb{E}_{n,s}^{\text{VCI}})]t\right) \cdot \exp(-\gamma t) \times \prod_{j=1}^{N_{\text{HO modes}}} \exp\left[-\frac{\Delta_j^2}{2} (1 - \exp(i\omega_j t))\right] dt \quad (80)$$

In this model, a sum over all  $N_{\text{states}}$  vibrational states in the final electronic state obtained by anharmonic VCI calculations is introduced. The energy of the corresponding vibronic state  $\mathbb{E}_{n,s}^{\text{VCI}}$  equals the sum of the electronic energy of this state and the VCI vibrational energy of the respective vibrational state. Similarly,  $\mathbb{E}_{m,0}^{\text{VCI}}$  corresponds to the electronic energy of the electronic state  $m$  plus the anharmonic zero-point energy. Further, the product accounting for vibrational broadening in the IMDHO is now restricted to those  $N_{\text{HO modes}}$  modes not accounted for in the VCI treatment. Physical effects that induce broadenings, such as finite temperature effects, are not explicitly included within this study. For the above-described hybrid scheme, the modes must be divided into two sets based on three criteria:

1. *Importance of the mode for vibrational progression*: Harmonic displacement ( $\Delta_i$ ) based on the IMDHO model.

The ranking of mode importance is determined by utilizing the harmonic displacements calculated from the gradient of the ground electronic state at the equilibrium structure of the excited state. Similar as in references [32] and [276], the idea behind this criteria is that within IMDHO model large displacements of the potential energy curve cause non-vanishing FC overlap for multiple states leading to pronounced vibrational progression. In contrast, with zero displacements, only a single peak is observed rather than a pronounced vibrational progression as there are only overlaps between the vibrational ground-states. Thus, this first measure estimates the impact of the mode under consideration on the vibrational progression, i.e., whether more than the 0–0 transitions are significant.

2. *Resolution of the vibrational progression with a given HWHM ( $\gamma$ )*: Crossing point between the Lorentzian broadening the 0–0 and 1–0 transition ( $\sigma$ ).

The resolution of these modes is additionally taken into account when assuming line broadening, ensuring that the additional peaks are not obscured beneath the broadened lines. In reference [32], the authors used a pure frequency criterion to discard the vibrational modes with frequencies smaller than half the chosen HWHM. This approach, however, does not account for

the relation between the height of the satellite peaks (in the IMDHO model determined by the displacement  $\Delta_i$ ) and the resolution. In the new scheme, the second criterion proposed is the HWHM at which the maximum of the  $1 \leftarrow 0$  peak broadened with a Lorentzian line shape lies on the Lorentzian band of the  $0 \leftarrow 0$  peak. Only the modes with greater  $\sigma_i$  than the chosen HWHM are included in the VCI calculations.

### 3. *Impact of one-mode anharmonicity on the vibrational profile:*

In conjunction with the aforementioned criteria, the third criterion is employed to assess whether individual modes necessitate anharmonic treatment or if a harmonic treatment within the IMDHO model would suffice. This aids in narrowing down the computations to more realistic and manageable scenarios. To do so, a the ratio  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$  between the squared anharmonic ( $\Delta_i^a$ ) and harmonic displacements ( $\Delta_i$ ) is proposed. The anharmonic displacements  $\Delta_i^a$  for all modes  $i$  are obtained by projecting the difference between the optimized structures in the initial and final electronic states on the normal modes employed in the IMDHO model. These two measures, however, do not correlate clearly as shown in the section 5.2.2. Consequently, the impact of anharmonicity is evaluated by comparing the vibrational profiles of individual modes within the IMDHO model and anharmonic treatment.

It is also observed that to accurately identify and treat those modes with high precision and relatively low computational cost, a combination of all three measures is preferable over relying on a single criterion alone. For example, the vibrational progression caused by a mode with a high displacement, may not be seen due to a poor resolution, or it might already be well described in the harmonic picture. This can be seen with more details in the next section.

## 5.2 RESULTS AND DISCUSSION

### 5.2.1 *Oligothiophenes, revisited*

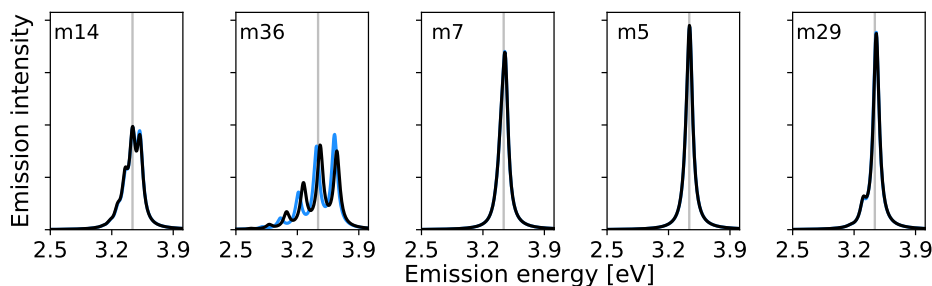
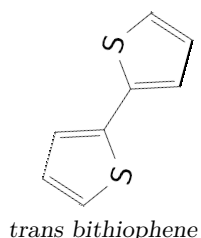
In the following, results are presented for mode rankings, the convergence of spectra for trans-oligothiophenes of different lengths using the reduced-space approach mentioned in reference [32], and the newly proposed VCI-in-IMDHO approach. The calculated spectra are then compared to the corresponding experimental ones [318] in ethanol at 77 K, that is, below the melting point of ethanol. Hence, the thermal broadening is likely suppressed in these spectra.

Following the outline of reference [32], the inclusion of anharmonic modes with decreasing displacements  $\Delta_i$  for the trans-bithiophene is evaluated. The most important modes for trans-bithiophene are shown in Table 1 with the one-mode

anharmonicity obtained by visual inspection of the respective spectra (see Figure 18).

Table 1: Mode ranking for trans-bithiophene with CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$ , anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-AH).

Mode (i)	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-AH
14	674	1.420	1.593	1.259	7473	weak
36	1658	1.350	0.422	0.098	5312	moderate
7	387	0.824	1.606	3.799	277	weak
5	290	0.529	0.415	0.615	117	weak
29	1214	0.446	0.303	0.461	403	weak



**Figure 18:** IMDHO-treated (blue) and VCI-treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of trans-bithiophene employing PESs from ref. [32]. The vertical transition energy is given as a vertical grey line.

By considering the reduced-space approach [32] with 0.2 as the displacement cut-off, there are five important modes (modes 14, 36, 7, 5, and 29). In the VCI-in-IMDHO scheme, modes 5 and 7, however, are neglected for the anharmonically treatment due to their resolution measure  $\sigma_i$ , which lies below the chosen HWHM  $\gamma$  value of  $322.6 \text{ cm}^{-1}$  (0.04 eV). In Figure 19 the vibrationally resolved emission spectra obtained by the pure IMDHO model and VCI are presented for different reduced spaces containing

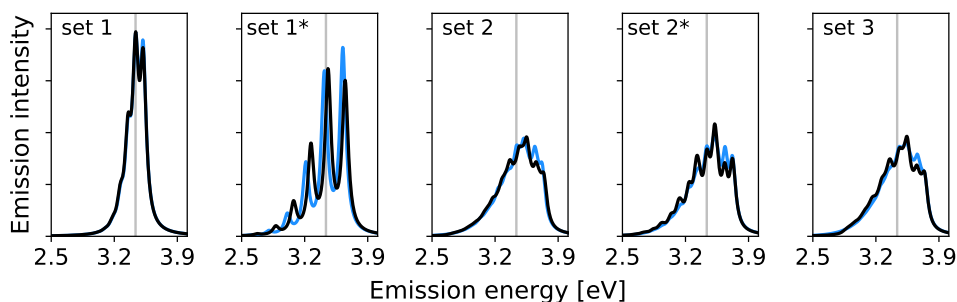
SET 1 only the mode with the largest IMDHO displacement, i.e., mode 14

SET 1\* only the mode with the most pronounced one-mode anharmonicity among the five modes with the largest IMDHO displacement, i.e., mode 36

SET 2 three modes with the largest IMDHO displacements according to the procedure in reference [32], i.e., modes 14, 36, 7

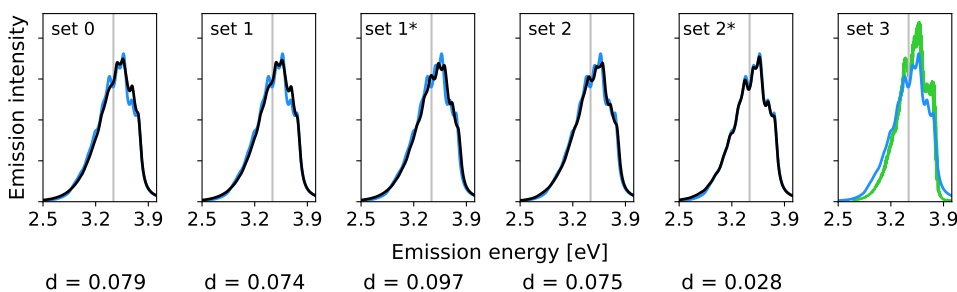
SET 2\* three modes with the largest IMDHO displacements and a resolution measure larger than the chosen half width at half maximum, i.e., modes 14, 36, and 29, and

SET 3 all five previously investigated modes, i.e., modes 14, 36, 7, 5, and 29.



**Figure 19:** *Reduced-space vibrational profile for the  $S_0 \leftarrow S_1$  emission for trans bithiophene with the IMDHO model (blue) and VCI (black). For the definition of the reduced spaces, see the main text. The vertical transition energy is given as a vertical grey line.*

It is observed, that the one-mode harmonic and anharmonic spectra are very similar (set 1) for mode 14, while they differ significantly for the mode with the most pronounced one-mode anharmonicity, i.e., mode 36. The same holds for all other sets, which all include the anharmonic mode 36.



**Figure 20:** *Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for trans bithiophene for the different sets of vibrational coordinates treated by VCI (black) and that for the largest set of modes set 3 (blue) with a HWHM of 0.04 eV compared to a corrected experimental spectrum of bithiophene at 77 K in ethanol [318] (green). For the definition of the reduced spaces, see the main text. The vertical transition energy is given as a vertical grey line. The experimental spectrum is shifted by 0.105 eV to get the same position of maximum as the calculated spectrum. Additionally, the deviation between the shown spectra according to equation (81) is quantified below the respective graphs.*

The calculated spectrum including both VCI treatment for the above-mentioned modes and IMDHO treatment for the rest is presented together with a pure IMDHO

treatment of all vibrational modes (set 0) and the experimental spectrum in Figure 20. No clear improvement is observed when mode 14 is included in the anharmonic calculation (set 1), but slight differences are noticed by including mode 36 in the VCI calculation. A similar trend was observed for the other sets. It is noticed that the vibrational profile for the hybrid scheme with set 3 closely resembles that of set 2\*. This observation is supported by the relatively small difference observed between the two spectra

$$d = \frac{\int_a^b |f(x) - g(x)| dx}{\int_a^b g(x) dx}. \quad (81)$$

This measure is used to obtain a more quantitative comparison of the convergence of the vibrational profiles to that obtained for the hybrid scheme with the different sets of modes. For that, the best model (hybrid calculation for set 3) is chosen as a reference spectrum ( $g(x)$ ). For the example of bithiophene, it can be concluded that considering anharmonicities, employing a refined resolution measure, and including less important modes within the IMDHO model can be advantageous for achieving increased convergence of the vibrational profile as the number of modes increases.

All-trans-terthiophene, all-trans-quaterthiophene and all-trans-pentathiophene show a similar convergence behavior. For these molecules, it is also possible to discard vibrational modes due to the more refined resolution measure and similar convergence behavior. In these cases, the agreement with the experiment is, however, only moderate mainly due to the solvation effect. Still, the VCI-in-IMDHO treatment leads to a systematic further improvement of this agreement. To conclude, in this section, it has been observed that reduced-space VCI vibrational profiles can be combined with IMDHO-type line shapes to the VCI-in-IMDHO model. This combination can lead to a faster convergence of the vibrational profile.

### 5.2.2 Assessment of the anharmonicity measure

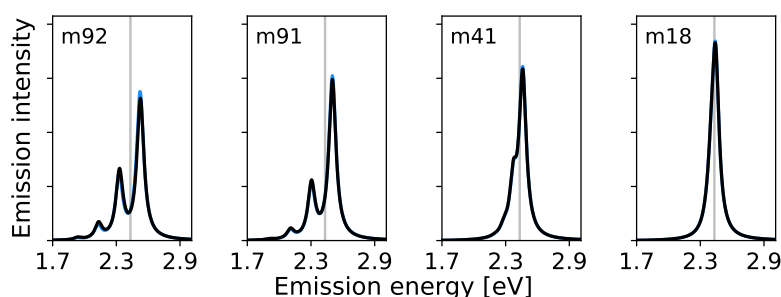
Anharmonic treatment is among the most challenging tasks of contemporary computational chemistry to obtain reliable results. To reduce the scaling problems of the extension of the anharmonicity and modes couplings to medium to large-sized systems, in this study an anharmonicity measure is introduced. It is obtained by comparing the IMDHO displacements to the anharmonic ones obtained by projecting the coordinate difference of the ground and excited state onto the respective vibrational coordinates  $\left(\frac{\Delta_i^a}{\Delta_i}\right)$ . If the squared ratio for all modes is equal to 1, a good approximation is expected for the IMDHO treatment, while large deviations suggest that the respective mode is more likely to exhibit anharmonic contributions. For thiophene test cases, all modes with significant (moderate) effect of one-mode anharmonicity on the one-mode vibrational profiles have an anharmonicity measure of  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2 < 0.1$ . For trans-pentathiophene (and the organic dyes), it is observed, however, that some modes with very small anharmonicity measures



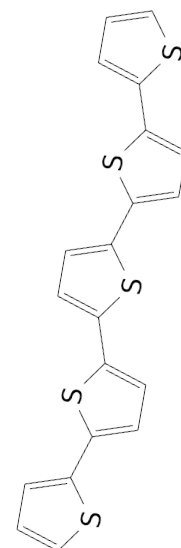
below the cut-off value that do not show significant one-mode anharmonicities as depicted in Table 2 and Figure 21.

Table 2: Mode ranking for all-trans-pentathiophene using CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$  and anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
92	1600	0.960	0.164	0.029	1479	weak
91	1592	0.835	0.191	0.052	1164	weak
41	711	0.795	0.989	1.549	483	weak
18	323	0.690	1.429	4.293	180	weak



**Figure 21:** *IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of all-trans-pentathiophene using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.*

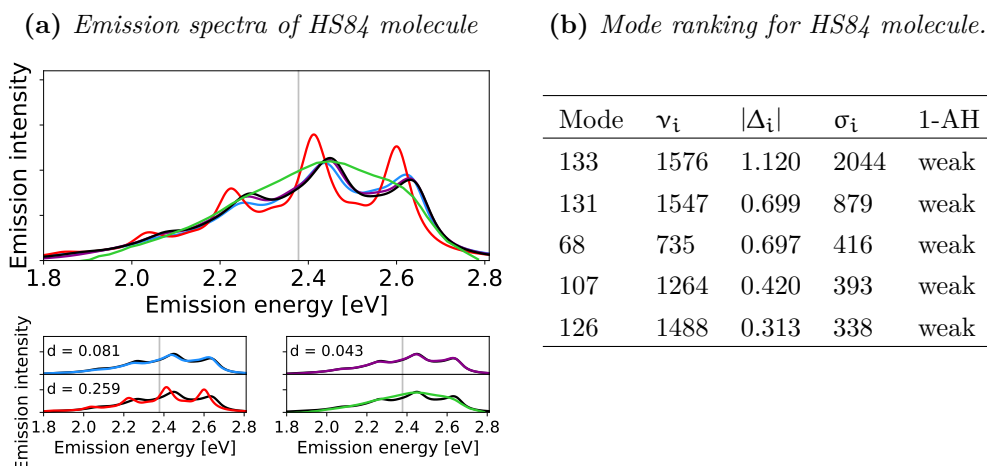


*all-trans-pentathiophene*

This observation might be related to that the FC region has a larger influence on the vibrational profile than the region around the equilibrium of the final state, as has been reported earlier[325]. For this reason, to assess the anharmonicity, the computationally more expensive comparison of the effect of the one-mode anharmonicity on the one-mode vibronic spectra directly is further applied. It is also noted that among the modes investigated for anharmonicity, those with high harmonic displacements also exhibit the largest anharmonicity effects on the vibrational profiles.

### 5.2.3 Organic biomarkers

Based on the results obtained in the previous section, the VCI-in-IMDHO scheme is now applied to the somewhat larger organic fluorescent dyes introduced in Figure 17 and compare their spectra to the experimental data. In contrast to the

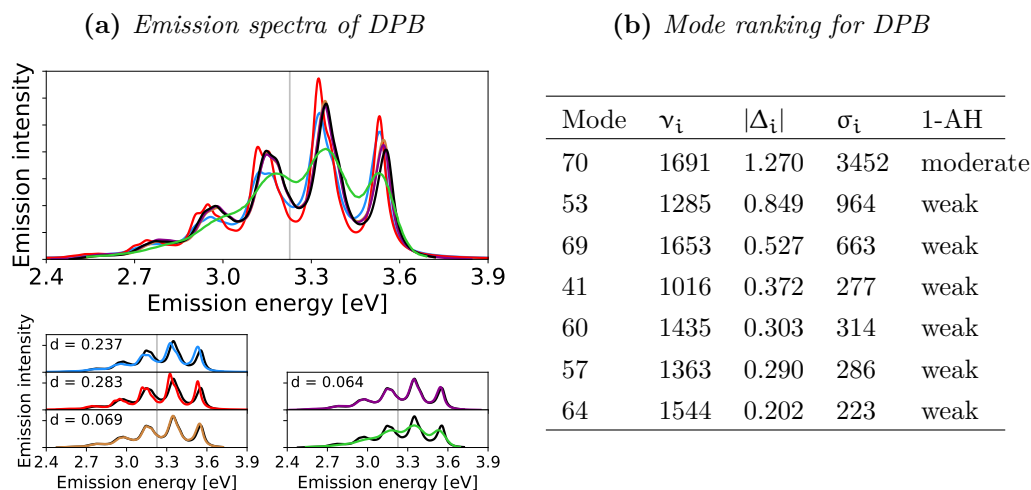


**Figure 22:** Calculated emission spectra of HS84 molecule (a) and the corresponding mode ranking (b) with CAM-B3LYP/def2-TZVP. See the caption of Table 1 for more details. (a): Calculated harmonic (blue), reduced space VCI (red), VCI-in-IMDHO without mode–mode couplings in the PES (magenta) and VCI-in-IMDHO (black) broadened emission spectra of HS84 molecule with a HWHM of 0.04 eV compared to a corrected experimental spectrum in PBS [49](green). Vertical electronic emission energy with grey lines. The VCI spectrum is obtained by conventional Lorentzian broadening. In the bottom part, the calculated spectra are directly compared to the best approximation (VCI-in-IMDHO) and the experimental spectrum, respectively. Additionally, the deviation between the calculated spectra according to equation 81 is quantified in the respective graphs. The modes treated by VCI in the reduced-space VCI are 131, 133, 68, 107, 126, out of which only the modes 133, 131, and 68 were coupled in the underlying PESs. The experimental spectra are shifted by 0.144 eV to get the same position of maximum as VCI-in-IMDHO spectra.

experimental spectra for oligothiophenes above, the broadening in the experimental spectra for the organic biomarkers can be assumed to be dominated by thermal broadening. Fewer details are resolved in these spectra. Furthermore, in interest of computational cost, only two-mode couplings were included in the PESs for modes, for which  $|\Delta_i| > 0.6$  for these larger examples.

For the oligothiophene-based fluorescent dye HS84, all modes are discarded from the anharmonic treatment due to the resolution measure and one-mode anharmonicity (see Figure 22-b). The corresponding calculated vibrational line shapes of the emission spectra for HS84 obtained for reduced-space VCI, IMDHO and VCI-in-IMDHO models are shown in Figure 22-a for a HWHM of 0.04 eV. The energy differences between the main peaks are similar in all approaches (between 0.180 eV and 0.187 eV). A slight deviation is obtained for the VCI-in-IMDHO spectrum with all the modes accounted for anharmonically including two-mode couplings for selected modes in the PES compared to that with only one-mode anharmonicities in the reduced-space PES (see Figure 22). This suggests that mode–mode coupling effects, similar to one-mode anharmonicities, play a small role here.

In 1,4-diphenylbutadiene (DPB) case, it is however noted that the impact of the strongest one-mode anharmonicity (mode 70) on the vibrational profile is larger than the mode–mode coupling contributions (see Figure 23-a). The impact of reduced-space anharmonicity holds in particular for the intensities, but also the energy differences are shifted. This shift in intensities mostly concerns an increase of the peak around 3.1 eV and a decreased intensity, of the peak around 3.5 eV when considering reduced-space anharmonicity. This shift leads to a better qualitative agreement with the experimental spectrum, similar to the conclusions drawn on the effect of full-space ground-state anharmonicity in reference [306]. A high degree of agreement with experimental data demonstrates the reliability and accuracy of the VCI-in-IMDHO scheme for predicting the emission spectra of these types of molecules.

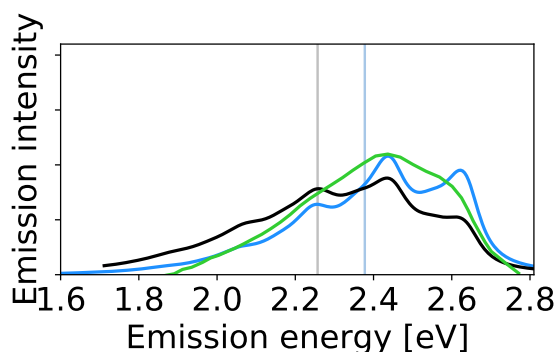


**Figure 23:** Calculated emission spectra of 1,4-diphenylbutadiene molecule (a) and the corresponding mode ranking (b) with CAM-B3LYP/def2-TZVP. See the caption of Table 1 for more details. **(a):** Calculated harmonic (blue), reduced space VCI (red), one mode VCI-in-IMDHO (orange), VCI-in-IMDHO without mode–mode couplings in the PES (magenta) and VCI-in-IMDHO (black) broadened emission spectra of HS84 molecule with a HWHM of 0.04 eV compared to a corrected experimental spectrum in hexane [306, 326](green). Vertical electronic emission energy with grey lines. The VCI spectrum is obtained by conventional Lorentzian broadening. In the bottom part, the calculated spectra are directly compared to the best approximation (VCI-in-IMDHO) and the experimental spectrum, respectively. Additionally, the deviation between the calculated spectra according to equation 81 is quantified in the respective graphs. The modes treated by VCI in the reduced-space VCI are 70, 53, 69, 41, 60, 57, and 64, out of which only the modes 70 and 53 were coupled in the underlying PESs. Only mode 70 is VCI treated in the one-mode VCI-in-IMDHO treatment. The experimental spectra are shifted by 1.549 eV to get the same position of maximum as VCI-in-IMDHO spectra.

anthracene diimide (ADI) was chosen as the final test system, as shown in Figure 17. The findings indicate a moderate impact of anharmonicity on ADI’s emission spectrum. This difference is lessened in the VCI-in-IMDHO treatment. The

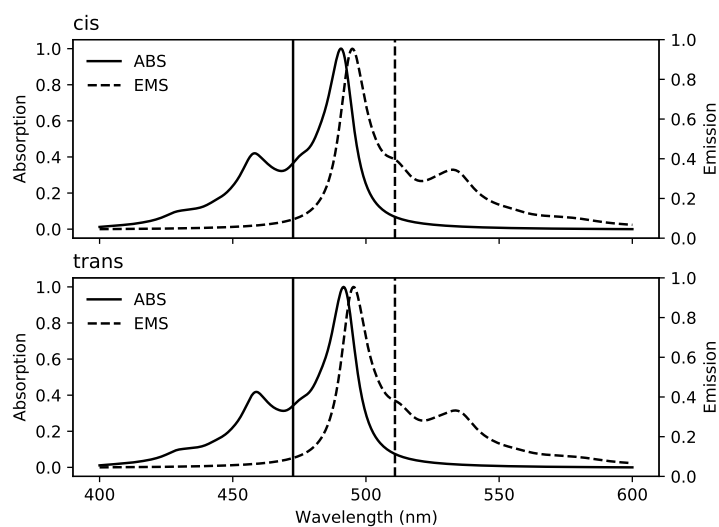
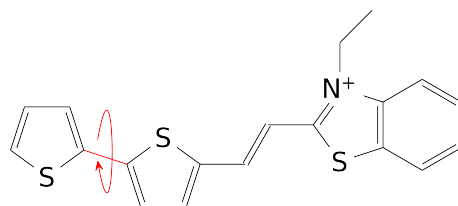
effect of mode–mode couplings in ADI is more significant than in previous examples.

As optical spectra measurements are typically performed in a solution, the positions of certain bands, as well as the tuning of their shapes by vibronic transitions are found to be sensitive to changes in the solvent [267]. Rough estimates of the effect of the solvent environment on the vibrational progression were made by performing pure IMDHO computations using the polarizable-continuum model (PCM) with the respective solvent that was used in the corresponding experimental spectrum for each dye. It was found that the impact of anharmonicities on the vibrational line shape is similar to that of the solvent, with the exception of a strong sensitivity of the pure IMDHO spectrum of HS84 when solvent effects are included (see Figure 24). However, all simulated spectra show significant discrepancies compared to the experimental reference.



**Figure 24:** *Calculated IMDHO harmonic emission spectra of HS84 in vacuum (blue), in water estimated with the PCM model (black) and a HWHM of  $\gamma = 0.04$  eV using CAM-B3LYP/def2-TZVP compared to a corrected experimental spectrum in PBS[49] (green). The vertical transition energy in vacuum is given as a vertical grey line. The experimental and the IMDHO spectrum in solvent as well as its corresponding excitation energy (vertical blue light line) are shifted by 0.130 eV and 0.211 eV, respectively, to get the same position of maximum as IMDHO harmonic spectra in vacuum. Additionally, the deviation between the calculated spectrum in vacuum and in solvent according to Equation 81 is 0.211 and 0.343, respectively.*

To further evaluate the effectiveness of this method, the calculation of absorption and emission spectra of different conformers of a novel fluorescence probe, bTVBT4, was conducted in gas phase (see Figure 25). It results in the identification of three vibrational stretching and angle-bending modes in the region of 1300–1600  $\text{cm}^{-1}$  to be of prime importance for the absorption spectrum profile. The 0–0 transition dominates the absorption spectrum, leading to a strongly inhomogeneous broadening [177]. With the VCI-in-IMDHO model, an accurate description of the nuclear motions of this ligand is provided, which is crucial for understanding the dynamics of the probe in more complex situations, such as in water at room temperature.



**Figure 25:** Computational vibrationally resolved absorption and emission spectra for *cis* and *trans* bTVBT<sub>4</sub> in vacuum (CAM-B3LYP/aug-cc-pVDZ). For the *cis* reference structure, three modes were included anharmonically in the absorption calculation and six modes for the emission. For the *trans* reference structure, three modes were treated anharmonically accordingly in the absorption and seven modes in the emission spectra calculation. All remaining vibrational modes were treated by the IMDHO model [177].

In conclusion, with the VCI-in-IMDHO model, a hybrid treatment for vibrational line shapes of optical spectra is introduced. This model uses a combination of highly accurate and anharmonic vibrational wave function methods (VCI) for selected modes and a very approximate IMDHO model for the remaining modes, providing a balance of accuracy and computational efficiency. For the separation of the modes into these two sets, the effect on the vibrational progression, previously utilized in previous research [32], was employed. Additionally, the resolution criterion  $\sigma_i$  was refined, excluding modes based on the expected poor resolution of the vibrational modes, rather than relying solely on a frequency criterion.

To assess the impact of anharmonicity a priori, a relation between the IMDHO displacements ( $\Delta_i$ ) and the anharmonic displacements ( $\Delta_a$ ) was proposed. This was achieved by projecting the coordinate difference of the ground and excited state onto the respective vibrational coordinates. Unfortunately, a clear correlation between this relation and the impact of anharmonicity on the one-mode vibrational profiles, especially for the anthracene diimide, could not be found. Thus, the computationally more expensive comparison of the effect of the one-mode anharmonicity was used for further studies. Still, with the presented VCI-in-IMDHO model, a reduction in computational cost is achieved compared to the reduced-space VCI approach [32]. This scheme can be a valuable tool for predicting the absorption and emission spectra of a wide range of fluorescence probes, and it is expected to be useful for the design and optimization of new fluorescence probes in the future. In addition, the effect of solvation was investigated using continuum solvation models. In most cases, the impact of anharmonicity and solvation effects are similar, except for HS84. This suggests that both effects need to be taken into account for accurate computations. The combination of accurate vibrational profiles with different environmental descriptions will be explored in the next section.

ENVIRONMENT EFFECT ON THE VIBRATIONALLY  
RESOLVED EMISSION SPECTRA

---

*This chapter contains sections sourced from a manuscript that is currently being prepared*

**T.M.N. Nguyen** and C. König.

*“The role of micro-solvation on the computed emission spectra: The case of oxazines.”*

Solvent plays a crucial role in the shape of the vibronic spectra of molecules. It influences the energy levels and transitions of the electronic states, leading to shifts in the absorption and emission wavelengths. Additionally, the solvent environment can affect the vibrational modes and their coupling with electronic transitions, resulting in changes in the overall spectral shape and intensity. [267] In the previous chapter, the impact of solvent on the vibronic spectra of HS84 was explored using the polarizable continuum model (PCM). The investigation revealed a significant sensitivity of the pure IMDHO spectrum of HS84 when solvent effects were taken into account. This highlights the crucial role of solvent in shaping the vibronic spectra and emphasizes the need to consider solvent interactions in accurate modeling and interpretation of optical properties. Similarly, other dyes like oxazine molecules exhibit notable solvchromism and are sensitive to their immediate chemical environment. [327–329] These characteristics make them promising fluorescent probes for a wide range of biological applications, [57, 63–65, 67, 68, 330, 331] including the detection of amyloid aggregation. [68]

Despite notable progress in electronic theories for molecules in solution in the last two decades, [29, 268, 269] effectively and accurately addressing solvation remains a challenging task in molecular modeling, even though it holds paramount importance. Implicit solvent models such as the PCM, [265–267] which capture mean-field solvent effects on the spectral position and associated vibronic structure, provide a computationally efficient approach. However, these models may not accurately account for the explicit molecular interactions between the solute and solvent molecules. [29, 268, 269] In contrast, explicit solvent models, such as the cluster model, provide a more comprehensive depiction of solute-solvent interactions. However, excited state calculations with explicit solvent molecules using the cluster model present several challenges, including increased computational costs due to the larger system size and the need to sample diverse solute-solvent configurations. [29, 268, 269] This limits the feasibility of the method to a

small number of solvent molecules. It should be emphasized that the influence of solute-solvent interactions is most notable within the first solvation shell, while their impact diminishes rapidly with increasing distance. Hence, for long-range interactions, a more approximate solvent model can often be employed.[29]

In this project, a microsolvation model was employed to capture the effects of the immediate solvation shell on the vibrationally resolved fluorescence spectra of oxazine dyes (c.f. Figure 3), namely nile blue (NB+), oxazine 1 (Ox1+), darrow red (DR+), crestyl violet (CV+), oxazine 4 (Ox4+), oxazine 170 (Ox170+). In section 6.1, the validation of the computational setup for the isolated oxazines in the gas phase is provided. Following that, in section 6.2, the impact of individual water molecules on Ox4+ and Ox170+ in both the gas phase and continuum water solvation environments is examined. Furthermore, in section 6.3, the influence of individual betaine molecules at different positions on the spectral shape is discussed.

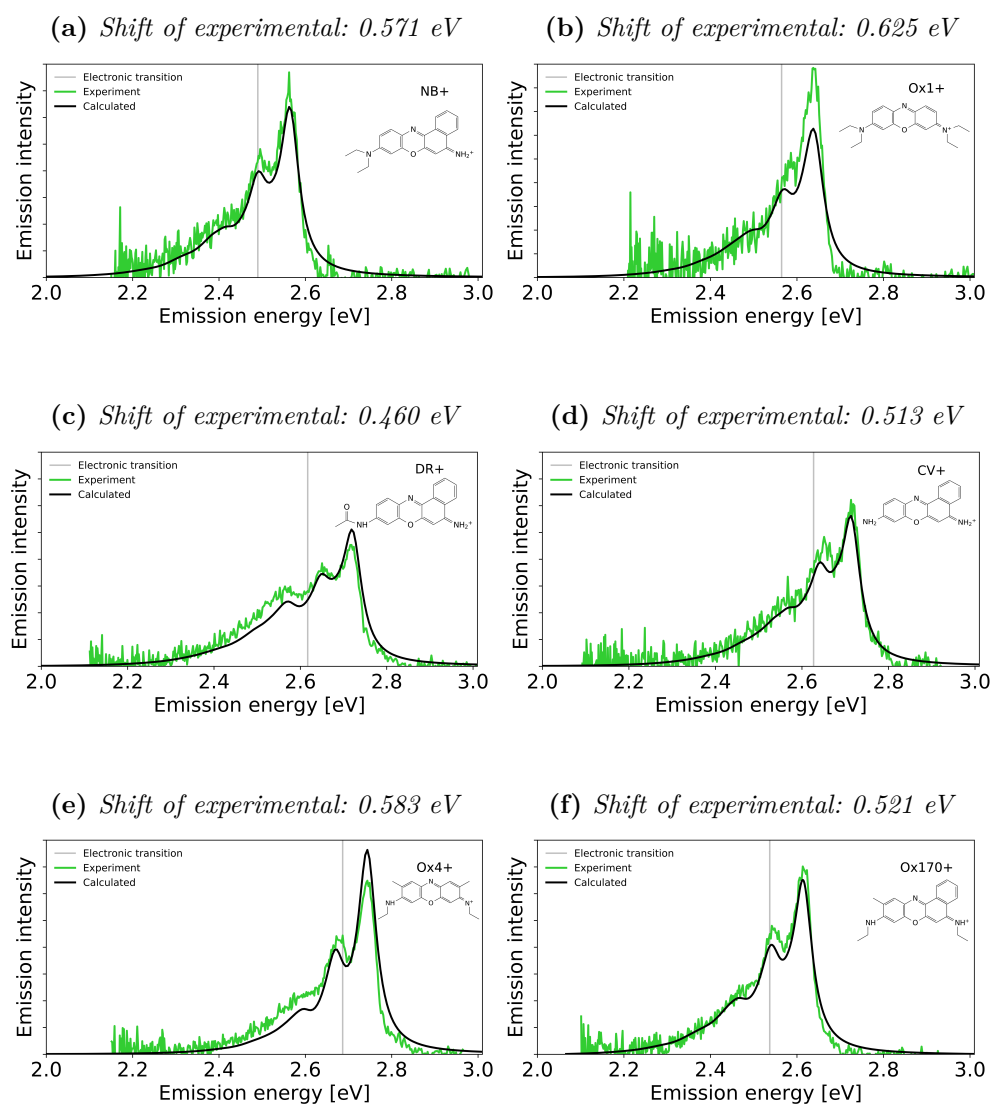
## 6.1 ISOLATED OXAZINE DYES

The computational setup was first validated by comparing the results with experimental data [332] obtained in the gas phase for isolated oxazines. The molecular structure of the investigated oxazines comprises a benzophenoxazine ring and amino-based side chains that can adopt various conformations. To determine the local energy minima, a conformational analysis was performed, and the resulting conformers were optimized using the CAM-B3LYP/def2-TZVP method. The emission spectra using the IMDHO lineshape of highly populated conformers exhibit remarkable similarity in the gas phase. This indicates that the electronic structure remains conserved across different conformers, even though they may differ in their structural arrangements. The calculated vibrationally resolved emission spectra of the predominant conformer for each oxazine molecule, presented in Figure 26, are compared with experimental reference spectra obtained in vacuum [332]. The close agreement between the calculated and experimental spectra confirms the accuracy of the computational methods employed in this study for analyzing the vibronic spectra of the oxazine molecules. This validation further strengthens the reliability of IMDHO computational approaches in the computation of the spectral lineshape in these cases.

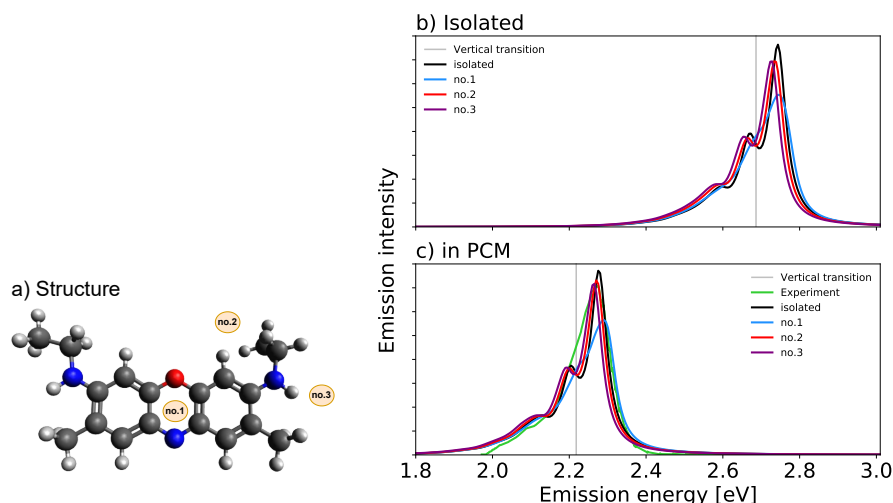
## 6.2 EFFECT OF INDIVIDUAL WATER MOLECULES

A deeper understanding of how vibronic emission is influenced by the microenvironment is sought by specifically examining Ox4+ and Ox170+ using two types of solvent molecules: water and betaine. As being the most prevalent molecule in biological systems, the former plays a crucial role in influencing the photophysical properties of fluorescent dyes within complex environments. The microsolvation





**Figure 26:** Calculated computational (of the most stable conformer, black) and experimental<sup>[332]</sup> (green) emission spectra in vacuum of a: nile blue (NB+), b: oxazine 1 (Ox1+), c: darrow red (DR+), d: crestyl violet (CV+), e: Ox4+, f: Ox170+ using CAM-B3LYP/def2-TZVP. The experiment spectra are shifted as shown in the respective caption to get the same maximum position as the calculated spectrum of the respective most stable conformer.



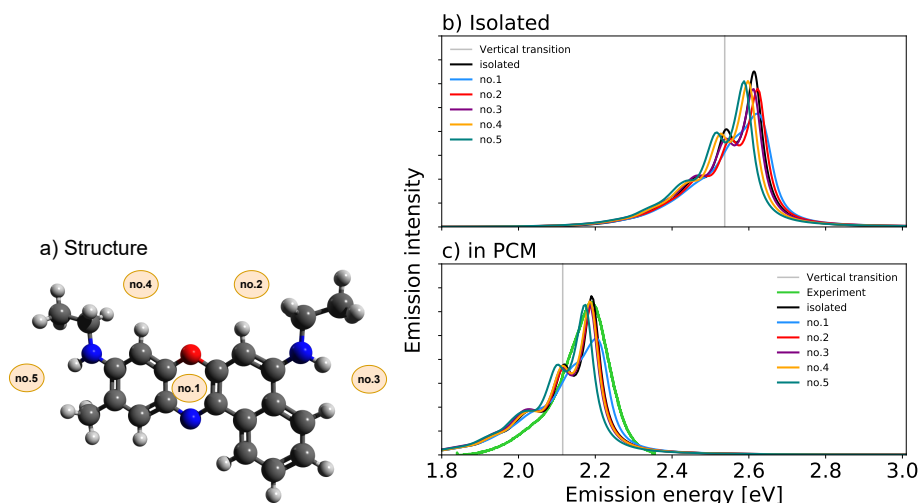
**Figure 27:** Optimized structure (a) and emission spectra of isolated (black), with the inclusion of a single water molecule at various positions no.1, no.2, no.3 (color) in vacuum (b), and in PCM (c) of Ox4+ using CAM-B3LYP/def2-TZVP with HWHM of 0.025 eV. The experimental spectra in water of Ox4+ [333] is shifted by 0.317 eV to get the same maximum position as the calculated spectrum of the isolated molecule in PCM.

model was employed, which involves selecting a small representative cluster consisting of the solute molecule surrounded by a few solvent molecules to capture the solute-solvent interactions within the local environment. One solvent molecule was positioned at various locations around the dye (see Figure 27 and 28) to investigate its influence on the system's properties.

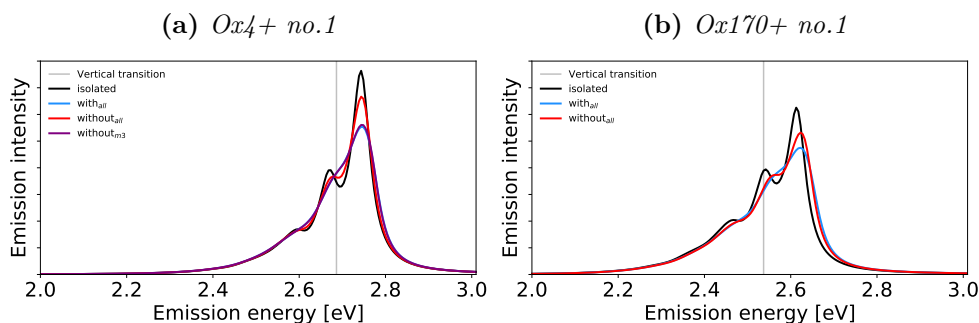
The calculated emission spectra for the bare dye (black) and the water-dye complexes (color) are also depicted in Figure 27 and 28. The emission spectra of both molecules show a slight blue shift in the presence of water-dye complexes with minimal alteration in the vibrational profile. However, when the water molecule is coordinated in a side-on position over the oxazine ring (position no.1), a larger change in the calculated emission profile is observed, with broadening of the peak and blurring of the shoulder pattern. These findings highlight the significant impact of the specific position of the water molecule on the vibronic emission properties.

The impact of additional vibrational modes from microsolvation on the vibrational profile of Ox4+ and Ox170+ was examined. Except for location no.1 (see in Figure 29), the influence of solvent modes on the vibronic spectra was found to be insignificant. In the case of Ox4+, this impact was attributed to mode 3, which exhibited large  $\Delta_i$  and  $\sigma_i$  values (such assignment is not as clear for Ox170+). It suggests that the key indicators obtained from the IMDHO model can also serve as an indication of the impact of environmental modes.

Additionally, the influence of attaching two water molecules at position 1 on the vibronic spectra of Ox4+ and Ox170+ is examined (Figure 30), revealing that



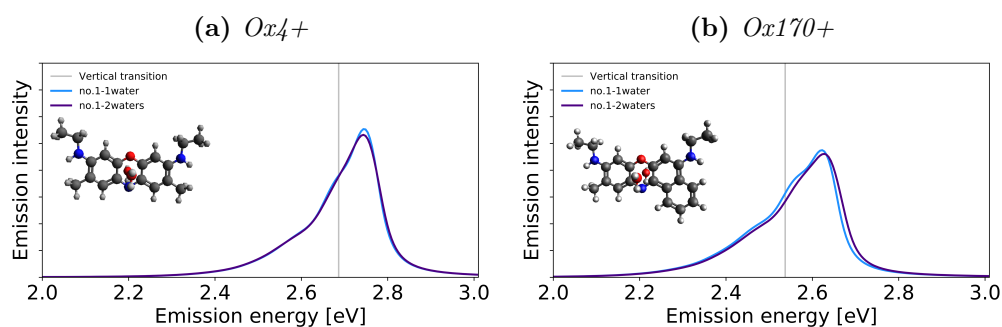
**Figure 28:** Optimized structure (a) and emission spectra of isolated (black), with the inclusion of a single water molecule at various positions no.1, no.2, no.3, no.4, no.5 (color) in vacuum (b), and in PCM (c) of Ox170+ using CAM-B3LYP/def2-TZVP with HWHM of 0.025 eV. The experimental spectra in water of Ox170+ [334] is shifted by 0.290 eV to get the same maximum position as the calculated spectrum of the isolated molecule in PCM.



**Figure 29:** Emission spectra of Ox4+ (a) and Ox170+ (b) with one water molecule at position no.1, illustrating the following scenarios: isolated dye (black), including all solvent modes (blue), excluding all solvent modes (red), and excluding the most significant water mode, namely mode number 3 ( $\omega_i = 49 \text{ cm}^{-1}$ , purple).

the number of water molecules at this position has negligible effects on the shape of the oxazine spectra.

Furthermore, the application of the PCM model in the calculation of spectra for the solvent-solute clusters is investigated. It results in a blue shift compared to spectra in vacuum, thus improving the agreement with experimental data in aqueous solutions. The overall vibrational profiles of the emission spectra remain largely similar between the vacuum and PCM calculations, both for isolated and microsolvated scenarios. Notably, the presence of a water molecule at the on-ring position (position no.1) exhibits a more pronounced effect on the vibrational line

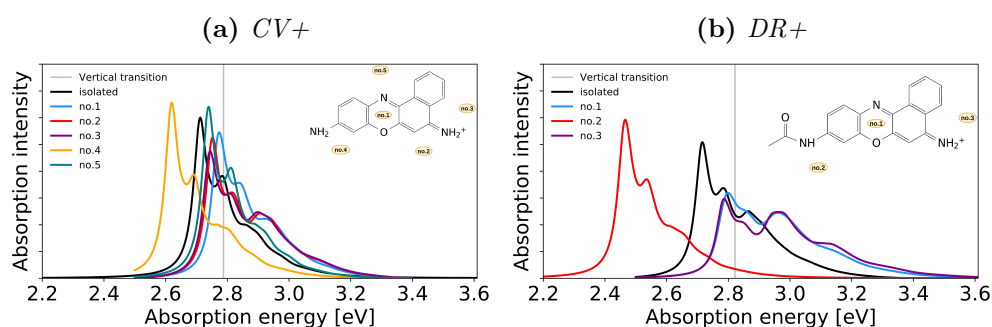


**Figure 30:** Comparison of the emission spectra of a)  $Ox4+$  and b)  $Ox170+$  with one water (blue), and two water molecules (violet) at position no.1 in vacuum.

shape, surpassing the impact of the continuum model (see Figure 27.c and 28.c). Importantly, the band shape obtained in this configuration bears resemblance to the experimental data, featuring a distinct and dominant main peak.

### 6.3 EFFECT OF INDIVIDUAL BETAINES MOLECULES

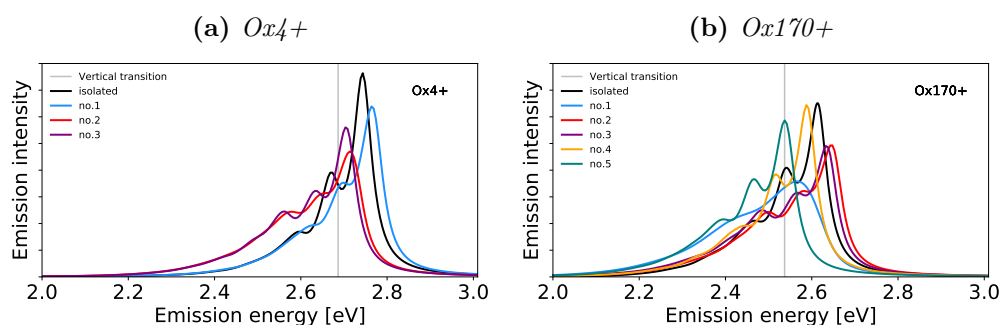
Additionally, organic compounds like betaines, characterized by a substantial dipole moment of 11.9 D, are valuable for stabilizing biological macromolecules [335, 336] and exploring the environmental sensitivity of fluorescent dyes [337]. A test study conducted by Nils Weber demonstrated the substantial influence of betaine on the absorption spectra shape of  $CV+$  and  $DR+$  molecules, as illustrated in Figure 31. Notably, the observed effects exhibited variations based on both the position of the solvent and the specific solute molecule involved. The findings suggest that betaine could have the potential to exert a substantial influence on the emission spectra of oxazine molecules.



**Figure 31:** Absorption spectra of isolated (black) a)  $CV+$  and b)  $DR+$  molecule, with one betaine cluster (color) in vacuum.<sup>1</sup>

<sup>1</sup> These calculations were performed by Nils Weber in a research internship under my supervision, Leibniz Universität Hannover, 2022–2023.

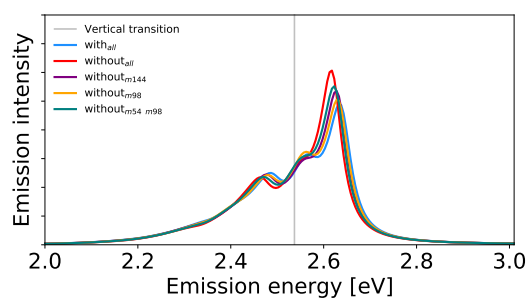
The impact of betaine on the emission spectra of Ox4+ and Ox170+ was then explored, anticipating a more pronounced effect due to the larger dipole moment of individual betaine molecules when attached to the oxazine dyes at the same positions. As shown in Figure 32, the presence of the betaine molecule induces a more pronounced blue shift in the emission spectra compared to water, and the overall band shape is noticeably altered for all betaine locations. In addition to the peak position, the most significant difference is a general broadening effect, for instance, with the most prominent broadening observed at location no.2 for Ox4+, followed by locations no.3 and no.1.



**Figure 32:** Emission spectra of isolated (black), with one betaine molecule (color) in vacuum using HWHM of 0.025 eV of Ox4+ (a) and Ox170+ (b) at CAM-B3LYP/def2-TZVP with solvent positions (no.1, no.2, no.3, no.4, no.5).

Furthermore, considering the analysis of the influence of additional vibrational DOFs on the vibrational profile, it can be observed that, for all Ox4+–betaine and the Ox170+–betaine complexes with betaine located at positions no.4 and no.5, the effect of the additional vibrational DOFs is minimal. So, it can be concluded that the observed changes in the vibrational profile caused by the aggregation of the betaine molecule are primarily attributable to electronic effects. In contrast, the Ox170+ with betaine at position no.3 exhibits the most significant effect on the spectral line shapes. Notably, among all the Ox170+–betaine complexes, only at position no.3 are vibrational modes (mode 54 and mode 98) identified with  $|\Delta_i| > 0.6$  and  $\sigma > 200 \text{ cm}^{-1}$ , which can be assigned to betaine modes. These modes account for approximately half of the observed impact on the betaine modes within the vibronic spectrum as shown in Figure 33.

In summary, the calculated harmonic spectra of oxazine molecules in a vacuum demonstrated good agreement with experimental results, validating the accuracy of the IMDHO model in predicting the emission spectra of oxazines. The study revealed that the position of the water molecule, especially when located at position no.1 on top of the 1,4-oxazine ring, had a notable influence on the spectra, resulting in broadened main peaks caused by a specific solvent mode. Moreover, the presence of betaine, with its large dipole moment, induced a larger blue shift in the emission spectra compared to water, significantly altering the band shape. These findings underscore the significant impact of treating explicit solvents including their DOFs for accurate calculation of vibrationally resolved emission spectra.



**Figure 33:** Calculated IMDHO emission spectra of *Ox170+* with the inclusion of a single betaine molecule at various positions using CAM-B3LYP/def2-TZVP. The spectra include all betaine modes (blue) and exclude betaine modes (red), mode 54 (purple), mode 98 (orange), and both modes (green).

MOLECULAR DYNAMICS AND BINDING ENERGY OF  
PYRENE-BASED FLUORESCENT AMYLOID LIGAND TO  
TRANSTHYRETIN

---

*This chapter includes sections originating from the submitted manuscript:*

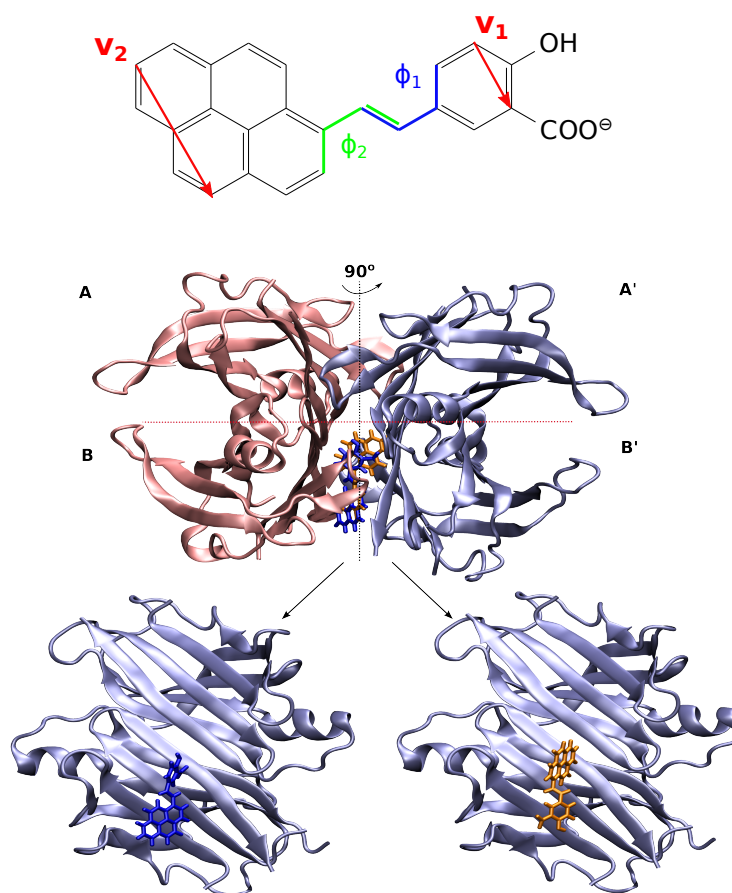
**T.M.N. Nguyen**, A. Begum, J. Zhang, P. Leira, Y. Todarwal, M. Linares, P. Norman, D. Derbyshire, E. von Castelmur, M. Lindgren, P. Hammarström and C. König, “*Binding of a pyrene-based fluorescent amyloid ligand to transthyretin: A combined crystallographic and molecular dynamics study*”, J. Phys. Chem. B., 2023, (Under revision)

*with some modifications.*

To facilitate the effort in early diagnosis and finding improved small molecule ligands as kinetic stabilizers of neurogenesis biomarker proteins, in this project, transthyretin (TTR) is used as a research platform to understand small molecule recognition and binding specificity. Despite a large number of solved structures, the binding modes of TTR ligands are still a matter of intense research. Previous work has identified a pyrene-based trans-stilbene ligand with a salicylic acid moiety (Py1SA) as an amyloid fibril probe for several different amyloid proteins. [80] Based on steady-state and time-resolved fluorescence spectroscopy at room temperature, in addition to fibrils, it shows that Py1SA can also bind to the native state of TTR through its T4 binding sites, also known as hydrophobic halogen binding pockets (HBPs). The crystallographic data of the complex is, however, not of sufficient resolution to unambiguously demonstrate the binding mode, mainly due to partial occupancy. That means the most likely orientation of the ligand in the binding pocket could not be resolved experimentally. Therefore to complement the experimental study molecular dynamics (MD) and umbrella sampling (US) calculations were performed. The computational details can be found in chapter 4.

## 7.1 MOLECULAR DYNAMICS SIMULATIONS

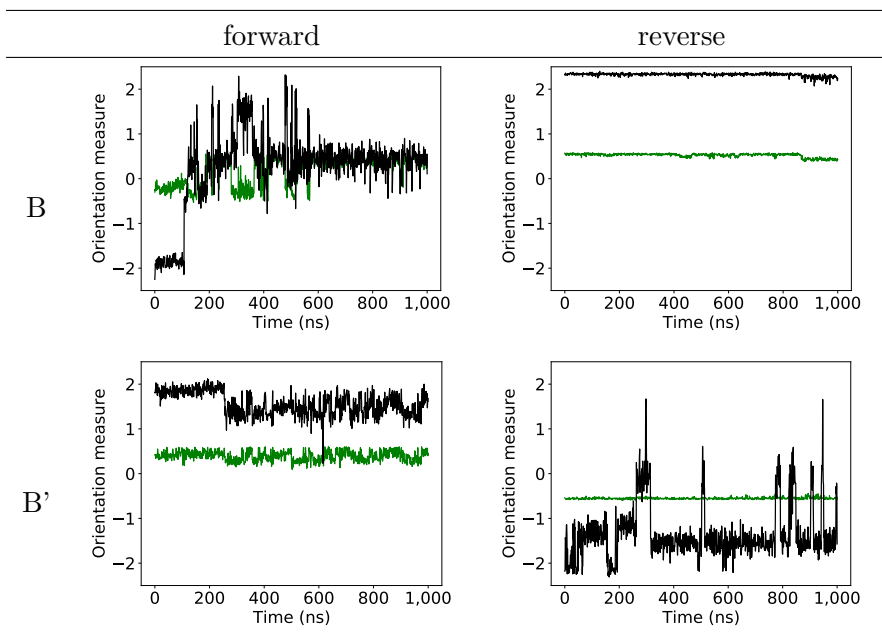
To clarify the uncertainty in the interpretation of X-ray data, MD simulations were performed using Gromacs 2019.3 [307] with the Amber ff14SB force field [114] for the TTR protein, TIP3P [103] for water, and a re-parametrized General Amber Force Field (GAFF) [308, 309] for Py1SA against B3LYP/6-31G(d,p) data. In the 1  $\mu$ s MD simulations, all atoms except for the ones that are within 4 Å



**Figure 34:** *Top: Schematic representation of Py1SA ligand with internal (dihedral  $\phi_1$ ,  $\phi_2$ ) and external ( $v_1$ ,  $v_2$ ) descriptors. Bottom: Possible structures of the TTR-Py1SA complex. The TTR tetramer is composed of monomers A, A', B, and B' assembled together around the 2-fold c-axis (dashed line). Py1SA (in sticks) binds in the T4 binding site in different modes. Forward is defined as when the pyrene moiety is mainly buried inside the HBP (orange); whilst in the reverse case, this moiety points away from the protein pocket (blue). Since the B and B' are symmetry equivalent, only the B' mode is shown.*

from the ligands are under harmonic constraints. As initial structures, four different possible structures obtained from the refinement of the X-ray data labeled forward-B, forward-B', reverse-B, and reverse-B' are used. The structures of the TTR-Py1SA complex in the binding cavity or the HBPs [73, 338] are displayed in Figure 34. Py1SA stabilizes different modes (forward/reverse) since it binds to one of the HBPs of TTR in various orientations. The forward mode is characterized by the pyrene moiety being primarily buried inside the HBPs, while the reverse mode is defined by the rotation of the ring away from the protein pocket. Note that forward-B and forward-B' as well as reverse-B and reverse-B' represent slightly different structures of symmetry-equivalent binding modes. By projecting the external descriptors (such as  $v_1$  and  $v_2$  in Figure 34) onto the reference vectors of the reverse-B mode during the simulation, it is demonstrated that there

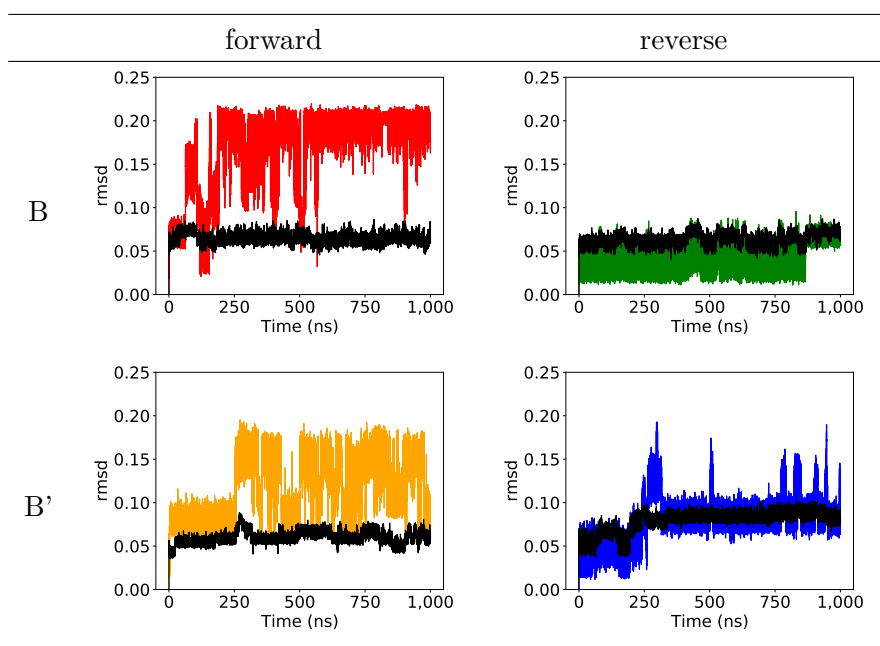




**Figure 35:** Projection of vector 1 ( $v_1$ , green, showing the orientation of the salicylic acid group) and vector 2 ( $v_2$ , black, showing the orientation of the pyrene group) on the respective reference vectors of reverse-B conformer during the simulation time as a measure of orientation. For the definition see Figure 34.

is no change in the orientation of both the salicylic acid group and the pyrene group (see Figure 35). This indicates that there is no rotation of the ligand inside the binding pocket. The relatively large change in orientation measures in the forward-B trajectory is accompanied by an unbinding/binding process. The observed rotations of the pyrene group in the reverse-B' trajectory represent a rotation of this group at the edge of the pocket, while the orientation of the more buried salicylic acid group is very stable in the binding pocket. Furthermore, the dihedral  $\phi_1$  and  $\phi_2$  exhibit few changes over the simulation time. Notably, a tendency to simultaneous switches of both dihedral angles for instance for forward-B' at around 250 ns and for reverse-B at about 850 ns is observed. This simultaneous switch may be favored over individual switches as the molecular shape remains similar during this procedure.

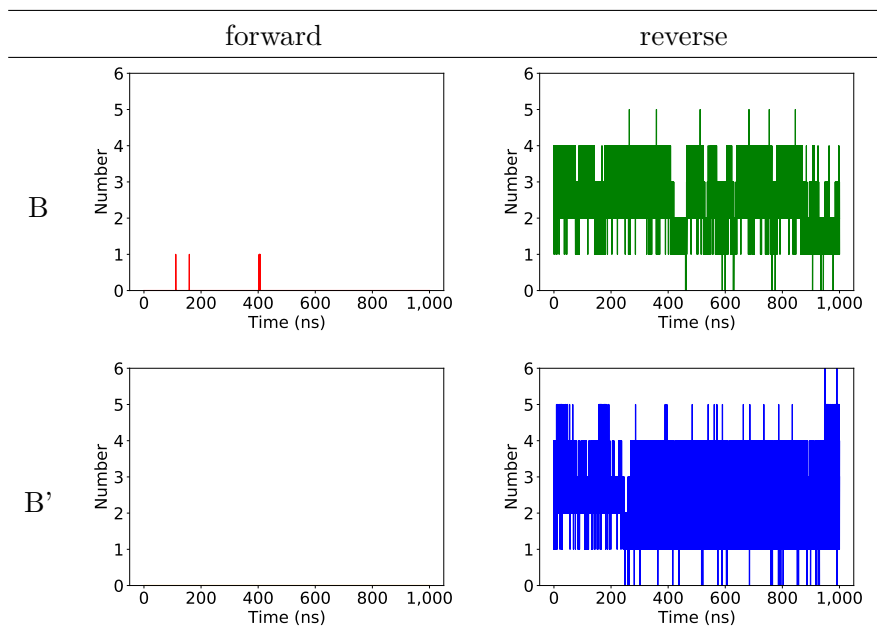
To evaluate the mobility of each protein–ligand complex, the root-mean-square deviation (RMSD) was calculated for the TTR-4 Å-pocket and its ligand as a function of time. As shown in Figure 36, the RMSD of the pocket is less than 0.1 Å for all modes within the full simulation, as was expected due to the position constraints of the surrounding protein. The results of the MD simulation analysis also show that the RMSD for the dye in reverse modes is smaller than the forward modes, especially for the reverse-B mode with the RMSD being smaller than 0.05 Å. By contrast, for the forward modes, deviations between the ligand structure and the corresponding crystal suggested a high alteration of the structures over the simulation course.



**Figure 36:** *Root mean square deviation (RMSD) of Py1SA (red for forward-B mode with respect to the respective crystal structure, green for reverse-B mode, orange for forward-B' mode, blue for reverse-B' mode) and pocket (black for protein) from 1  $\mu$ s simulation.*

To further investigate the binding sites, the number of hydrogen bonds (H-bonds) between Py1SA and the pocket of four selected structures was calculated using the MD simulation trajectories as shown in Figure 37. Overall, the number of H-bonds of the reverse-B mode is similar to that of the reverse-B' mode, while being significantly larger in that of the forward modes. This is expected because TTR-bound Py1SA in the forward modes is mostly oriented with the pyrene ring inside the cavity, which does not form H-bonds with the pocket. Differently from the forward modes, the reverse modes show a higher number of H-bonds, which can assist the stability of these modes. Notably, in the reverse modes, the residues Ser-117 and Thr-119 are particularly important since they are involved in the ligand-binding stabilization in which their hydroxyl groups form a hydrogen bond to both of the hydroxyl groups of the phenyl ring of the ligand. This is consistent with previous studies that have shown these residues are capable of forming hydrogen bonds with the natural ligand, T4, to stabilize the TTR-T4 complex [73, 339].

Throughout the MD trajectory, the interaction energy of the Py1SA ligand with the surrounding protein residues within 4  $\text{\AA}$  was calculated. The Lennard-Jones short-range (LJ-SR), Coulombic short-range (Coul-SR) potential, and their total sum are listed in Figure 38. The average attractive contribution to the interaction energy of LJ-SR in the forward modes (-116 kJ/mol for forward-B, -175 kJ/mol for forward-B') is greater than that of the Coul-SR potential (-57 kJ/mol and 9.8 kJ/mol, respectively). Nevertheless, the contribution of these interactions in the



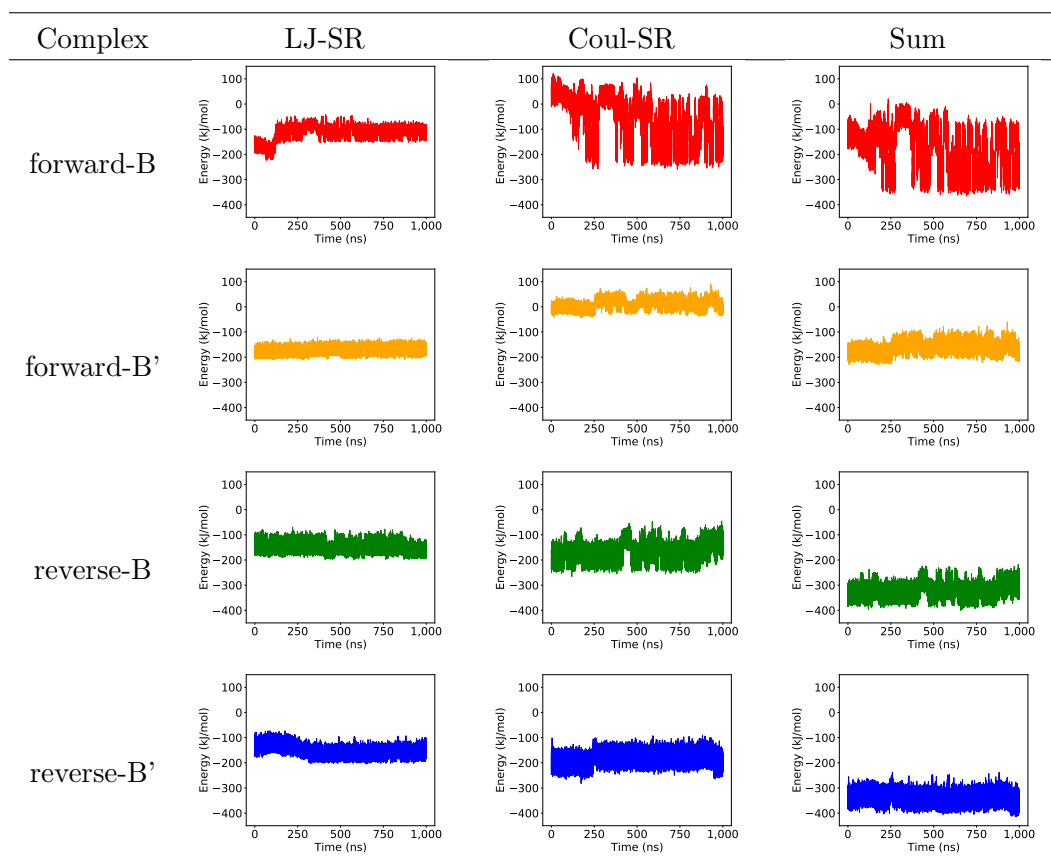
**Figure 37:** Number of hydrogen bonds during 1  $\mu$ s simulation (red for forward-B mode, green for reverse-B mode, orange for forward-B' mode, and blue for reverse-B' mode), as defined by the Gromacs utilities. Significantly more hydrogen bonds between Py1SA and the 4 Å-pocket for the reverse mode than for the forward mode is observed.

reverse modes is very similar. In general, the overall interaction energy for the forward mode is higher than that for the reverse one by about 160–170 kJ/mol.

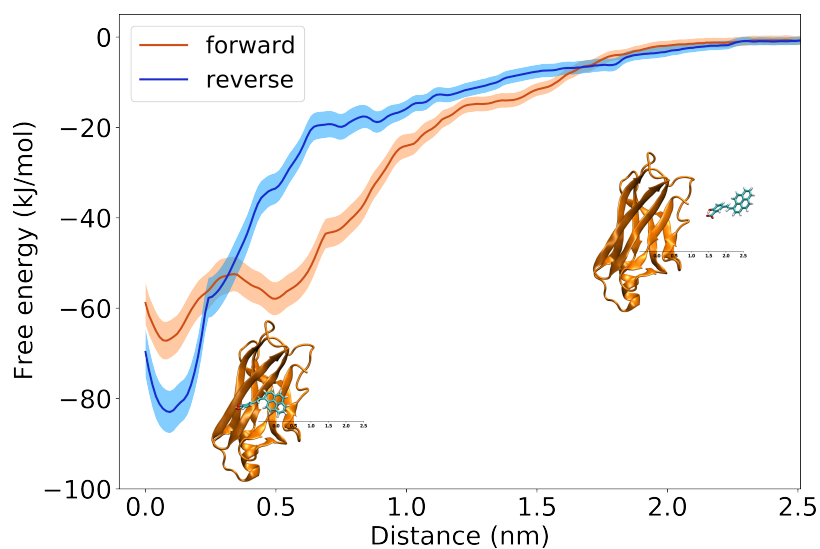
## 7.2 UMBRELLA SAMPLING SIMULATIONS

From Figure 38, it is also noticed that the binding energy for the reverse mode is more favorable than that for the other forward trajectory by around 160–170 kJ/mol. However, this difference in binding energy does not conclusively indicate the actual free binding energies. Therefore, US simulations were additionally performed using Gromacs version 2021.3 to construct the potential of mean force (PMF) surface starting from all four initial structures.

To study the dissociation pathways of the Py1SA–TTR complex, umbrella sampling simulations are used to construct the PMF surface. As reaction coordinate, the center-of-mass (COM) distance between the pyrene or benzene group of the Py1SA ligand and 4 Å-pocket for the reverse and forward modes, respectively, is chosen. The ligand is free to move on the sphere with a radius of the COM distance in the corresponding window. In the US simulations, all backbone movements were constrained. In this section, the results from the forward-B' and reverse-B' modes are presented, while others are excluded due to their abnormal events. In Figure 39, the averaged PMFs profiles for the forward and reverse modes



**Figure 38:** Short-range interaction (Lennard-Jones short-range (LJ-SR), Coulombic short-range (Coul-SR)) energy between ligand and protein during 1  $\mu$ s MD simulations.



**Figure 39:** Free energy profiles (potentials of mean force) for Py1SA in the forward (orange, binding free energy =  $-65$  kJ/mol) and reverse (blue, binding free energy =  $-85$  kJ/mol) modes obtained by the potential of mean force approach by pulling the ligand from the binding site to become free in solution.

are obtained by evaluating the distance histograms of the simulations by employing the weighted histogram analysis method (WHAM) algorithm [164, 165]. The binding free energy for the forward mode was calculated to be 65 kJ/mol, while the reverse mode had a binding free energy of 80–85 kJ/mol. This suggests that the reverse mode is more stable and dominant, which aligns with the findings from the MD simulations and previous studies on the importance of carboxylic acid and its position in the ligand structure for activity [74].

In summary, for diagnosing and monitoring patients with amyloidosis, TTR in TTR amyloidosis is a particularly interesting target due to its abundance of accessible blood samples and numerous successful available treatments. Despite the fact that this protein and its associated diseases have been researched for a long time, there is currently no approved TTR-based biomarker to distinguish misfolded TTR from native tetrameric TTR. Additionally, detailed knowledge of the binding modes of various small molecule ligands towards the native TTR tetramer can lead to new kinetic stabilizers as alternatives to currently approved drugs. The native TTR tetramer with its intrinsic symmetry poses a challenge for X-ray crystallography. One of the important features of this work thus is the interaction of experiment and computation in studies of protein aggregation, demonstrating the power of computational simulations as analytical tools to obtain and digest information from nature and to use it to make useful predictions. The results of these calculations provide a detailed understanding of the binding mechanism and the interactions between the Py1SA ligand and the protein, which could be useful in the development of new therapeutics for the treatment of amyloidosis caused by wild-type transthyretin (TTRwt).



SUMMARY, CONCLUSIONS, AND OUTLOOK

---

The aggregation of amyloid- $\beta$  ( $A\beta$ ) into oligomers, fibrils, and plaques plays a central role in the molecular pathogenesis of several misfolding diseases such as Alzheimer's disease (AD). It is the main target of AD drug development. To gain a more comprehensive understanding of  $A\beta$ 's role in disease pathogenesis and to track the effectiveness of drugs targeting  $A\beta$  in clinical trials, it is essential to have biomarkers that can directly monitor  $A\beta$  metabolism and aggregation in patients. The theoretical description of complex systems, such as those found in amyloid aggregation, is highly challenging mainly due to the size of the systems, which far exceeds the capacity of standard quantum-chemical methods. Nowadays, researchers are constantly seeking to improve the accuracy and computational efficiency of these models, as well as to develop new methods to simulate larger and more complex systems.

This thesis addresses two interrelated aspects of developing amyloid biomarkers. Firstly, it involves the computation of vibrationally resolved spectra of small fluorescent dyes to detect amyloid aggregation in gas and condensed phases. Secondly, it examines the binding and unbinding processes of a novel ligand to the target protein.

The computational study of vibronic spectra of small fluorescent dyes used to detect protein aggregation is crucial. However, the inclusion of anharmonic vibrations on the spectral line shape poses a significant challenge in the computation of optical spectra. Calculating vibrationally resolved spectra with an anharmonic vibrational description is only practical for small systems with a few degrees of freedom. For medium to large-sized systems, computational cost constraints lead to a need to sacrifice accuracy, resulting in more qualitative descriptions. In chapter 5, a hybrid model for vibrational line shapes of optical spectra, called VCI-in-IMDHO, is introduced. This model enables the treatment of selected modes using highly accurate and anharmonic vibrational wave function methods (VCI) while treating the remaining modes using the approximate IMDHO model. No mode-mode coupling is considered between the two sets of modes or within the modes treated by IMDHO. The Franck-Condon (FC) factors for the reduced vibrational space obtained by anharmonic VCI are broadened using a system-specific line shape derived from the IMDHO autocorrelation function, with a given half-width at half maximum considering all modes except the ones treated by VCI.

The criteria introduced in reference 32 for separating modes into two sets have been refined and extended. These criteria include the impact on the vibrational progression ( $|\Delta_i|$ ) estimated by the displacement of the minima in the two elec-

tronic states within the IMDHO model, as well as the resolution criterion  $\sigma_i$ , which excludes modes due to expected poor resolution of their vibrational progression. In the case of oligothiophene examples, it is demonstrated that modes with large  $|\Delta_i|$  but a  $\sigma_i$  below the chosen half-width at half maximum can be treated harmonically without significant deviations in the resulting VCI-in-IMDHO spectra. Anharmonicity is evaluated by comparing the displacement between optimized minima of the two electronic states projected on the respective mode ( $|\Delta_i^a|$ ) to  $|\Delta_i|$ , but no clear correlation is found. Anharmonic effects have a greater impact on FC factors than vibrational energies, and the influence of anharmonicity varies among the dyes studied. It is important to note that The VCI-in-IMDHO model reduces the computational cost compared to the reduced-space VCI method and allows for the calculation of emission line shapes of organic dyes with anharmonicity in both involved electronic states. However, estimating anharmonic effects and mode-mode couplings remains challenging. Furthermore, improved agreement is observed for 1,4-diphenylbutadiene, while the performance for HS84 and anthracene diimide varies due to experimental limitations and potential methodological constraints.

To further evaluate the effectiveness of this hybrid method, calculations of absorption and emission spectra for different conformers of a novel fluorescence probe of the LCOs family, bTVBT4, was performed. The analysis revealed that, in gas phase, three vibrational stretching and angle-bending modes in the range of 1300-1600  $\text{cm}^{-1}$  significantly contribute to the absorption spectrum profile. The 0-0 transition dominates the absorption spectrum, resulting in a pronounced inhomogeneous broadening effect. Thus, the VCI-in-IMDHO model offers a precise depiction of the ligand's nuclear motions, which is essential for comprehending the probe's dynamics in intricate scenarios like water at room temperature, as demonstrated in reference [177].

For the organic dyes, the vibrational profiles in continuum solvation models were also calculated, and the effect was found to be significant, particularly for HS84. This suggests that both anharmonicity and solvation effects need to be considered for accurate calculations of vibrational profiles. Furthermore, for the calculation of vibronic profiles in heterogeneous environments, such as HS84 at an amyloid fibril, more multi-level aspects, e.g., on the electronic-structure calculations and dynamics of the environment are required.

Therefore, chapter 6 delves deeper into the effect of the environment on the emission spectra of the widely used fluorescent dyes namely oxazines. It is found that the gas phase emission spectra of the different conformers of each oxazine are similar, indicating that their electronic structure remains conserved despite structural differences. Thus, the study focuses on the most stable conformer of each oxazine molecule. The calculated harmonic spectra in a vacuum align well with the experimental results, supporting the accuracy of the computational methods employed to study the vibronic spectra. To accurately predict the photophysical properties of the oxazine molecules in practical applications, the interaction between the dyes and their environment is carefully explored. The effect of different solvents,



including water and betaine, is studied on Ox4+ and Ox170+. The position of the water molecule has a significant impact on the spectra, particularly at position no.1 (on top of the oxazine ring), which broadens the main peaks of the emission spectra of Ox4+ by the presence of one particular mode from the solvent. The presence of a nearby solvent with a larger dipole moment, such as betaine, leads to a larger blue shift in the emission spectra compared to water, with the band shape noticeably altered. Overall, this study provides valuable insights into the role of the environment in determining the spectral properties of oxazine dyes.

To better understand the molecular basis of protein misfolding diseases and the function of fluorescent biomarkers, it is essential to gain insights into the ligand's mode of binding with its target protein, in addition to the intrinsic properties of the dye in both gas and condensed phases. In particular, there is an urgent need to diagnose and monitor patients with amyloidosis during treatment with various modalities. TTR in TTR amyloidosis is an interesting target due to its abundance of accessible blood samples and numerous successful available treatments. However, there is currently no clinically approved TTR-based biomarker, and distinguishing misfolded TTR from native tetrameric TTR is a significant challenge. The current work on fluorescent amyloid ligands that can distinguish misfolded fibrillar TTR and native TTR is a step towards developing a clinically approved biomarker.[340] Additionally, understanding the binding modes of various small molecule ligands towards the native TTR tetramer can lead to new kinetic stabilizers as alternatives to currently approved anti-TTR amyloid drugs[341] such as diflunisal and tafamidis. However, X-ray crystallography poses a challenge for the native TTR tetramer due to its intrinsic symmetry, resulting in partial occupancy observed due to the crystal lattice. Therefore, this study uses a combination of biophysical and structural methods, including molecular dynamics simulations, to establish the most plausible ligand-TTR complex structure at the atomic resolution of Py1SA-TTR. The potentials of mean force (PMFs) obtained with an enhanced sampling technique i.e. umbrella sampling (US) suggest clear domination of the reverse binding mode, in which the pyrene moiety of the Py1SA is directed along hydrophobic halogen binding pocket toward the solvent. This finding is in line with the MD results discussed above and a previous study where the structure-activity relationships [74] revealed the importance of the presence of the carboxylic acid as well as its position in the ligand structure in its activity. Hence, the methodology successfully mitigates the issue of partial ligand occupancy.

In summary, these two interconnected factors - vibrationally resolved spectra calculation and exploration of binding free energy - offer valuable insights into the spectroscopic properties of fluorescent dyes as well as provide information on the dissociation pathway, aiding in understanding ligand-protein interactions and facilitating the development of improved fluorescent probes. Overall, this thesis contributes to the progress of amyloid biomarker research, with implications for improved diagnostics and treatment strategies in the realm of amyloid-related diseases.

Looking ahead, further research can build upon the findings and methodologies presented in this thesis. For the computation of vibrationally resolved spectra, expanding the study to larger and more complex systems can enhance our understanding of amyloid aggregation in diverse environments. Specially, investigating the vibronic spectra of fluorescent probes in aqueous environments and within protein binding pockets holds great promise. This knowledge can aid in the development of more realistic and accurate models for simulating the behavior of fluorescent probes in complex biological environments. Furthermore, the binding pocket provides a unique environment with specific electrostatic, hydrophobic, and steric properties that can significantly influence the spectral characteristics of the probe. [177, 342, 343] By examining the vibronic spectra within protein binding pockets, researchers can gain insights into the structural changes, conformational dynamics, and intermolecular interactions occurring during ligand binding. [342, 343] This information is crucial for the rational design of ligands with improved binding affinity and selectivity. However, it should be noted that, in quantum chemical calculations, explicitly including the environment can be computationally challenging, especially for large systems. To overcome this limitation, embedding methods treat the molecule of interest, the environment, or their interaction separately, allowing for the computation of larger systems while maintaining accuracy for the essential components. One prominent embedding scheme is the quantum mechanics/molecular mechanics (QM/MM) approach, which combines quantum chemical calculations with molecular mechanics to account for the electronic and mechanical properties of the system, respectively. [343–346] Within this model, the dye is subjected to quantum mechanical calculations while its surrounding environment is described using classical approaches. In recent years QM/MM methods have been successfully applied to the simulation of electronic spectroscopies of biological systems. [29, 267, 343, 347–350]

In addition to the binding affinity, which is governed by the relative free energy difference between the two states, the kinetics of ligand binding and unbinding processes are important quantities for predicting drug/biomarker action *in vivo*. This binding kinetics is affected by the intricate pathway that connects these states. Depending on the application, both faster and slower binding kinetics have been shown to be desirable objectives for rational drug design [351–354]. We should bear in mind that PMF provides information about the thermodynamics of ligand–protein association or dissociation, but it does not directly provide information about the kinetics of such processes. In principle, if a good reaction coordinate is identified, and the PMF is in fact well-converged, including in the barrier regions, then the kinetic rate constants can also be estimated from the recovered barrier height between the metastable states using transition state theory (TST). [355, 356] However, a fundamental issue in using a PMF derived from umbrella sampling, metadynamics, or any other method for rate calculation is that of dynamical corrections to the TST rate. [356, 357] The transition rate calculated via TST provides only a crude upper bound to the true rate, and a transmission coefficient needs to be calculated, for example, by launching trajectories from the top of every barrier identified on the PMF. [356]

The other outstanding and unresolved issue concerns the impact of solvent environments such as osmolyte solutions on the binding processes. Since many biological reactions occur in living cells and not in pure water, a liquid environment is crucial to clarify the discrepancies between theoretical estimates and *in vivo* data. The cytoplasm is dense with macromolecules, small organic molecules, and ions, and structural and dynamic aspects can vary greatly from those in pure water. Molecular simulations are capable of investigating these intricate systems directly. As a first step, one can study the molecular basis of the near-ideal solution properties of a urea solution at various concentrations, as well as the solvation/transfer free energy of a peptide in various osmolyte solutions. To obtain dependable results from molecular simulations, accurate force-field parameters must be developed to express chemical activity in concentrated solutions. General force-field refinement procedures for protein [358, 359], and Kirkwood-Buff force field (KBFF) models [360–362] are also promising approaches using thermodynamic data to determine dependable force-field parameters for biomolecular simulations in realistic liquid conditions.

Overall, through computational studies, a deeper understanding of fluorescent amyloid biomarkers can be achieved, leading to advancements in their design and application for diagnostic and therapeutic purposes.



## ACRONYMS

---

<b>A<math>\beta</math></b>	amyloid $\beta$
<b>ADI</b>	anthracene diimide
<b>AD</b>	Alzheimer's disease
<b>ADGA</b>	adaptive density guided approach
<b>ATTR</b>	transthyretin amyloidosis
<b>ATTRwt</b>	wild-type transthyretin amyloidosis
<b>BBB</b>	blood-brain barrier
<b>BOA</b>	Born-Oppenheimer approximation
<b>CAM</b>	coulomb attenuating method
<b>CC</b>	coupled cluster
<b>CI</b>	configuration interaction
<b>Coul-SR</b>	Coulombic short-range
<b>CV</b>	collective variable
<b>CV+</b>	cresyl violet
<b>DFT</b>	density functional theory
<b>DOF</b>	degree of freedom
<b>DPB</b>	1,4-diphenylbutadien
<b>DR+</b>	darrow red
<b>DZ</b>	double- $\zeta$
<b>ES</b>	excited-state
<b>EXX</b>	exact-exchange
<b>FQ</b>	first quantization
<b>FF</b>	force field
<b>FC</b>	Franck-Condon
<b>GAFF</b>	General Amber Force Field
<b>GGA</b>	generalized gradient approximation
<b>GS</b>	ground-state
<b>HF</b>	Hartree-Fock
<b>HBP</b>	halogen binding pocket
<b>HK</b>	Hohenberg-Kohn
<b>HO</b>	harmonic oscillator
<b>HWHM</b>	half-width at half-maximum
<b>IEF-PCM</b>	integral equation formalism of the polarizable-continuum model
<b>IMDHO</b>	independent mode, displaced harmonic oscillator
<b>KS</b>	Kohn-Sham
<b>LCAO</b>	linear combination of atomic orbitals
<b>LCOs</b>	luminescent conjugated oligothiophenes
<b>LDA</b>	local density approximation
<b>LINCS</b>	linear constraint solver
<b>LJ-SR</b>	Lennard-Jones short-range

**LL** Levy–Lieb  
**LR** linear response  
**MC** mode combination  
**MCR** mode combination range  
**MD** molecular dynamics  
**MM** molecular mechanics  
**MO** molecular orbital  
**ONV** occupation number vector  
**Ox1+** oxazine 1  
**Ox170+** oxazine 170  
**Ox4+** oxazine 4  
**PBC** periodic boundary conditions  
**PCM** polarizable continuum model  
**PD** Parkinson’s disease  
**PDB** Protein Data Bank  
**PEF** potential energy function  
**PES** potential energy surface  
**PME** particle–mesh Ewald  
**PMF** potential of mean force  
**TZ** triple- $\zeta$   
**QM** quantum mechanics  
**QZ** quadruple- $\zeta$   
**RC** reaction coordinate  
**RMSD** root-mean-square deviation  
**SCF** self-consistent field  
**SE** Schrödinger equation  
**SIE** self-interaction error  
**SMD** solvation model based on density  
**SSA** senile systemic amyloidosis  
**SQ** second quantization  
**TTR** transthyretin  
**TTRwt** wild-type transthyretin  
**TD** time dependent  
**TD-DFT** time-dependent density functional theory  
**TI** time-independent  
**TF** Thomas–Fermi  
**US** umbrella sampling  
**VCI** vibrational configuration interaction  
**VSCF** variational self-consistent field  
**WHAM** weighted histogram analysis method  
**WFT** wave function theory  
**XC** exchange–correlation

## BIBLIOGRAPHY

---

- [1] M. Vendruscolo, J. Zurdo, C. E. MacPhee, and C. M. Dobson, *Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems*, Phil. Trans. R. Soc. Lond. **361**, 1205 (2003).
- [2] C. M. Dobson, *Protein folding and misfolding*, Nature **426**, 884 (2003).
- [3] J. N. Onuchic and P. G. Wolynes, *Theory of protein folding*, Curr. Opin. Struct. Biol. **14**, 70 (2004).
- [4] J. Adamcik and R. Mezzenga, *Amyloid polymorphism in the protein folding and aggregation energy landscape*, Angew. Chem., Int. Ed. **57**, 8370 (2018).
- [5] A. Morriss-Andrews and J.-E. Shea, *Computational studies of protein aggregation: methods and applications*, Annu. Rev. Phys. Chem. **66**, 643 (2015).
- [6] V. Bellotti and M. Stoppini, *Protein misfolding diseases*, Open Biol. **2** (2009).
- [7] P. Westermark, M. D. Benson, J. N. Buxbaum, A. S. Cohen, B. Frangione, S.-I. Ikeda, C. L. Masters, G. Merlini, M. J. Saraiva, and J. D. Sipe, *Amyloid: Toward terminology clarification report from the nomenclature committee of the international society of amyloidosis*, Amyloid **12**, 1 (2005).
- [8] F. Chiti and C. M. Dobson, *Protein misfolding, functional amyloid, and human disease*, Annu. Rev. Biochem. **75**, 333 (2006).
- [9] H. Ecroyd and J. A. Carver, *Unraveling the mysteries of protein folding and misfolding*, IUBMB life **60**, 769 (2008).
- [10] T. P. Knowles, M. Vendruscolo, and C. M. Dobson, *The amyloid state and its association with protein misfolding diseases*, Nat. Rev. Mol. Cell Biol. **15**, 384 (2014).
- [11] C. A. Ross and M. A. Poirier, *Protein aggregation and neurodegenerative disease*, Nat. Med. **10**, S10 (2004).
- [12] A. W. Fitzpatrick et al., *Atomic structure and hierarchical assembly of a cross- $\beta$  amyloid fibril*, Proc. Natl. Acad. Sci. **110**, 5468 (2013).
- [13] J. Näslund, A. Schierhorn, U. Hellman, L. Lannfelt, A. D. Roses, L. O. Tjernberg, J. Silberring, S. E. Gandy, B. Winblad, and P. Greengard, *Relative abundance of Alzheimer A $\beta$  amyloid peptide variants in Alzheimer disease and normal aging.*, Proc. Natl. Acad. Sci. USA **91**, 8378 (1994).

- [14] M. Pepys, *Pathogenesis, diagnosis and treatment of systemic amyloidosis*, Philos. Trans. R. Soc. Lond. B Biol. Sci. **356**, 203 (2001).
- [15] A. D. Wechalekar, J. D. Gillmore, and P. N. Hawkins, *Systemic amyloidosis*, The Lancet **387**, 2641 (2016).
- [16] M. Oerlemans, K. Rutten, M. Minnema, R. Raymakers, F. Asselbergs, and N. de Jonge, *Cardiac amyloidosis: the need for early diagnosis*, Neth. Heart J. **27**, 525 (2019).
- [17] D. Adams, H. Koike, M. Slama, and T. Coelho, *Hereditary transthyretin amyloidosis: a model of medical progress for a fatal disease*, Nat. Rev. Neurol. **15**, 387 (2019).
- [18] A. Gonzalez-Duarte and A. Ulloa-Aguirre, *A brief journey through protein misfolding in transthyretin amyloidosis (ATTR amyloidosis)*, Int. J. Mol. Sci. **22**, 13158 (2021).
- [19] A. Quintas, M. J. M. Saraiva, and R. M. Brito, *The tetrameric protein transthyretin dissociates to a non-native monomer in solution: a novel model for amyloidogenesis*, J. Biol. Chem. **274**, 32943 (1999).
- [20] M. R. Sawaya et al., *Atomic structures of amyloid cross- $\beta$  spines reveal varied steric zippers*, Nature **447**, 453 (2007).
- [21] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko, *A structural model for Alzheimer's  $\beta$ -amyloid fibrils based on experimental constraints from solid state NMR*, Proc. Natl. Acad. Sci. **99**, 16742 (2002).
- [22] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek, *3D structure of Alzheimer's amyloid- $\beta$  (1-42) fibrils*, Proc. Natl. Acad. Sci. **102**, 17342 (2005).
- [23] J. K. Lim, Q.-X. Li, Z. He, A. J. Vingrys, V. H. Wong, N. Currier, J. Mullen, B. V. Bui, and C. T. Nguyen, *The eye as a biomarker for Alzheimer's disease*, Front. Neurosci. **10**, 536 (2016).
- [24] A. J. Atkinson Jr et al., *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework*, Clin. Pharmacol. Ther. **69**, 89 (2001).
- [25] E. Bistaffa, F. Tagliavini, P. Matteini, and F. Moda, *Contributions of molecular and optical techniques to the clinical diagnosis of Alzheimer's disease*, Brain Sci. **10**, 815 (2020).
- [26] S. Sargazi, I. Fatima, M. H. Kiani, V. Mohammadzadeh, R. Arshad, M. Bilal, A. Rahdar, A. M. Díez-Pascual, and R. Behzadmehr, *Fluorescent-based nanosensors for selective detection of a wide range of biological macromolecules: A comprehensive review*, Int. J. Biol. Macromol. **206**, 115 (2022).



- [27] A. De Simone, M. Naldi, D. Tedesco, M. Bartolini, L. Davani, and V. Andrisano, *Advanced analytical methodologies in Alzheimer’s disease drug discovery*, J. Pharm. Biomed. Anal. **178**, 112899 (2020).
- [28] W.-T. Dou, H.-H. Han, A. C. Sedgwick, G.-B. Zhu, Y. Zang, X.-R. Yang, J. Yoon, T. D. James, J. Li, and X.-P. He, *Fluorescent probes for the detection of disease-associated biomarkers*, Sci. Bull. (2022).
- [29] T. J. Zuehlsdorff and C. M. Isborn, *Modeling absorption spectra of molecules in solution*, Int. J. Quantum Chem. **119**, e25719 (2019).
- [30] F. Santoro and V. Barone, *Computational approach to the study of the lineshape of absorption and electronic circular dichroism spectra*, Int. J. Quantum Chem. **110**, 476 (2010).
- [31] E. Kahl, J. Berengut, M. Laatiaoui, E. Eliav, and A. Borschevsky, *High-precision ab initio calculations of the spectrum of  $Lr+$* , Phys. Rev. A **100**, 062505 (2019).
- [32] D. Madsen, O. Christiansen, P. Norman, and C. König, *Vibrationally resolved emission spectra of luminescent conjugated oligothiophenes from anharmonic calculations*, Phys. Chem. Chem. Phys. **21**, 17410 (2019).
- [33] C. König, *Tailored multilevel approaches in vibrational structure theory: A route to quantum mechanical vibrational spectra for complex systems*, Int. J. Quantum Chem. **121**, e26375 (2021).
- [34] A. Leach, *Molecular modelling: principles and applications*, Pearson, 2 edition, 2001.
- [35] C. Puzzarini, J. Bloino, N. Tasinato, and V. Barone, *Accuracy and interpretability: The devil and the holy grail. New routes across old boundaries in computational spectroscopy*, Chem. Rev. **119**, 8131 (2019).
- [36] J. Bloino, A. Baiardi, and M. Biczysko, *Aiming at an accurate prediction of vibrational and electronic spectra for medium-to-large molecules: an overview*, Int. J. Quantum Chem. **116**, 1543 (2016).
- [37] T. L. Jansen, *Computational spectroscopy of complex systems*, J. Chem. Phys. **155**, 170901 (2021).
- [38] M. Dierksen and S. Grimme, *Density functional calculations of the vibronic structure of electronic absorption spectra*, J. Chem. Phys. **120**, 3544 (2004).
- [39] J. M. Luis, D. M. Bishop, and B. Kirtman, *A different approach for calculating Franck–Condon factors including anharmonicity*, J. Chem. Phys. **120**, 813 (2004).
- [40] J. M. Luis, B. Kirtman, and O. Christiansen, *A variational approach for calculating Franck–Condon factors including mode-mode anharmonic coupling*, J. Chem. Phys. **125**, 154114 (2006).

- [41] J. Cerezo, D. Aranda, F. J. Avila Ferrer, G. Prampolini, and F. Santoro, *Adiabatic-molecular dynamics generalized vertical hessian approach: a mixed quantum classical method to compute electronic spectra of flexible molecules in the condensed phase*, J. Chem. Theory Comput. **16**, 1215 (2019).
- [42] H. Naiki, K. Higuchi, M. Hosokawa, and T. Takeda, *Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavine T*, Anal. Biochem. **177**, 244 (1989).
- [43] H. Puchtler and F. Sweat, *Congo red as a stain for fluorescence microscopy of amyloid*, . Histochem. Cytochem. **13**, 693 (1965).
- [44] T. Klingstedt, H. Shirani, K. A. Åslund, N. J. Cairns, C. J. Sigurdson, M. Goedert, and K. P. R. Nilsson, *The structural basis for optimal performance of oligothiophene-based fluorescent amyloid ligands: Conformational flexibility is essential for spectral assignment of a diversity of protein aggregates*, Chem. Eur. J. **19**, 10179 (2013).
- [45] C. J. Sigurdson, K. P. R. Nilsson, S. Hornemann, G. Manco, M. Polymenidou, P. Schwarz, M. Leclerc, P. Hammarström, K. Wüthrich, and A. Aguzzi, *Prion strain discrimination using luminescent conjugated polymers*, Nat. Methods **4**, 1023 (2007).
- [46] T. Klingstedt, A. Åslund, R. A. Simon, L. B. Johansson, J. J. Mason, S. Nyström, P. Hammarström, and K. P. R. Nilsson, *Synthesis of a library of oligothiophenes and their utilization as fluorescent ligands for spectral assignment of protein aggregates*, Org. Biomol. Chem **9**, 8356 (2011).
- [47] T. Klingstedt, H. Shirani, J. Mahler, B. M. Wegenast-Braun, S. Nyström, M. Goedert, M. Jucker, and K. P. R. Nilsson, *Distinct spacing between anionic groups: An essential chemical determinant for achieving thiophene-based ligands to distinguish  $\beta$ -amyloid or tau polymorphic aggregates*, Chem. Eur. J. **21**, 9072 (2015).
- [48] R. A. Simon, H. Shirani, K. A. Åslund, M. Bäck, V. Haroutunian, S. Gandy, and K. P. R. Nilsson, *Pentameric thiophene-based ligands that spectrally discriminate amyloid- $\beta$  and tau aggregates display distinct solvatochromism and viscosity-induced spectral shifts*, Chem. Eur. J. **20**, 12537 (2014).
- [49] H. Shirani, M. Linares, C. J. Sigurdson, M. Lindgren, P. Norman, and K. P. R. Nilsson, *A palette of fluorescent thiophene-based ligands for the identification of protein aggregates*, Chem. - A Eur. J. **21**, 15133 (2015).
- [50] A. Åslund et al., *Novel pentameric thiophene derivatives for in vitro and in vivo optical imaging of a plethora of protein aggregates in cerebral amyloidoses*, ACS Chem. Biol. **4**, 673 (2009).
- [51] H. Shirani, H. Appelqvist, M. Bäck, T. Klingstedt, N. J. Cairns, and K. P. R. Nilsson, *Synthesis of thiophene-based optical ligands that selectively detect tau pathology in Alzheimer's disease*, Chem. Eur. J. **23**, 17127 (2017).

- [52] S. E. Wallace-Williams, S. Moeller, R. A. Goldbeck, K. M. Hanson, J. W. Lewis, W. A. Lee, and D. S. Kliger, *Excited-state s-cis rotamers produced by extreme red edge excitation of all-trans-1, 4-diphenyl-1, 3-butadiene*, J. Phys. Chem. **97**, 9587 (1993).
- [53] Y.-P. Sun, C. E. Bunker, P. L. Wickremesinghe, H. W. Rollins, and G. E. Lawson, *Quantitative study of dual fluorescence from s-cis and s-trans conformers of trans, trans-1, 4-diphenyl-1, 3-butadiene*, J. Phys. Chem. **99**, 3423 (1995).
- [54] D. Morris and T. L. Gustafson, *Solvent solute interactions probed by picosecond transient Raman spectroscopy: S 1 1, 4-diphenyl-1, 3-butadiene in the linear alkanes*, Appl. Phys. B **59**, 389 (1994).
- [55] R. Burnett, I. Brown, and J. Findlay, *Cresyl fast violet staining method for Campylobacter like organisms.*, J. Clin. Pathol. **40**, 353 (1987).
- [56] L. Zhu, W. Wu, M.-Q. Zhu, J. J. Han, J. K. Hurst, and A. D. Li, *Reversibly photoswitchable dual-color fluorescent nanoparticles as new tools for live-cell imaging*, J. Am. Chem. Soc. **129**, 3524 (2007).
- [57] H. Ju, Y. Ye, and Y. Zhu, *Interaction between Nile blue and immobilized single-or double-stranded DNA and its application in electrochemical recognition*, Electrochim. acta **50**, 1361 (2005).
- [58] H. Cao, D. I. Diaz, N. DiCesare, J. R. Lakowicz, and M. D. Heagy, *Monoboronic acid sensor that displays anomalous fluorescence sensitivity to glucose*, Org. Lett. **4**, 1503 (2002).
- [59] Y. Che, X. Yang, S. Loser, and L. Zang, *Expedient vapor probing of organic amines using fluorescent nanofibers fabricated from an n-type organic semiconductor*, Nano Lett. **8**, 2219 (2008).
- [60] X. Zhan, A. Facchetti, S. Barlow, T. J. Marks, M. A. Ratner, M. R. Wasielewski, and S. R. Marder, *Rylene and related diimides for organic electronics*, Adv. Mater. **23**, 268 (2011).
- [61] C. Li and H. Wonneberger, *Perylene imides for organic photovoltaics: Yesterday, today, and tomorrow*, Adv. Mater. **24**, 613 (2012).
- [62] H. Langhals, *Cyclic carboxylic imide structures as structure elements of high stability. Novel developments in perylene dye chemistry*, Heterocycles **1**, 477 (1995).
- [63] X.-D. Liu, C. Fan, R. Sun, Y.-J. Xu, and J.-F. Ge, *Nile-red and Nile-blue-based near-infrared fluorescent probes for in-cellulo imaging of hydrogen sulfide*, Anal. Bioanal. Chem. **406**, 7059 (2014).
- [64] S. J. Isak and E. M. Eyring, *Fluorescence quantum yield of cresyl violet in methanol and water as a function of concentration*, J. Phys. Chem. **96**, 1738 (1992).

- [65] Q. Wan, Y. Song, Z. Li, X. Gao, and H. Ma, *In vivo monitoring of hydrogen sulfide using a cresyl violet-based ratiometric fluorescence probe*, Chem-Comm. **49**, 502 (2013).
- [66] R. Gvishi and R. Reisfeld, *An investigation of the equilibrium between various forms of oxazine-170 by means of absorption and fluorescence spectroscopy*, Chem. Phys. Lett. **156**, 181 (1989).
- [67] J. Jose and K. Burgess, *Benzophenoxazine-based fluorescent dyes for labeling biomolecules*, Tetrahedron **62**, 11021 (2006).
- [68] M. Hintersteiner et al., *In vivo detection of amyloid- $\beta$  deposits by near-infrared imaging using an oxazine-derivative probe*, Nat. Biotechnol. **23**, 577 (2005).
- [69] M. A. Liz, T. Coelho, V. Bellotti, M. I. Fernandez-Arias, P. Mallaina, and L. Obici, *A narrative review of the role of transthyretin in health and disease*, Neurol. Ther. **9**, 395 (2020).
- [70] J. Magalhães, J. Eira, and M. A. Liz, *The role of transthyretin in cell biology: Impact on human pathophysiology*, Cell. Mol. Life Sci. **78**, 6105 (2021).
- [71] D. Bellovino, T. Morimoto, F. Tosetti, and S. Gaetani, *Retinol binding protein and transthyretin are secreted as a complex formed in the endoplasmic reticulum in HepG2 human hepatocarcinoma cells*, Exp. Cell Res. **222**, 77 (1996).
- [72] G. J. Miroy, Z. Lai, H. A. Lashuel, S. A. Peterson, C. Strang, and J. W. Kelly, *Inhibiting transthyretin amyloid fibril formation via protein stabilization*, Proc. Natl. Acad. Sci. U.S.A. **93**, 15051 (1996).
- [73] A. Wojtczak, V. Cody, J. R. Luft, and W. Pangborn, *Structures of human transthyretin complexed with thyroxine at 2.0 Å resolution and 3', 5'-dinitro-N-acetyl-L-thyronine at 2.2 Å resolution*, Acta Crystallogr. D Biol. Crystallogr. **52**, 758 (1996).
- [74] V. B. Oza, C. Smith, P. Raman, E. K. Koepf, H. A. Lashuel, H. M. Petrassi, K. P. Chiang, E. T. Powers, J. Sacchettini, and J. W. Kelly, *Synthesis, structure, and activity of diclofenac analogues as transthyretin amyloid fibril formation inhibitors*, J. Med. Chem. **45**, 321 (2002).
- [75] S. L. Adamski-Werner, S. K. Palaninathan, J. C. Sacchettini, and J. W. Kelly, *Diflunisal analogues stabilize the native state of transthyretin. Potent inhibition of amyloidogenesis*, J. Med. Chem. **47**, 355 (2004).
- [76] H. E. Purkey, S. K. Palaninathan, K. C. Kent, C. Smith, S. H. Safe, J. C. Sacchettini, and J. W. Kelly, *Hydroxylated polychlorinated biphenyls selectively bind transthyretin in blood and inhibit amyloidogenesis: rationalizing rodent PCB toxicity*, Chem. Biol. **11**, 1719 (2004).

- [77] C. Blake, M. Geisow, S. Oatley, B. Rerat, and C. Rerat, *Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å*, J. Mol. Biol. **121**, 339 (1978).
- [78] M. Lindgren, K. Sörgjerd, and P. Hammarström, *Detection and characterization of aggregates, prefibrillar amyloidogenic oligomers, and protofibrils using fluorescence spectroscopy*, Biophys. J. **88**, 4200 (2005).
- [79] K. Sörgjerd, T. Klingstedt, M. Lindgren, K. Kågedal, and P. Hammarström, *Prefibrillar transthyretin oligomers and cold stored native tetrameric transthyretin are cytotoxic in cell culture*, Biochem. Biophys. Res. Commun. **377**, 1072 (2008).
- [80] J. Zhang, J. Wang, A. Sandberg, X. Wu, S. Nyström, H. LeVine III, P. Konradsson, P. Hammarström, B. Durbeej, and M. Lindgren, *Intramolecular proton and charge transfer of pyrene-based trans-stilbene salicylic acids applied to detection of aggregated proteins*, ChemPhysChem **19**, 3001 (2018).
- [81] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, volume 1, Elsevier, 2001.
- [82] M. Karplus and J. A. McCammon, *Molecular dynamics simulations of biomolecules*, Nat. Struct. Biol. **9**, 646 (2002).
- [83] S. A. Hollingsworth and R. O. Dror, *Molecular dynamics simulation for all*, Neuron **99**, 1129 (2018).
- [84] B. J. Alder and T. E. Wainwright, *Phase transition for a hard sphere system*, J. Chem. Phys. **27**, 1208 (1957).
- [85] R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, and D. E. Shaw, *Pathway and mechanism of drug binding to G-protein-coupled receptors*, Proc. Natl. Acad. Sci. **108**, 13118 (2011).
- [86] I. Buch, T. Giorgino, and G. De Fabritiis, *Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations*, Proc. Natl. Acad. Sci. **108**, 10184 (2011).
- [87] A. González, T. Perez-Acle, L. Pardo, and X. Deupi, *Molecular basis of ligand dissociation in  $\beta$ -adrenergic receptors*, PLoS one **6**, e23815 (2011).
- [88] D. Wacker et al., *Crystal structure of an LSD-bound human serotonin receptor*, Cell **168**, 377 (2017).
- [89] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich, *Identifying the protein folding nucleus using molecular dynamics*, J. Mol. Biol. **296**, 1183 (2000).
- [90] C. D. Snow, H. Nguyen, V. S. Pande, and M. Gruebele, *Absolute comparison of simulated and experimental protein-folding dynamics*, nature **420**, 102 (2002).

- [91] J. P. Ulmschneider, M. B. Ulmschneider, and A. Di Nola, *Monte Carlo vs molecular dynamics for all-atom polypeptide folding simulations*, J. Phys. Chem. B **110**, 16733 (2006).
- [92] W. A. Eaton, *Modern kinetics and mechanism of protein folding: A retrospective*, J. Phys. Chem. B **125**, 3452 (2021).
- [93] M. M. Dedmon, K. Lindorff-Larsen, J. Christodoulou, M. Vendruscolo, and C. M. Dobson, *Mapping long-range interactions in  $\alpha$ -synuclein using spin-label NMR and ensemble molecular dynamics simulations*, J. Am. Chem. Soc. **127**, 476 (2005).
- [94] E. S. Groban, A. Narayanan, and M. P. Jacobson, *Conformational changes in protein loops and helices induced by post-translational phosphorylation*, PLoS Comput. Biol. **2**, 32 (2006).
- [95] T. J. Harpole and L. Delemotte, *Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations*, Biochim. Biophys. Acta - Biomembr. **1860**, 909 (2018).
- [96] J. R. Perilla and K. Schulten, *Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations*, Nat. Commun **8**, 15959 (2017).
- [97] A. Acharya et al., *Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19*, Journal of Chemical Information and Modeling **60**, 5832 (2020).
- [98] M. Born and W. Heisenberg, *Zur Quantentheorie der Molekeln*, Ann. d. Phys. , 216 (1985).
- [99] P. Derreumaux and G. Vergoten, *A new spectroscopic molecular mechanics force field. Parameters for proteins*, The Journal of chemical physics **102**, 8586 (1995).
- [100] H. Sun, *COMPASS: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds*, J. Phys. Chem. B **102**, 7338 (1998).
- [101] C. S. Ewig et al., *Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds*, J. Comput. Chem. **22**, 1782 (2001).
- [102] K. Vanommeslaeghe et al., *Molecular mechanics*, Curr. Pharm. Des. **20**, 3281 (2014).
- [103] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water*, J. Chem. Phys. **79**, 926 (1983).
- [104] W. F. van Gunsteren and H. J. Berendsen, *Groningen molecular simulation (GROMOS) library manual*, Biomos, Groningen **24**, 13 (1987).

- [105] N. L. Allinger, *Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms*, J. Am. Chem. Soc. **99**, 8127 (1977).
- [106] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system*, Proteins: Struct. Funct. Genet. **4**, 31 (1988).
- [107] F. Momany, R. F. McGuire, A. Burgess, and H. A. Scheraga, *Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids*, J. Phys. Chem. **79**, 2361 (1975).
- [108] B. R. Brooks, R. E. Bruccoleri, and B. D. Olafson, *States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations*, J. comput. Chem **4**, 187 (1983).
- [109] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *An all atom force field for simulations of proteins and nucleic acids*, J. Comput. Chem. **7**, 230 (1986).
- [110] J. Hermans, H. J. Berendsen, W. F. Van Gunsteren, and J. P. Postma, *A consistent empirical potential for water-protein interactions*, Biopolymers: Original Research on Biomolecules **23**, 1513 (1984).
- [111] E. Platt and B. Robson, *Ab-initio refinement of an orbital-centred force field for biomolecules. test cases including peptides, a sulphonamide and modelling of DNA helices*, J. Theor. Biol. **96**, 381 (1982).
- [112] S. Lifson, A. Hagler, and P. Dauber, *Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O...H-hydrogen bonds*, J. Am. Chem. Soc. **101**, 5111 (1979).
- [113] A. Donchev, V. Ozrin, M. Subbotin, O. Tarasov, and V. Tarasov, *A quantum mechanical polarizable force field for biomolecular interactions*, Proc. Natl. Acad. Sci. **102**, 7829 (2005).
- [114] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, *ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB*, J. Chem. Theory Comput. **11**, 3696 (2015).
- [115] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *Development and testing of a general amber force field*, J. Comput. Chem. **25**, 1157 (2004).
- [116] X. He, V. H. Man, W. Yang, T.-S. Lee, and J. Wang, *A fast and high-quality charge model for the next generation general AMBER force field*, J. Chem. Phys. **153**, 114502 (2020).

- [117] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model*, J. Phys. Chem. **97**, 10269 (1993).
- [118] W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, *Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation*, J. Am. Chem. Soc. **115**, 9620 (2002).
- [119] S. D. Fried, L.-P. Wang, S. G. Boxer, P. Ren, and V. S. Pande, *Calculations of the electric fields in liquid solutions*, J. Phys. Chem. B **117**, 16236 (2013).
- [120] L. Verlet, *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*, Phys. Rev. **159**, 98 (1967).
- [121] D. Beeman, *Some multistep methods for use in molecular dynamics calculations*, J. Comput. Phys. **20**, 130 (1976).
- [122] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*, McGraw-Hill, New York, first edition, 1981.
- [123] W. F. Van Gunsteren and H. J. Berendsen, *Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry*, Angew. Chem., Int. Ed. Engl. **29**, 992 (1990).
- [124] P. H. Hünenberger, *Thermostat algorithms for molecular dynamics simulations*, Advanced computer simulation: Approaches for soft matter sciences I , 105 (2005).
- [125] M. Parrinello and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*, J. Appl. Phys. **52**, 7182 (1981).
- [126] G. J. Martyna, D. J. Tobias, and M. L. Klein, *Constant pressure molecular dynamics algorithms*, J. Chem. Phys. **101**, 4177 (1994).
- [127] G. Bussi, D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*, J. Chem. Phys. **126**, 014101 (2007).
- [128] G. Bussi, T. Zykova-Timan, and M. Parrinello, *Isothermal-isobaric molecular dynamics using stochastic velocity rescaling*, J. Chem. Phys. **130**, 074101 (2009).
- [129] H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. DiNola, and J. R. Haak, *Molecular dynamics with coupling to an external bath*, J. Chem. Phys. **81**, 3684 (1984).
- [130] S. Lifson and I. Oppenheim, *Neighbor interactions and internal rotations in polymer molecules. IV. Solvent effect on internal rotations*, J. Chem. Phys. **33**, 109 (1960).
- [131] F. Fraternali and W. Van Gunsteren, *An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution*, J. Mol. Biol. **256**, 939 (1996).



- [132] G. N. Phillips Jr and B. Montgomery Pettitt, *Structure and dynamics of the water around myoglobin*, *Protein Sci.* **4**, 149 (1995).
- [133] F. H. Stillinger and A. Rahman, *Improved simulation of liquid water by molecular dynamics*, *J. Chem. Phys.* **60**, 1545 (1974).
- [134] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, Interaction models for water in relation to protein hydration, in *Intermolecular forces: proceedings of the fourteenth Jerusalem symposium on quantum chemistry and biochemistry held in jerusalem, israel, april 13–16, 1981*, pages 331–342, Springer, 1981.
- [135] H. Berendsen, J. Grigera, and T. Straatsma, *The missing term in effective pair potentials*, *J. Phys. Chem.* **91**, 6269 (1987).
- [136] M. Levitt, M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett, *Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution*, *J. Phys. Chem. B* **101**, 5051 (1997).
- [137] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *A smooth particle mesh Ewald method*, *J. Chem. Phys.* **103**, 8577 (1995).
- [138] Z. Rinkevicius et al., *VeloxChem: A Python-driven density-functional theory program for spectroscopy simulations in high-performance computing environments*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **10**, 1457 (2020).
- [139] A. Barducci, M. Bonomi, and M. Parrinello, *Metadynamics*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 826 (2011).
- [140] A. F. Voter, F. Montalenti, and T. C. Germann, *Extending the time scale in atomistic simulation of materials*, *Annu. Rev. Mater. Res.* **32**, 321 (2002).
- [141] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- [142] Z. Cournia, C. Chipot, B. Roux, D. M. York, and W. Sherman, Free energy methods in drug discovery—introduction, in *Free energy methods in drug discovery: Current state and future directions*, pages 1–38, ACS Publications, 2021.
- [143] P. Kollman, *Free energy calculations: applications to chemical and biochemical phenomena*, *Chem. Rev.* **93**, 2395 (1993).
- [144] D. L. Beveridge and F. M. Dicapua, *Free energy via molecular simulation: applications to chemical and biomolecular systems*, *Annu. Rev. Biophys. Chem.* **18**, 431 (1989).

- [145] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren, *Basic ingredients of free energy calculations: a review*, J. Comput. Chem. **31**, 1569 (2010).
- [146] Z. Cournia, B. Allen, and W. Sherman, *Relative binding free energy calculations in drug discovery: recent advances and practical considerations*, J. Chem. Inf. Model. **57**, 2911 (2017).
- [147] J. Kästner, *Umbrella sampling*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **1**, 932 (2011).
- [148] K. Shing and K. Gubbins, *The chemical potential in dense fluids and fluid mixtures via computer simulation*, Mol. Phys. **46**, 1109 (1982).
- [149] J. Powles, W. Evans, and N. Quirke, *Non-destructive molecular-dynamics simulation of the chemical potential of a fluid*, Mol. Phys. **46**, 1347 (1982).
- [150] J. Kästner, *Umbrella integration in two or more reaction coordinates*, J. Chem. Phys. **131**, 034109 (2009).
- [151] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, *Role of molecular dynamics and related methods in drug discovery*, J. Med. Chem. **59**, 4035 (2016).
- [152] O. Valsson, P. Tiwary, and M. Parrinello, *Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint*, Annu. Rev. Phys. Chem. **67**, 159 (2016).
- [153] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, and Y. Q. Gao, *Enhanced sampling in molecular dynamics*, J. Chem. Phys. **151**, 070902 (2019).
- [154] Y. Q. Gao, *An integrate-over-temperature approach for enhanced sampling*, J. Chem. Phys. **128**, 064105 (2008).
- [155] L. Yang, C.-W. Liu, Q. Shao, J. Zhang, and Y. Q. Gao, *From thermodynamics to kinetics: Enhanced sampling of rare events*, Acc. Chem. Res. **48**, 947 (2015).
- [156] Y. Wang, C. B. Harrison, K. Schulten, and J. A. McCammon, *Implementation of accelerated molecular dynamics in NAMD*, Comput. Sci. Discov. **4**, 015002 (2011).
- [157] A.-h. Wang, Z.-c. Zhang, and G.-h. Li, *Advances in enhanced sampling molecular dynamics simulations for biomolecules*, Chinese J. Chem. Phys. **32**, 277 (2019).
- [158] G. M. Torrie and J. P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling*, J. Comput. Phys. **23**, 187 (1977).
- [159] E. Papaleo, G. Saladino, M. Lambrugh, K. Lindorff-Larsen, F. L. Gervasio, and R. Nussinov, *The role of protein loops and linkers in conformational dynamics and allostery*, Chem. Rev. **116**, 6391 (2016).

- [160] D. J. Tobias, S. F. Sneddon, and C. L. Brooks III, *Reverse turns in blocked dipeptides are intrinsically unstable in water*, J. Mol. Biol. **216**, 783 (1990).
- [161] A. M. Ferrenberg and R. H. Swendsen, *New Monte Carlo technique for studying phase transitions*, Phys. Rev. Lett. **61**, 2635 (1988).
- [162] A. M. Ferrenberg and R. H. Swendsen, *Optimized Monte Carlo data analysis*, Comput. phys. **3**, 101 (1989).
- [163] J. Kästner and W. Thiel, *Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”*, J. Chem. Phys. **123**, 144104 (2005).
- [164] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method*, J. Comput. Chem. **13**, 1011 (1992).
- [165] J. S. Hub, B. L. de Groot, and D. van der Spoel, *g\_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates*, J. Chem. Theory Comput. **6**, 3713 (2010).
- [166] J. McCarty and M. Parrinello, *A variational conformational dynamics approach to the selection of collective variables in metadynamics*, J. Chem. Phys. **147**, 204109 (2017).
- [167] D. Mendels, G. Piccini, and M. Parrinello, *Collective variables from local fluctuations*, J. Phys. Chem. Lett. **9**, 2776 (2018).
- [168] A. Ma and A. R. Dinner, *Automatic method for identifying reaction coordinates in complex systems*, J. Phys. Chem. B **109**, 6769 (2005).
- [169] J. Rogal, E. Schneider, and M. E. Tuckerman, *Neural-network-based path collective variables for enhanced sampling of phase transformations*, Phys. Rev. Lett. **123**, 245701 (2019).
- [170] L. Bonati, V. Rizzi, and M. Parrinello, *Data-Driven Collective Variables for Enhanced Sampling*, J. Phys. Chem. Lett. **11**, 2998 (2020).
- [171] J. Comer, J. C. Gumbart, J. Hénin, T. Lelièvre, A. Pohorille, and C. Chipot, *The adaptive biasing force method: Everything you always wanted to know but were afraid to ask*, J. Phys. Chem. B **119**, 1129 (2015).
- [172] T. Straatsma and J. McCammon, *Multiconfiguration thermodynamic integration*, J. Chem. Phys. **95**, 1175 (1991).
- [173] N. Bansal, Z. Zheng, D. S. Cerutti, and K. M. Merz, *On the fly estimation of host–guest binding free energies using the movable type method: participation in the SAMPL5 blind challenge*, J. Comput. Aided Mol. Des. **31**, 47 (2017).
- [174] H. Sun, S. Tian, S. Zhou, Y. Li, D. Li, L. Xu, M. Shen, P. Pan, and T. Hou, *Revealing the favorable dissociation pathway of type II kinase inhibitors via*

- enhanced sampling simulations and two-end-state calculations*, Sci. Rep. **5**, 8457 (2015).
- [175] J.-K. Zhou, D.-Y. Yang, and S.-Y. Sheu, *The molecular mechanism of ligand unbinding from the human telomeric G-quadruplex by steered molecular dynamics and umbrella sampling simulations*, Phys. Chem. Chem. Phys. **17**, 12857 (2015).
- [176] W. You, Z. Tang, and C.-E. A. Chang, *Potential mean force from umbrella sampling simulations: What can we learn and what is missed?*, J. Chem. Theory Comput. **15**, 2433 (2019).
- [177] Y. Todarwal et al., *Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands*, J. Phys. Chem. B **125**, 11628 (2021).
- [178] E. Schrödinger, *An undulatory theory of the mechanics of atoms and molecules*, Phys. Rev. **28**, 1049 (1926).
- [179] T. Engel, *Physical chemistry*, Pearson Education India, 2006.
- [180] L. Serrano-Andrés and M. Merchán, *Quantum chemistry of the excited state: 2005 overview*, J. Mol. Struct-Theochem. **729**, 99 (2005).
- [181] J. C. Slater, *Note on Hartree’s method*, Phys. Rev. **35**, 210 (1930).
- [182] V. Fock, *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*, Z. Phys. **61**, 126 (1930).
- [183] D. R. Hartree, *The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods*, in *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 89–110, Cambridge university press, 1928.
- [184] P. Jensen and P. R. Bunker, *Computational molecular spectroscopy*, Wiley, 2000.
- [185] W. Ritz, *Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik.*, (1909).
- [186] J. A. Pople, R. Seeger, and R. Krishnan, *Variational configuration interaction methods and comparison with perturbation theory*, Int. J. Quantum Chem. **12**, 149 (1977).
- [187] J. Pople, R. Krishnan, H. Schlegel, and J. Binkley, *Electron correlation theories and their application to the study of simple reaction potential surfaces*, Int. J. Quantum Chem. **14**, 545 (1978).
- [188] C. Møller and M. S. Plesset, *Note on an approximation treatment for many-electron systems*, Phys. Rev. **46**, 618 (1934).
- [189] R. A. Friesner, *Ab initio quantum chemistry: Methodology and applications*, Proc. Natl. Acad. Sci. U.S.A. **102**, 6648 (2005).

- [190] J. C. Slater, *Atomic shielding constants*, Phys. Rev. **36**, 57 (1930).
- [191] S. F. Boys, *Electronic wave functions-I. A general method of calculation for the stationary states of any molecular system*, Proc. R. Soc. Lond. A Math. Phys. Sci. **200**, 542 (1950).
- [192] W. J. Hehre, R. Ditchfield, and J. A. Pople, *Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules*, J. Chem. Phys. **56**, 2257 (1972).
- [193] F. Weigend and R. Ahlrichs, *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
- [194] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, *6-31G\* basis set for third-row atoms*, J. Comput. Chem. **22**, 976 (2001).
- [195] F. Weigend, *Accurate Coulomb-fitting basis sets for H to Rn*, Phys. Chem. Chem. Phys. **8**, 1057 (2006).
- [196] J. Zheng, X. Xu, and D. G. Truhlar, *Minimally augmented Karlsruhe basis sets*, Theor. Chem. Acc. **128**, 295 (2011).
- [197] J. C. Zapata and L. K. McKemmish, *Computation of dipole moments: a recommendation on the choice of the basis set and the level of theory*, J. Phys. Chem. A **124**, 7538 (2020).
- [198] P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, Phys. Rev. **136**, B864 (1964).
- [199] M. Levy, *Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the  $v$ -representability problem*, Proc. Natl. Acad. Sci. U.S.A. **76**, 6062 (1979).
- [200] M. Levy, *Electron densities in search of Hamiltonians*, Phys. Rev. A **26**, 1200 (1982).
- [201] E. Fermi, *Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente*, Z. Phys **48**, 73 (1928).
- [202] L. H. Thomas, *The calculation of atomic fields*, in *Mathematical proceedings of the Cambridge philosophical society*, volume 23, pages 542–548, Cambridge University Press, 1927.
- [203] S. Nordholm, *Analysis of covalent bonding by nonergodic Thomas–Fermi theory*, J. Chem. Phys. **86**, 363 (1987).
- [204] W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*, Phys. Rev. **140**, A1133 (1965).

- [205] M. E. Casida and M. Huix-Rotllant, *Progress in time-dependent density-functional theory*, Annu. Rev. Phys. Chem. **63**, 287 (2012).
- [206] P. Norman, K. Ruud, and T. Saue, *Principles and practices of molecular properties: Theory, modeling, and simulations*, John Wiley & Sons, 2018.
- [207] S. Ghosh, P. Verma, C. J. Cramer, L. Gagliardi, and D. G. Truhlar, *Combining wave function methods with density functional theory for excited states*, Chem. Rev. **118**, 7249 (2018).
- [208] K. Ruedenberg, L. Cheung, and S. Elbert, *MCSCF optimization through combined use of natural orbitals and the Brillouin–Levy–Berthier theorem*, Int. J. Quantum Chem. **16**, 1069 (1979).
- [209] E. Runge and E. K. Gross, *Density-functional theory for time-dependent systems*, Phys. Rev. Lett. **52**, 997 (1984).
- [210] A. Dreuw and M. Head-Gordon, *Single-reference ab initio methods for the calculation of excited states of large molecules*, Chem. Rev. **105**, 4009 (2005).
- [211] R. Van Leeuwen, *Mapping from densities to potentials in time-dependent density-functional theory*, Phys. Rev. Lett. **82**, 3863 (1999).
- [212] E. Gross and W. Kohn, *Local density-functional theory of frequency-dependent linear response*, Phys. Rev. Lett. **55**, 2850 (1985).
- [213] E. K. Gross and W. Kohn, *Density functional theory of many-fermion systems*, Adv. Quantum Chem. **21**, 255 (1990).
- [214] J. Olsen and P. Jørgensen, *Linear and nonlinear response functions for an exact state and for an MCSCF state*, J. Chem. Phys. **82**, 3235 (1985).
- [215] M. E. Casida, *Time-dependent density functional response theory for molecules*, in *Recent Advances In Density Functional Methods: (Part I)*, pages 155–192, World Scientific, 1995.
- [216] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, *An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules*, J. Chem. Phys. **109**, 8218 (1998).
- [217] M. E. Casida, *Time-dependent density-functional theory for molecules and molecular solids*, J. Mol. Struct-Theochem. **914**, 3 (2009).
- [218] R. Bauernschmitt and R. Ahlrichs, *Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory*, Chem. Phys. Lett. **256**, 454 (1996).
- [219] S. Sharma, J. Dewhurst, S. Shallcross, G. Madjarova, and E. Gross, *Excitons in organics using time-dependent density functional theory: PPV, pentacene, and picene*, J. Chem. Theory Comput. **11**, 1710 (2015).

- [220] A. Dreuw and M. Head-Gordon, *Failure of time-dependent density functional theory for long-range charge-transfer excited states: the zincbacteriochlorin- bacteriochlorin and bacteriochlorophyll- spheroidene complexes*, J. Am. Chem. Soc. **126**, 4007 (2004).
- [221] B. G. Levine, C. Ko, J. Quenneville, and T. J. Martínez, *Conical intersections and double excitations in time-dependent density functional theory*, Mol. Phys. **104**, 1039 (2006).
- [222] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, *Spurious fractional charge on dissociated atoms: Pervasive and resilient self-interaction error of common density functionals*, J. Chem. Phys. **125**, 194112 (2006).
- [223] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Challenges for density functional theory*, Chem. Rev. **112**, 289 (2012).
- [224] D. K. Mok, R. Neumann, and N. C. Handy, *Dynamical and nondynamical correlation*, J. Phys. Chem. **100**, 6225 (1996).
- [225] A. Dreuw, G. R. Fleming, and M. Head-Gordon, *Chlorophyll fluorescence quenching by xanthophylls*, Phys. Chem. Chem. Phys. **5**, 3247 (2003).
- [226] D. J. Tozer and N. C. Handy, *Improving virtual Kohn–Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities*, J. Chem. Phys. **109**, 10180 (1998).
- [227] D. J. Tozer and N. C. Handy, *On the determination of excitation energies using density functional theory*, Phys. Chem. Chem. Phys. **2**, 2117 (2000).
- [228] N. A. Besley, M. J. Peach, and D. J. Tozer, *Time-dependent density functional theory calculations of near-edge X-ray absorption fine structure with short-range corrected functionals*, Phys. Chem. Chem. Phys. **11**, 10350 (2009).
- [229] T. Yanai, D. P. Tew, and N. C. Handy, *A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)*, Chem. Phys. Lett. **393**, 51 (2004).
- [230] Z.-L. Cai, M. J. Crossley, J. R. Reimers, R. Kobayashi, and R. D. Amos, *Density functional theory for charge transfer: The nature of the N-bands of porphyrins and chlorophylls revealed through CAM-B3LYP, CASPT2, and SAC-CI calculations*, J. Phys. Chem. B **110**, 15624 (2006).
- [231] H. Paul, *Introduction to quantum optics: from light quanta to quantum teleportation*, Cambridge University Press, 2004.
- [232] G. Onida, L. Reining, and A. Rubio, *Electronic excitations: density-functional versus many-body Green’s-function approaches*, Rev. Mod. Phys. **74**, 601 (2002).

- [233] D. Jacquemin, E. A. Perpete, I. Ciofini, and C. Adamo, *Accurate simulation of optical properties in dyes*, *Acc. Chem. Res.* **42**, 326 (2009).
- [234] J. Kim, J. Y. Lee, S. Lee, B. J. Mhin, and K. S. Kim, *Harmonic vibrational frequencies of the water monomer and dimer: Comparison of various levels of ab initio theory*, *J. Chem. Phys.* **102**, 310 (1995).
- [235] P. R. Bunker and P. Jensen, *Molecular symmetry and spectroscopy*, volume 46853, NRC research press, 2006.
- [236] J. K. Watson, *Simplification of the molecular vibration-rotation Hamiltonian*, *Mol. Phys.* **15**, 479 (1968).
- [237] A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. Ipatova, *Theory of lattice dynamics in the harmonic approximation*, volume 3, Academic press New York, 1963.
- [238] F. Pascale, C. M. Zicovich-Wilson, F. López Gejo, B. Civalleri, R. Orlando, and R. Dovesi, *The calculation of the vibrational frequencies of crystalline compounds and its implementation in the CRYSTAL code*, *J. Comput. Chem.* **25**, 888 (2004).
- [239] S. Carter, S. J. Culik, and J. M. Bowman, *Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of CO adsorbed on Cu (100)*, *J. Chem. Phys.* **107**, 10458 (1997).
- [240] G. Rauhut, *Efficient calculation of potential energy surfaces for the generation of vibrational wave functions*, *J. Chem. Phys.* **121**, 9313 (2004).
- [241] J. Kongsted and O. Christiansen, *Automatic generation of force fields and property surfaces for use in variational vibrational calculations of anharmonic vibrational energies and zero-point vibrational averaged properties*, *J. Chem. Phys.* **125**, 124108 (2006).
- [242] O. Christiansen, I. H. Godtlielsen, E. M. Gras, W. Györffy, M. B. Hansen, M. B. Hansen, E. L. Klinting, J. Kongsted, C. König, S. A. Losilla, D. Madsen, N. K. Madsen, P. Seidler, K. Sneskov, M. Sparta, B. Thomsen, D. Toffoli, and A. Zocante, *Molecular Interactions Dynamics And Simulation C++ (MidasCPP) package*, 2019.
- [243] M. Sparta, D. Toffoli, and O. Christiansen, *An adaptive density-guided approach for the generation of potential energy surfaces of polyatomic molecules*, *Theor. Chem. Acc.* **123**, 413 (2009).
- [244] M. Sparta, I.-M. Høyvik, D. Toffoli, and O. Christiansen, *Potential energy surfaces for vibrational structure calculations from a multiresolution adaptive density-guided approach: Implementation and test calculations*, *J. Phys. Chem. A* **113**, 8712 (2009).



- [245] O. Christiansen and J. M. Luis, *Beyond vibrational self-consistent-field methods: Benchmark calculations for the fundamental vibrations of ethylene*, Int. J. Quantum Chem. **104**, 667 (2005).
- [246] O. Christiansen, *Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties*, Phys. Chem. Chem. Phys. **9**, 2942 (2007).
- [247] O. Christiansen, *Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations*, Phys. Chem. Chem. Phys. **14**, 6672 (2012).
- [248] J. M. Bowman, *The self-consistent-field approach to polyatomic vibrations*, Acc. Chem. Res. **19**, 202 (1986).
- [249] D. Toffoli, M. Sparta, and O. Christiansen, *Accurate multimode vibrational calculations using a B-spline basis: theory, tests and application to dioxirane and diazirinone*, Mol. Phys. **109**, 673 (2011).
- [250] J. M. Bowman, K. Christoffel, and F. Tobin, *Application of SCF-SI theory to vibrational motion in polyatomic molecules*, J. Chem. Phys. **83**, 905 (1979).
- [251] O. Christiansen, *A second quantization formulation of multimode dynamics*, J. Chem. Phys. **120**, 2140 (2004).
- [252] P. Atkins, P. W. Atkins, and J. de Paula, *Atkins' physical chemistry*, Oxford university press, 11 edition, 2014.
- [253] S. Grimme, *Calculation of the electronic spectra of large molecules*, Rev. Comput. Chem. **20**, 153 (2004).
- [254] E. U. Condon, *Nuclear motions associated with electron transitions in diatomic molecules*, Phys. Rev. **32**, 858 (1928).
- [255] E. U. Condon, *The Franck-Condon principle and related topics*, Am. J. Phys. **15**, 365 (1947).
- [256] J. Franck and E. Dymond, *Elementary processes of photochemical reactions*, Trans. Faraday Soc. **21**, 536 (1926).
- [257] A. B. Myers, R. A. Mathies, D. J. Tannor, and E. J. Heller, *Excited state geometry changes from preresonance Raman intensities: isoprene and hexatriene*, J. Chem. Phys. **77**, 3857 (1982).
- [258] A. B. Myers and R. A. Mathies, *Resonance Raman intensities: a probe of excited-state structure and dynamics*, in *Biological Applications of Raman Spectroscopy*, pages 1–58, Wiley: New York, 1987.
- [259] T. Sharp and H. Rosenstock, *Franck-Condon factors for polyatomic molecules*, J. Chem. Phys. **41**, 3453 (1964).

- [260] P. M. Champion and A. C. Albrecht, *On band shapes of electronic transitions in the multimode weak coupling limit*, J. Chem. Phys. **72**, 6498 (1980).
- [261] G. Herzberg and E. Teller, *Schwingungsstruktur der Elektronenübergänge bei mehratomigen Molekülen*, Z. Phys. Chem. Abt. **21**, 410 (1933).
- [262] E. Doktorov, I. Malkin, and V. Man'Ko, *Dynamical symmetry of vibronic transitions in polyatomic molecules and the Franck-Condon principle*, J. Mol. Spectrosc. **64**, 302 (1977).
- [263] A. Hazra and M. Nooijen, *Comparison of various Franck-Condon and vibronic coupling approaches for simulating electronic spectra: The case of the lowest photoelectron band of ethylene*, Phys. Chem. Chem. Phys. **7**, 1759 (2005).
- [264] K. K. Lehmann, *Lorentzian line shape due to an inhomogeneous size distribution without relaxation*, J. Chem. Phys. **126**, 024108 (2007).
- [265] S. Miertuš, E. Scrocco, and J. Tomasi, *Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects*, Chem. Phys. **55**, 117 (1981).
- [266] B. Mennucci, *Polarizable continuum model*, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 386 (2012).
- [267] T. J. Zuehlsdorff, S. V. Shedge, S. Y. Lu, H. Hong, V. P. Aguirre, L. Shi, and C. M. Isborn, *Vibronic and environmental effects in simulations of optical spectroscopy*, Annu. Rev. Phys. Chem. **72**, 165 (2021).
- [268] D. Loco and L. Cupellini, *Modeling the absorption lineshape of embedded systems from molecular dynamics: A tutorial review*, Int. J. Quantum Chem. **119**, e25726 (2019).
- [269] A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, *Modeling solvent effects on electronic excited states*, J. Phys. Chem. Lett **2**, 2184 (2011).
- [270] D.-S. Ahn, S.-W. Park, I.-S. Jeon, M.-K. Lee, N.-H. Kim, Y.-H. Han, and S. Lee, *Effects of microsolvation on the structures and reactions of neutral and zwitterion alanine: computational study*, J. Phys. Chem. B **107**, 14109 (2003).
- [271] S. Raugai, G. Cardini, and V. Schettino, *Microsolvation effect on chemical reactivity: The case of the  $Cl^- + CH_3Br$   $S_N2$  reaction*, J. Chem. Phys. **114**, 4089 (2001).
- [272] C. Loison, R. Antoine, M. Broyer, P. Dugourd, J. Guthmuller, and D. Simon, *Microsolvation effects on the optical properties of crystal violet*, Chem. Eur. J. **14**, 7351 (2008).

- [273] S. Venkatesan and S.-L. Lee, *Computational investigation on microsolvation of the osmolyte glycine betaine [GB (H<sub>2</sub>O)<sub>1-7</sub>]*, J. Mol. Model. **18**, 5017 (2012).
- [274] F. Neese, T. Petrenko, D. Ganyushin, and G. Olbrich, *Advanced aspects of ab initio theoretical optical spectroscopy of transition metal complexes: Multiplets, spin-orbit coupling and resonance Raman intensities*, Coord. Chem. Rev. **251**, 288 (2007).
- [275] T. Petrenko, O. Krylova, F. Neese, and M. Sokolowski, *Optical absorption and emission properties of rubrene: insight from a combined experimental and theoretical study*, New J. Phys. **11**, 015001 (2009).
- [276] D. Barton, C. König, and J. Neugebauer, *Vibronic-structure tracking: A shortcut for vibrationally resolved UV/Vis-spectra calculations*, J. Chem. Phys. **141**, 164115 (2014).
- [277] F. Duschinsky, *On the interpretation of electronic spectra of polyatomic molecules*, Acta Physicochim. URSS **7**, 551 (1937).
- [278] M. Biczysko, J. Bloino, F. Santoro, and V. Barone, *Computational strategies for spectroscopy: from small molecules to nano systems*, chapter Time-independent approaches to simulate electronic spectra lineshapes: From small molecules to macrosystems, page 361–443, John Wiley & Sons, 2012.
- [279] D. J. Tannor and E. J. Heller, *Polyatomic Raman scattering for general harmonic potentials*, J. Chem. Phys. **77**, 202 (1982).
- [280] T. Petrenko and F. Neese, *Efficient and automatic calculation of optical band shapes and resonance Raman spectra for larger molecules within the independent mode displaced harmonic oscillator model*, J. Chem. Phys. **137**, 234107 (2012).
- [281] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Funnels, pathways, and the energy landscape of protein folding: a synthesis*, Proteins **21**, 167 (1995).
- [282] C. M. Dobson and M. Karplus, *The fundamentals of protein folding: bringing together theory and experiment*, Curr. Opin. Struct. Biol. **9**, 92 (1999).
- [283] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson, and M. Karplus, *Understanding protein folding via free-energy surfaces from theory and experiment*, Trends Biochem. Sci. **25**, 331 (2000).
- [284] T. R. Jahn and S. E. Radford, *The Yin and Yang of protein folding*, FEBS J. **272**, 5962 (2005).
- [285] J. Tyedmers, A. Mogk, and B. Bukau, *Cellular strategies for controlling protein aggregation*, Nat. Rev. Mol. Cell Biol. **11**, 777 (2010).

- [286] J. S. Valastyan and S. Lindquist, *Mechanisms of protein-folding diseases at a glance*, Dis. Models Mech. **7**, 9 (2014).
- [287] M. Pitschke, R. Prior, M. Haupt, and D. Riesner, *Detection of single amyloid  $\beta$ -protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy*, Nat. Med. **4**, 832 (1998).
- [288] R. Tycko, *Progress towards a molecular-level structural understanding of amyloid fibrils*, Curr. Opin. Struct. Biol. **14**, 96 (2004).
- [289] A. Dusa, J. Kaylor, S. Edridge, N. Bodner, D.-P. Hong, and A. L. Fink, *Characterization of oligomers during  $\alpha$ -synuclein aggregation using intrinsic tryptophan fluorescence*, Biochem. **45**, 2752 (2006).
- [290] C. Frieden, *Protein aggregation processes: In search of the mechanism*, Protein Sci. **16**, 2334 (2007).
- [291] M. Lindgren and P. Hammarström, *Amyloid oligomers: spectroscopic characterization of amyloidogenic protein states*, FEBS J. **277**, 1380 (2010).
- [292] M. Wehrle, M. Šulc, and J. Vaníček, *On-the-fly ab initio semiclassical dynamics: Identifying degrees of freedom essential for emission spectra of oligothiophenes*, J. Chem. Phys. **140**, 244114 (2014).
- [293] M. J. Frisch et al., Gaussian16 Revision B.01, 2016, Gaussian Inc. Wallingford CT.
- [294] T. H. Dunning Jr, *Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen*, J. Chem. Phys. **90**, 1007 (1989).
- [295] D. Madsen, *Vibrational and vibronic spectra computed with mode-coupling approximations*, PhD thesis, Aarhus University, 2019.
- [296] T. Moitra, D. Madsen, O. Christiansen, and S. Coriani, *Vibrationally resolved coupled-cluster x-ray absorption spectra from vibrational configuration interaction anharmonic calculations*, J. Chem. Phys. **153**, 234111 (2020).
- [297] J. M. Bowman, X. Huang, L. B. Harding, and S. Carter, *The determination of molecular properties from MULTIMODE with an application to the calculation of Franck-Condon factors for photoionization of CF<sub>3</sub>*, Mol. Phys. **104**, 33 (2007).
- [298] V. Rodriguez-Garcia, K. Yagi, K. Hirao, S. Iwata, and S. Hirata, *Franck-Condon factors based on anharmonic vibrational wave functions of polyatomic molecules*, J. Chem. Phys. **125**, 014109 (2006).
- [299] G. Rauhut, *Anharmonic Franck-Condon factors for the  $X^2B_1 \leftarrow X^1A_1$  photoionization of ketene*, J. Phys. Chem. A **119**, 10264 (2015).

- [300] R. Piessens, E. de Doncker-Kapenga, C. W. Überhuber, and D. K. Kahaner, *Quadpack: a subroutine package for automatic integration*, Springer-Verlag, Berlin Heidelberg, 1983.
- [301] E. Cancès, B. Mennucci, and J. Tomasi, *A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics*, J. Chem. Phys. **107**, 3032 (1997).
- [302] J. Tomasi, B. Mennucci, and R. Cammi, *Quantum mechanical continuum solvation models*, Chem. Rev. **105**, 2999 (2005).
- [303] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions*, J. Phys. Chem. B . **113**, 6378 (2009).
- [304] Ponder, JW, Tinker–software tools for molecular design, Washington University: St. Louis, 2015.
- [305] F. J. Avila Ferrer, J. Cerezo, E. Stendardo, R. Improta, and F. Santoro, *Insights for an accurate comparison of computational data to experimental absorption and emission spectra: beyond the vertical transition approximation*, J. Chem. Theory Comput. **9**, 2072 (2013).
- [306] A. Charaf-Eddin, A. Planchat, B. Mennucci, C. Adamo, and D. Jacquemin, *Choosing a functional for computing absorption and fluorescence band shapes with TD-DFT*, J. Chem. Theory Comput. **9**, 2749 (2013).
- [307] H. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. **91**, 43 (1995).
- [308] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, *Automatic atom type and bond type perception in molecular mechanical calculations*, J. Mol. Graph **25**, 247 (2006).
- [309] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *Development and testing of a general amber force field*, J. Comput. Chem. **25**, 1157 (2004).
- [310] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, *LINCS: A linear constraint solver for molecular simulations*, J. Comput. Chem. **18**, 1463 (1997).
- [311] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, J. Mol. Graph. **14**, 33 (1996).
- [312] J. Neugebauer, E. Jan Baerends, M. Nooijen, and J. Autschbach, *Importance of vibronic effects on the circular dichroism spectrum of dimethyloxirane*, J. Chem. Phys. **122**, 234305 (2005).

- [313] Y. Nakai, T. Mori, and Y. Inoue, *Circular dichroism of (di) methyl- and diaza [6] helicenes. A combined theoretical and experimental study*, J. Phys. Chem. A **117**, 83 (2013).
- [314] D. Padula, J. Cerezo, G. Pescitelli, and F. Santoro, *The shape of the electronic circular dichroism spectrum of (2, 6-dimethylphenyl)(phenyl) methanol: Interplay between conformational equilibria and vibronic effects*, Phys. Chem. Chem. Phys. **19**, 32349 (2017).
- [315] S. Höfener, B. A. Günther, M. E. Harding, and L. H. Gade, *Understanding UV-vis spectra of halogenated tetraazaperopyrenes (TAPPs): A computational study*, J. Phys. Chem. A **123**, 3160 (2019).
- [316] D. Jacquemin, E. Brémond, I. Ciofini, and C. Adamo, *Impact of vibronic couplings on perceived colors: two anthraquinones as a working example*, J. Phys. Chem. Lett. **3**, 468 (2012).
- [317] S. Di Tommaso, D. Bousquet, D. Moulin, F. Baltenneck, P. Riva, H. David, A. Fadli, J. Gomar, I. Ciofini, and C. Adamo, *Theoretical approaches for predicting the color of rigid dyes in solution*, J. Comput. Chem. **38**, 998 (2017).
- [318] R. S. Becker, J. Seixas de Melo, A. L. Macanita, and F. Elisei, *Comprehensive evaluation of the absorption, photophysical, energy transfer, structural, and theoretical properties of  $\alpha$ -oligothiophenes with one to seven rings*, J. Phys. Chem. **100**, 18683 (1996).
- [319] R. Improta, F. J. Ferrer, E. Stendardo, and F. Santoro, *Quantum-classical calculation of the absorption and emission spectral shapes of oligothiophenes at low and room temperature by first-principle calculations*, ChemPhysChem **15**, 3320 (2014).
- [320] J. Cerezo, F. J. Avila Ferrer, G. Prampolini, and F. Santoro, *Modeling solvent broadening on the vibronic spectra of a series of coumarin dyes. From implicit to explicit solvent models*, J. Chem. Theory Comput. **11**, 5810 (2015).
- [321] R. Zalesny, N. A. Murugan, F. Gel'mukhanov, Z. Rinkevicius, B. Osmałowski, W. Bartkowiak, and H. Ågren, *Toward fully nonempirical simulations of optical band shapes of molecules in solution: A case study of heterocyclic ketoimine difluoroborates*, J. Phys. Chem. A **119**, 5145 (2015).
- [322] A. Segalina, J. Cerezo, G. Prampolini, F. Santoro, and M. Pastore, *Accounting for vibronic features through a mixed quantum-classical scheme: Structure, dynamics, and absorption spectra of a perylene diimide dye in solution*, J. Chem. Theory Comput. **16**, 7061 (2020).
- [323] P. P. Fehér, Á. Madarász, and A. Stirling, *Multiscale modeling of electronic spectra including nuclear quantum effects*, J. Chem. Theory Comput. **17**, 6340 (2021).

- [324] T. M. N. Nguyen and C. König, *Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers*, Phys. Chem. Chem. Phys. **24**, 14825 (2022).
- [325] W. Domcke, L. Cederbaum, H. Köppel, and W. von Niessen, *A comparison of different approaches to the calculation of Franck-Condon factors for polyatomic molecules*, Mol. Phys. **34**, 1759 (1977).
- [326] H. Du, R.-C. A. Fuh, J. Li, L. A. Corkan, and J. S. Lindsey, *PhotochemCAD: A computer-aided design and research tool in photochemistry*, Photochem. Photobiol. **68**, 141 (1998).
- [327] J. F. Deye, T. A. Berger, and A. G. Anderson, *Nile Red as a solvatochromic dye for measuring solvent strength in normal liquids and mixtures of normal liquids with supercritical and near critical fluids*, Anal. Chem. **62**, 615 (1990).
- [328] A. Ghanadzadeh, A. Zeini, A. Kashef, and M. Moghadam, *Solvent polarizability and anisotropy effects on the photophysical behavior of oxazine 1: An appropriate polarizability indicator dye*, Spectrochim. Acta A Mol. Biomol. Spectrosc. **73**, 324 (2009).
- [329] H. Tajalli, A. G. Gilani, M. Zakerhamidi, and P. Tajalli, *The photophysical properties of Nile red and Nile blue in ordered anisotropic media*, Dyes Pigm. **78**, 15 (2008).
- [330] J.-P. Knemeyer, N. Marmé, and M. Sauer, *Probes for detection of specific DNA sequences at the single-molecule level*, Anal. Chem. **72**, 3717 (2000).
- [331] J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, and P. Tinnefeld, *Controlling the fluorescence of ordinary oxazine dyes for single-molecule switching and superresolution microscopy*, Proc. Natl. Acad. Sci. **106**, 8107 (2009).
- [332] C. Kjær and S. B. Nielsen, *Luminescence spectroscopy of oxazine dye cations isolated in vacuo*, Phys. Chem. Chem. Phys. **21**, 4600 (2019).
- [333] A. Czímerová, N. Iyi, and J. Bujdák, *Energy transfer between rhodamine 3B and oxazine 4 in synthetic-saponite dispersions and films*, J. Colloid Interface Sci. **306**, 316 (2007).
- [334] S. Saha, D. Majhi, K. Bhattacharyya, N. Preeyanka, A. Datta, and M. Sarkar, *Evidence of homo-FRET in quantum dot–dye heterostructured assembly*, Phys. Chem. Chem. Phys. **20**, 9523 (2018).
- [335] C. Scheibler, *Ueber das betain, eine im safte der zuckerrüben (Beta vulgaris) vorkommende pflanzenbase*, Ber. Dtsch. Chem. Ges. **2**, 292 (1869).
- [336] C. Scheibler, *Ueber das Betain und seine Constitution*, Ber. Dtsch. Chem. Ges. **3**, 155 (1870).

- [337] C. Kjær, J. M. Lisy, and S. B. Nielsen, *Gas-Phase Ion Spectroscopy of Congo Red Dianions and Their Complexes with Betaine*, *J. Phys. Chem. A* **122**, 3211 (2018).
- [338] C. Blake and S. Oatley, *Protein–DNA and protein–hormone interactions in prealbumin: a model of the thyroid hormone nuclear receptor?*, *Nature* **268**, 115 (1977).
- [339] J. Hamilton and M. Benson, *Transthyretin: a review from a structural perspective*, *Cell. Mol. Life Sci.* **58**, 1491 (2001).
- [340] R. I. Campos, X. Wu, M. Elgland, P. Konradsson, and P. Hammarström, *Novel trans-stilbene-based fluorophores as probes for spectral discrimination of native and protofibrillar transthyretin*, *ACS Chem. Neurosci.* **7**, 924 (2016).
- [341] M. L. Müller, J. Butler, and B. Heidecker, *Emerging therapies in transthyretin amyloidosis—a new wave of hope after years of stagnancy?*, *Eur. J. Heart Fail.* **22**, 39 (2020).
- [342] L. C. Menikarachchi and J. A. Gascón, *QM/MM approaches in medicinal chemistry research*, *Curr. Top. Med. Chem.* **10**, 46 (2010).
- [343] H. M. Senn and W. Thiel, *QM/MM methods for biomolecular systems*, *Angew. Chem., Int. Ed.* **48**, 1198 (2009).
- [344] A. Warshel and M. Levitt, *Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme*, *J. Mol. Biol.* **103**, 227 (1976).
- [345] H. Lin and D. G. Truhlar, *QM/MM: what have we learned, where are we, and where do we go from here?*, *Theoretical Chemistry Accounts* **117**, 185 (2007).
- [346] M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini, and B. Menucci, *Polarizable embedding QM/MM: the future gold standard for complex (bio) systems?*, *Phys. Chem. Chem. Phys.* **22**, 14433 (2020).
- [347] A. Altun, S. Yokoyama, and K. Morokuma, *Quantum mechanical/molecular mechanical studies on spectral tuning mechanisms of visual pigments and other photoactive proteins*, *Photochem. Photobiol.* **84**, 845 (2008).
- [348] N. H. List, J. M. H. Olsen, and J. Kongsted, *Excited states in large molecular systems through polarizable embedding*, *Phys. Chem. Chem. Phys.* **18**, 20234 (2016).
- [349] E. Boulanger and J. N. Harvey, *QM/MM methods for free energies and photochemistry*, *Curr. Opin. Struct. Biol.* **49**, 72 (2018).
- [350] F. Segatta, L. Cupellini, M. Garavelli, and B. Menucci, *Quantum chemical modeling of the photoinduced activity of multichromophoric biosystems: focus review*, *Chem. Rev.* **119**, 9361 (2019).



- [351] R. A. Copeland, D. L. Pompliano, and T. D. Meek, *Drug–target residence time and its implications for lead optimization*, Nat. Rev. Drug Discov. **5**, 730 (2006).
- [352] N. Yin, J. Pei, and L. Lai, *A comprehensive analysis of the influence of drug binding kinetics on drug action at molecular and systems levels*, Mol. Biosyst. **9**, 1381 (2013).
- [353] R. A. Copeland, *The drug–target residence time model: a 10-year retrospective*, Nat. Rev. Drug Discov. **15**, 87 (2016).
- [354] Q. Shao and W. Zhu, *Exploring the ligand binding/unbinding pathway by selectively enhanced sampling of ligand in a protein–ligand complex*, J. Phys. Chem. B **123**, 7974 (2019).
- [355] P. Tiwary and A. van de Walle, *A Review of enhanced sampling approaches for accelerated molecular dynamics*, pages 195–221, Springer International Publishing, Cham, 2016.
- [356] J. M. L. Ribeiro, S.-T. Tsai, D. Pramanik, Y. Wang, and P. Tiwary, *Kinetics of ligand–protein dissociation from all-atom simulations: Are we there yet?*, Biochem. **58**, 156 (2018).
- [357] E. Vanden-Eijnden and F. A. Tal, *Transition state theory: Variational formulation, dynamical corrections, and error estimates*, J. Chem. Phys. **123**, 184103 (2005).
- [358] Y. Sakae and Y. Okamoto, *Optimization of protein force-field parameters with the Protein Data Bank*, Chem. Phys. Lett. **382**, 626 (2003).
- [359] Y. Sakae and Y. Okamoto, *Secondary-structure design of proteins by a backbone torsion energy*, J. Phys. Soc. Japan **75**, 054802 (2006).
- [360] S. Weerasinghe and P. E. Smith, *A Kirkwood–Buff derived force field for mixtures of urea and water*, J. Phys. Chem. B **107**, 3891 (2003).
- [361] S. Weerasinghe and P. E. Smith, *A Kirkwood–Buff derived force field for sodium chloride in water*, J. Chem. Phys. **119**, 11342 (2003).
- [362] M. B. Gee, N. R. Cox, Y. Jiao, N. Benteinitis, S. Weerasinghe, and P. E. Smith, *A Kirkwood–Buff derived force field for aqueous alkali halides*, J. Chem. Theory Comput. **7**, 1369 (2011).



APPENDIX

---

A.1 TAILORED ANHARMONIC–HARMONIC VIBRATIONAL PROFILES FOR  
FLUORESCENT BIOMARKERS

The publication:

**Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers,**

T.M.N. Nguyen and C. König,

*Phys. Chem. Chem. Phys.*, 2022, 24, 14825-14835,

DOI: 10.1039/d2cp01486f,

is included in this thesis under the terms of the Attribution 3.0 Unported (CC BY 3.0) license.


 Cite this: *Phys. Chem. Chem. Phys.*,  
 2022, 24, 14825

## Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers†

 Nghia Nguyen Thi Minh and Carolin König \*

We propose a hybrid anharmonic–harmonic scheme for vibrational broadenings, which embeds a reduced-space vibrational configuration interaction (VCI) anharmonic wave function treatment in the independent-mode displaced harmonic oscillator (IMDHO) model. The resulting systematically-improvable VCI-in-IMDHO model allows including the vibronic effects of all vibrational degrees of freedom, while focusing the effort on the important degrees of freedom with minimal extra computational effort compared to a reduced-space VCI treatment. We show for oligothiophene examples that the VCI-in-IMDHO approach can yield accurate vibrational profiles employing smaller vibrational spaces in the VCI part than the reduced-space VCI approach. By this, the VCI-in-IMDHO model enables accurate calculation of vibrational profiles of common fluorescent dyes with more than 100 vibrational degrees of freedom. We illustrate this for three examples of fluorescent biomarkers of current interest. These are the oligothiophene-based fluorescent dye called HS84, 1,4-diphenylbutadiene, and an anthracene diimide. For all examples, we assess the impact of the anharmonic treatment on the vibrational broadening, which we find to be more pronounced for the intensities than for the peak positions.

 Received 30th March 2022,  
 Accepted 4th June 2022

DOI: 10.1039/d2cp01486f

[rsc.li/pccp](https://rsc.li/pccp)

### 1 Introduction

The calculation of accurate vibrational line shapes in optical spectra is of interest in various ways: they are needed for direct comparison of theoretical to experimental spectra<sup>1–4</sup> as well as for elucidating the colour of a dye.<sup>5,6</sup> Optical line shapes are therefore also essential for the rational design of dyes.<sup>7</sup> Moreover, the line shape of emission peaks can significantly be altered within different environments. This is for example the case for the oligothiophene-based fluorescent dye denoted HS84: in solution, a broad fluorescent signal is obtained, while a structured fluorescent signal occurs for HS84 mixed with recombinant A $\beta$  1–42 amyloid-like fibrils.<sup>8</sup> Deposition of amyloid fibrils is a hallmark for neurodegenerative diseases such as Alzheimer's and Parkinson's disease.<sup>9</sup> The aggregation of the HS84 dye to the amyloid fibril is associated with a change of the fluorescence colour. This colour change is of particular relevance as it is a promising route for improved detection of amyloid fibrils also *in vivo*.<sup>10</sup> Such luminescent conjugated

oligothiophenes (LCOs) have, hence, been developed as candidates for improved diagnosis of these wide-spread diseases by fluorescence imaging. LCOs have been shown to detect a wider range of disease-associated protein aggregates<sup>10–12</sup> than more conventional ligands such as thioflavin S or congo red.

Given that the line shape of fluorescent spectra of the oligothiophenes<sup>13</sup> can well be recovered from theoretical models for vibrational broadenings,<sup>14–16</sup> the observed structure in the fluorescence spectrum can be assigned to vibronic couplings. Still, supporting computational studies to understand the mechanism behind the colour change rely fully on conformational averaging of vertical transition and cannot recover the line shape in the emission spectrum.<sup>17</sup>

The main reason, why vibrational line shapes are typically not considered in such studies is their large computational cost, particularly when anharmonic effects need to be accounted for. In ref. 16, a reduced-space anharmonic model is proposed, in which a selected vibrational subspace is treated by means of vibrational configuration interaction (VCI) wave functions<sup>18–21</sup> using anharmonic potential energy surfaces accounting for mode–mode couplings and correlations. It was found that the vibrational progression of oligothiophenes is dominated by only a small number of modes.<sup>15,16</sup> Notably, the number of important modes does not increase significantly with an increasing number of thiophene rings.<sup>16</sup>

For estimating the impact of individual modes on the vibronic profile, we previously applied the harmonic displacement

*Institut für Physikalische Chemie und Elektrochemie, Leibniz Universität Hannover, Callinstr. 3A, 30167 Hannover, Germany. E-mail: carolin.koenig@pci.uni-hannover.de*

† Electronic supplementary information (ESI) available: Analysis of one-mode anharmonicity on the vibrational profile for all tested molecules; mode ranking of all-*trans* quaterthiophene, all-*trans* pentathiophene, and the organic dyes; details of vibrational profiles of all-*trans* quaterthiophene, all-*trans* pentathiophene; pure IMDHO vibronic spectrum of all investigated organic dyes in a respective solvent. See DOI: <https://doi.org/10.1039/d2cp01486f>



between the minima of the two involved electronic states within the independent mode displaced harmonic oscillator (IMDHO) model.<sup>22,23</sup> The IMDHO model is a rather crude approximation, that allows the calculation of vibrational line shapes from harmonic frequencies of the initial state and the atomic gradients of the final state at the equilibrium structure of the initial state alone.

In this work, we suggest using the information on the IMDHO-type vibrational line shape not only for the assessment of the impact of a mode on the vibrational spectrum, but also for treating all modes that are considered less important or to a high degree harmonic. By this, we obtain a hybrid scheme for vibrational broadenings, which we denote VCI-in-IMDHO model. This means, we treat a selected set of vibrational modes by VCI methods and all others within the IMDHO model. In contrast to other hybrid schemes for vibrational line shapes,<sup>14,24–29</sup> the VCI-in-IMDHO is not a mixed quantum–classical approach, but rather a mixed quantum–quantum scheme. So far, this model is restricted to Frank-Condon factors, which can be assumed to be a good approximation for the bright transitions considered in this study.

After validation of the VCI-in-IMDHO method for oligothiophenes with two to five thiophene rings, we apply it to the fluorescent dyes depicted in Fig. 1. Among those is the HS84 fluorescent marker for amyloid fibrils already mentioned above. It is a pentameric oligothiophene derivative that belongs to the LCO dyes.<sup>30</sup> Additionally, we study 1,4-diphenylbutadiene (DPB) and an anthracene diimide (ADI) derivate. DPB is a fluorescent derivative of linear polyenes, which have been extensively investigated as a model for the photochemical and photobiological systems.<sup>31,32</sup> ADIs and related compounds are functional organic dyes with outstanding chemical, thermal, and photochemical properties.<sup>33–35</sup> During the past decade, this class of dyes has attracted increasing interest as promising scaffolds for several exciting applications such as photovoltaic

cells,<sup>34,36</sup> chemosensors,<sup>37,38</sup> fluorescence dyes in biological media.<sup>39</sup>

This work is organized as follows. First, we present the VCI-in-IMDHO model and the measures for the division of the vibrational spaces (Section 2). Second, the computational details are presented in Section 3. The results are discussed and compared to the experiment in Section 4 and followed by the conclusion (Section 5).

## 2 Methodology

In the VCI-in-IMDHO model, we treat all important anharmonic modes with an accurate VCI wave function, which accounts for the one-mode anharmonicity as well as for mode–mode couplings and correlations, usually restricted to low orders. For all other modes, we employ the IMDHO model, which is purely harmonic and neglects frequency differences in the involved electronic states as well as Duschinsky rotations.<sup>40</sup> It, however, has the advantage that it offers closed formulas for the Franck-Condon factors:<sup>22,23,41–43</sup> the Franck-Condon factors from the vibrational ground state (0) of the initial electronic state to the  $n$ th vibrational state of the final electronic state can in the IMDHO model be expressed in atomic units as<sup>23,41</sup>

$$f_{n-0} = \frac{1}{n!} \left( \frac{A_i^2}{2} \right)^n \exp\left(-\frac{A_i^2}{2}\right) \quad (1)$$

where  $A_i$  is the dimensionless normal-mode displacement of the two involved electronic states along the considered normal mode  $i$ . Furthermore, the vibronic profile of an emission band within the IMDHO model can directly be calculated *via* its auto-correlation function as<sup>41,44,45</sup>

$$I_{n-m}(\omega) = \frac{4\pi\omega}{3c} |\mu^{\text{el}}|^2 \Re \int_0^{\infty} \exp(i[\omega - (E_{m,0} - E_{n,0})]t) \cdot \exp(-\gamma t) \times \prod_{j=1}^{N_{\text{modes}}} \exp\left[-\frac{A_j^2}{2}(1 - \exp(i\omega_j t))\right] dt, \quad (2)$$

where  $c$  is the speed of light,  $\mu^{\text{el}}$  is the electronic transition dipole moment, which is assumed to be constant in the present work.  $\omega_j$  is the angular vibrational frequency of normal mode  $j$  in the initial state and  $A_j$  the harmonic displacement of normal mode  $j$  in the same state. Both,  $\omega_j$  and  $A_j$ , are assumed to be identical in both involved electronic states.  $E_{n,0}$  is the energy of vibrational level 0 in electronic state  $n$ . In the IMDHO model, the difference  $E_{m,0} - E_{n,0}$  is equal to the adiabatic excitation energy between the electronic states  $m$  and  $n$ .  $N_{\text{modes}}$  is the total number of normal modes.  $\gamma$  is the half width at half maximum (HWHM) for the underlying Lorentz line-shape function of the individual vibronic peaks. Hence,  $\gamma$  determines the broadening of the vibronic peaks. It is chosen prior to the computation of the IMDHO line shape. This time-dependent approach is particularly beneficial for large systems, as the accumulated intensity can be calculated for every frequency of

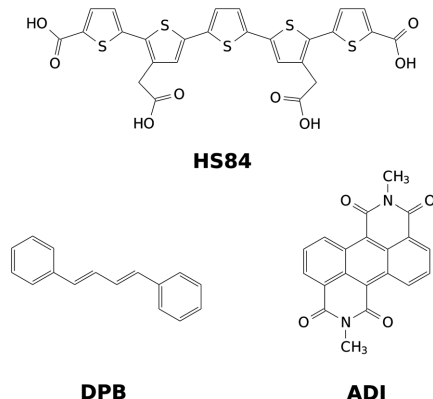


Fig. 1 Representation of the investigated dyes: pentameric oligothiophene derivative (HS84), 1,4-diphenylbutadiene (DPB) and anthracene diimide (ADI).

interest rather than calculating a large number of Franck–Condon factors.<sup>45,46</sup>

In our hybrid scheme, we chose a small subset of modes for the VCI calculations leading to a set of Franck–Condon factors. The resulting stick spectra are then broadened with a line shape obtained from the IMDHO autocorrelation function [eqn (2)] for the remaining subset of modes,

$$I_{n \leftarrow m}(\omega) = \frac{4\pi\omega}{3c} |\mu^{el}|^2 \sum_s^{N_{\text{states}}} \Re \int_0^\infty \exp\left(i\left[\omega - \left(E_{m,0}^{\text{VCI}} - E_{n,s}^{\text{VCI}}\right)t\right]\right) \times \exp(-\gamma t) \prod_{j=1}^{N_{\text{HO modes}}} \exp\left[-\frac{A_j^2}{2}(1 - \exp(i\omega_j t))\right] dt, \quad (3)$$

where we introduce a sum over all  $N_{\text{states}}$  vibrational states in the final electronic state obtained by anharmonic VCI calculations. The energy of the corresponding vibronic state  $E_{n,s}^{\text{VCI}}$  equals the sum of the electronic energy of this state and the VCI vibrational energy of the respective vibrational state  $s$ . Similarly,  $E_{m,0}^{\text{VCI}}$  corresponds to the electronic energy of the electronic state  $m$  plus the anharmonic zero-point energy. Further, the product accounting for vibrational broadening in the IMDHO is now restricted to those  $N_{\text{HO modes}}$  modes not accounted for in the VCI treatment. Note that in case all modes are accounted for anharmonically, the Lorentzian line shape for all peaks is recovered within this methodology.

For the above-described hybrid scheme, the modes must be divided into two sets. This division is done based on three criteria:

(1) Importance of the mode for vibrational progression: Harmonic displacement ( $A_j$ ) based on the IMDHO model.

This first measure estimates the impact of the mode under consideration on the vibrational progression, *i.e.*, whether more than the 0–0 transition are significant. Similar as in ref. 16 and 47, we use the harmonic displacements  $A_j$  to rank the modes with decreasing importance and assess the threshold by the convergence of the theoretical spectra.

(2) Resolution of the vibrational progression with a given HWHM ( $\gamma$ ): crossing point between the Lorentzian broadening the 0–0 and 1–0 transition ( $\sigma$ ).

Besides the ratio of the Franck–Condon factors, it is also important that their effect is also resolved when assuming a line broadening, *i.e.*, that the additional peaks are not hidden below the broadened lines. In ref. 16, this was accounted for by discarding all vibrational modes with frequencies smaller than half the chosen HWHM. This approach, however, does not account for the relation between the height of the satellite peaks (in the IMDHO model determined by the displacement  $A_j$ ) and the resolution. We therefore replace this pure frequency criterion by a more nuanced criterion, which is the HWHM at which the maximum of the  $1 \leftarrow 0$  peak broadened with a Lorentzian line shape lies on the Lorentzian band of the  $0 \leftarrow 0$  peak. Within the IMDHO approximation, this value is given by

$$\sigma_i = \frac{\nu_i}{\sqrt{\frac{2}{A_i^2} - 1}}, \quad (4)$$

where  $\nu_i$  is the harmonic frequency of the mode  $i$ . We only include modes for which  $\sigma_i$  is greater than the chosen HWHM.

(3) Impact of one-mode anharmonicity on the vibrational profile: with this third measure for anharmonicity, we aim at an estimate, whether individual modes need to be treated anharmonically or whether a harmonic treatment within the IMDHO model is sufficient. We assess this by comparing the one-mode vibrational profiles within the IMDHO model and anharmonic treatment. Additionally, we also consider the ratio  $\left(\frac{A_i^a}{A_i}\right)^2$  between the squared anharmonic ( $A_i^a$ ) and harmonic displacements ( $A_i$ ). For this, we project the difference between the optimized structures in the initial and final electronic states on the normal modes employed in the IMDHO model. By this procedure, we obtain the anharmonic displacements  $A_i^a$  for all modes  $i$ . As analysed below, these two measures do not correlate clearly and we focus by default on the comparison of the one-mode profiles.

By a combination of these measures, we identify those modes that are important to be treated accurately. We stress that one criterion alone cannot achieve this. For example, the vibrational progression caused by a mode with a high displacement, may not be seen due to a poor resolution, or it might already be well described in the harmonic picture. In these cases, a computationally expensive VCI treatment is not in proportion to the minor improvements to be expected. Similarly, some very anharmonic modes, such as certain torsion motions,<sup>14,27</sup> do not necessarily contribute significantly to the vibrational progression. Also here, a VCI treatment may not be beneficial due to the accuracy–cost considerations.

Within this study, no attempts were made to include finite-temperature effects.

### 3 Computational details

All electronic-structure calculations have been performed with the Gaussian16 program package<sup>48</sup> using (time-dependent) density-functional theory [(TD)DFT] with the CAM-B3LYP<sup>49</sup> density functional and a def2-TZVP basis set.<sup>50,51</sup> The calculations are performed for the dyes in vacuum, if not specified otherwise. To get a rough estimate of the impact of the solvent environment on the vibrational progression, we performed the pure IMDHO computations additionally for each dye using the integral equation formalism of the polarizable-continuum model (IEF-PCM)<sup>52,53</sup> with the default dielectric constants in Gaussian16. We choose for each dye the solvent of the respective experimental spectrum compared to, that is, water for HS84 (note that the experiment was done in a phosphate buffered saline), hexane for 1,4-diphenylbutadiene and dichloromethane for the anthracene diimide.

For oligothiophenes, we have used the potential energy surfaces (PESs) from ref. 16. For new combinations of coordinates treated anharmonically, not present in ref. 16, we have deleted the respective regions from potential energy surfaces generated in ref. 16.



For the PES generation and anharmonic vibrational wave function calculations, we employed a locally modified version of the Molecular Interactions Dynamics And Simulation C++ package (MidasCpp)<sup>54</sup> version 2019.04.0. For all vibrational profile calculations, we employed the normal modes and vibrational frequencies of the first excited state. The vibrational coordinates of the excited state were also employed for the reduced-space anharmonic PESs for both the ground and excited state. The ground- and excited-state PESs for the fluorescent dyes were obtained with the multi-state extension<sup>55,56</sup> to the adaptive density guided approach (ADGA)<sup>57</sup> obtaining reliable PESs for at least the lowest six vibrational states in both electronic states. Additionally, mode–mode coupling up to second order in  $n$ -mode expansion is considered in the PES generation within a selected set of modes. The relative ADGA convergence criterion threshold was set to  $1.0 \times 10^{-2}$  while the absolute ADGA convergence criterion threshold was set to  $1.0 \times 10^{-6}$ . Both was increased by a factor of 10 for the two-mode coupling part of the PES. The PES cuts were fitted to polynomials with the maximal polynomial order of eight. All vibrational self-consistent field (VSCF) calculations were performed with B-spline basis sets<sup>58</sup> with a basis set density of 0.8.

The anharmonic Franck–Condon factors were obtained from vibrational configuration interaction (VCI) calculations. For all cases with four or more modes in the VCI treatment up to quadruply substituted configurations are included in the VCI treatment (VCI[4]). For smaller vibrational subspaces, full vibrational configuration interaction (FVCI) calculations have been conducted. The VCI Franck–Condon factors<sup>59–62</sup> were obtained as overlaps of the anharmonic VCI wave functions of the different vibrational states within the final electronic state with the vibrational ground state of the initial electronic state. The sum of FCs was required to be more than 0.98 to ensure that most of the intensity was captured in the calculated spectra. To reach this, different numbers modals were included in the configurational space for different molecules. The calculations of the Franck–Condon factors were performed using the respective implementation<sup>16,56</sup> in MidasCpp.<sup>54</sup>

The different measures as well as the IMDHO line shapes according to eqn (2) have been implemented in a python framework. Therefore, `scipy.integrate.quad` with techniques from the Fortran library QUADPACK<sup>63</sup> is used for the integration parts.

For a consistent theory–experiment comparison, the experimental spectra measured in wavelength scale are transformed in line shapes by applying an intensity correction proportional to  $\omega^2$ .<sup>64,65</sup> We further normalized the area under the peak to 1 in all spectral comparisons and shifted the maxima of the experiment to the same position as the maxima of the respective most advanced calculation.

## 4 Results and discussion

### 4.1 Oligothiophenes, revisited

The above-mentioned reduced-space approach for anharmonic vibrational broadenings was assessed for oligothiophenes of different lengths.<sup>16</sup> In this previous work, all chosen vibrations

(*i.e.* all with  $A_i \leq 0.6$  and  $\nu_i > 161 \text{ cm}^{-1}$ ) were included anharmonically and all others were neglected. We have chosen these examples as first testbed for our hybrid approach and the refined criteria described above. In view of the similar spectra obtained with different conformers of the oligothiophenes,<sup>16</sup> we restrict the discussion here to the all-*trans* conformers. The experimental spectra we compare to ref. 13 are taken in ethanol at 77 K, that is, below the melting point of ethanol. Hence, thermal broadening is likely suppressed in these spectra.

**4.1.1 In-depth analysis of *trans* bithiophene.** Following the outline of ref. 16, we first evaluate the inclusion of anharmonic modes with decreasing  $A_i$  for the *trans* bithiophene. The corresponding mode ranking is listed in Table 1. This contains next to the previous criteria (frequency  $\nu_i$  and harmonic displacement  $A_i$ ) also the projected anharmonic displacements

$|A_i^a|$  and squared ratio  $\left(\frac{A_i^a}{A_i}\right)^2$  as well as the resolution measure  $\sigma_i$  and the qualitative impact of one-mode anharmonicity in the respective anharmonic calculation for this mode only. The latter is obtained by visual inspection of the respective spectra, which can be found in Fig. S1 in the ESI†

In the previous work, modes 14, 36, and 7 were deemed important. Additionally, modes 5 and 29 were assessed. The resolution measure  $\sigma_i$  for the modes 5 and 7 lies below the chosen  $\gamma$  value of  $322.6 \text{ cm}^{-1}$  (0.04 eV). This suggests that mode 5 and 7 are likely not well resolved and may therefore be neglected. Fig. 2 shows the vibrational broadening of the emission spectrum obtained with the IMDHO model and by VCI for the following reduced spaces, containing

set 1: only the mode with the largest IMDHO displacement, *i.e.*, mode 14,

set 1\*: only the mode with most pronounced one-mode anharmonicity among the five modes with the largest IMDHO displacement, *i.e.*, mode 36,

set 2: three modes with the largest IMDHO displacements according to the procedure in ref. 16, *i.e.*, modes 14, 36, 7,

set 2\*: three modes with the largest IMDHO displacements and a resolution measure larger than the chosen half width at half maximum, *i.e.*, modes 14, 36, and 29, and

set 3: all five previously investigated modes, *i.e.*, modes 14, 36, 7, 5, and 29.

**Table 1** Mode ranking for *trans* bithiophene with CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|A_i|$ , anharmonic  $|A_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{A_i^a}{A_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH)

Mode (i)	$\nu_i/\text{cm}^{-1}$	$ A_i $	$ A_i^a $	$\left(\frac{A_i^a}{A_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-Mode AH
14	674	1.420	1.593	1.259	7473	Weak
36	1658	1.350	0.422	0.098	5312	Moderate
7	387	0.824	1.606	3.799	277	Weak
5	290	0.529	0.415	0.615	117	Weak
29	1214	0.446	0.303	0.461	403	Weak



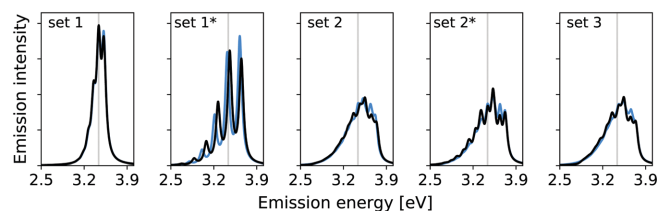


Fig. 2 Reduced-space vibrational profile for the  $S_0 \leftarrow S_1$  emission for *trans* bithiophene with the IMDHO model (blue) and VCI (black). For the definition of the reduced spaces, see the main text. The vertical transition energy is given as a vertical grey line.

For mode 14, which has the largest harmonic displacement, the one-mode harmonic and anharmonic spectra are very similar (set 1), while the respective spectra for only mode 36 (set 1\*) differ significantly. The same also holds for all other sets, that are set 2, set 2\*, and set 3, which all include the anharmonic mode 36. We further observe significant differences between set 2 and set 2\*, which both contain three modes.

Fig. 3 shows the vibrational profiles, for which the above-mentioned sets of modes were treated by VCI and all other modes by the IMDHO model. Additionally, set 0 is depicted, which is a pure IMDHO treatment of all vibrational modes. The respective vibrational profile is hardly altered, when treating mode 14 anharmonically (set 1), but slight differences are obtained for treating only mode 36 (set 1\*) anharmonically. Similar slight modifications of the vibrational profiles are observed for set 2 and set 2\*. The vibrational profile for the hybrid scheme with set 3 is almost perfectly resembled by that for set 2\*. For a more quantitative comparison of the convergence of the vibrational profiles to that obtained for the hybrid scheme with the different sets of modes, we quantify the difference between the two spectra by

$$d = \frac{\int_a^b |f(x) - g(x)| dx}{\int_a^b g(x) dx} \quad (5)$$

here  $f(x)$  and  $g(x)$  is the approximated and reference spectrum, respectively. As reference spectrum, we choose our best model,

*i.e.*, the hybrid calculation for set 3. The  $d$  values are shown below the respective spectra in Fig. 3. The  $d$  value for set 2\* is lower than that for set 2. For the example of bithiophene we, hence, conclude that both the consideration of anharmonicities and the refined resolution measure as well as the inclusion of the less important modes within the IMDHO model can be beneficial for increased convergence of the vibrational profile with increasing number of modes.

**4.1.2 All-*trans* terthiophene.** For the terthiophene, we obtain the importance and anharmonicity measures shown in Table 2. All these modes have been considered in the previous work.<sup>16</sup> The set including all these modes will be denoted set 3 in the following. Two of these modes, that are mode 8 and mode 10, may be discarded due to our resolution criterion  $\sigma_i$  being lower than the chosen half width at half maximum. The resulting set of modes includes modes 55, 24 and is denoted set 2. Mode 55 is the mode with the largest IMDHO displacement and exhibits the most pronounced one-mode anharmonicity among the listed modes. Mode 55 is the only mode in set 1.

The calculated hybrid spectra treating above-described sets of modes anharmonically are depicted in Fig. 4. Again, we see a convergence of the spectrum towards the set 3 results already for set 2. This convergence is illustrated by the decreasing  $d$  measures, also shown in Fig. 4. In this case, however, the consideration of the remaining modes within the IMDHO model has only little influence on the result compared to the reduced-space VCI result already presented in ref. 16. The

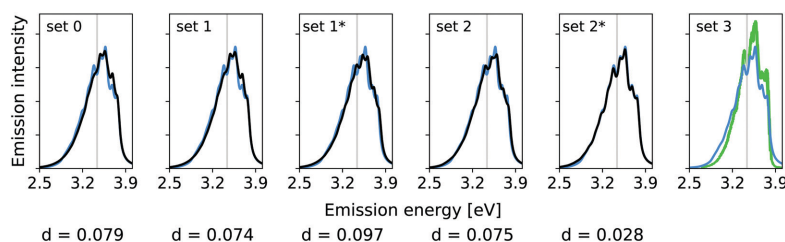
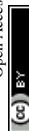


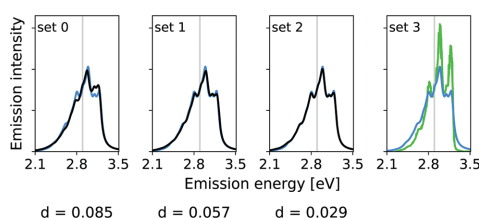
Fig. 3 Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for *trans* bithiophene for the different sets of vibrational coordinates treated by VCI (black) and that for the largest set of modes set 3 (blue) with a HWHM of 0.04 eV compared to a corrected experimental spectrum of bithiophene at 77 K in ethanol<sup>15</sup> (green). For the definition of the reduced spaces, see the main text. The experimental spectrum is shifted by 0.105 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eqn (5) is quantified below the respective graphs.





**Table 2** Mode ranking for all-*trans* terthiophene with CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|A_i|$ , anharmonic  $|A_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{A_i^a}{A_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$  and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH)

Mode	$\nu_i/\text{cm}^{-1}$	$ A_i $	$ A_i^a $	$\left(\frac{A_i^a}{A_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-Mode AH
55	1631	1.31	0.342	0.068	4009	Moderate
24	697	1.050	1.233	1.378	772	Weak
8	210	0.977	0.629	1.116	181	Weak
10	350	0.741	1.371	3.423	215	Weak



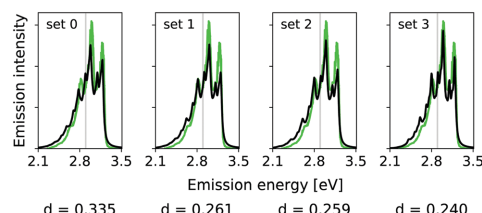
**Fig. 4** Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* terthiophene for the different sets of vibrational coordinates treated by VCI (black) and that for the largest set of modes set 3 (blue) with a HWHM of 0.04 eV compared to a corrected experimental spectrum of terthiophene at 77 K in ethanol<sup>13</sup> (green). For the definition of the reduced spaces, see the main text. The vertical transition energy is given as a vertical grey line. The experimental spectrum is shifted by 0.093 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eqn (5) is quantified below the respective graphs.

agreement with experiment (most right graph in Fig. 4), is, however, only moderate. We note that the experimental data is obtained in ethanol solution, so that many factors may play a role, that are not accounted for in our present computational setup. Note also that better agreement with the experiment is obtained with a reduced HWHM value of  $\gamma = 0.02$  eV (see Fig. 5).

In this case, however, already the pure IMDHO model (set 0) results in good agreement with experiment. Still, the VCI-in-IMDHO treatment leads to a systematic further improvement of this agreement as illustrated by the  $d$  values.

**4.1.3 All-*trans* quaterthiophene and all-*trans* pentathiophene.** The results for the all-*trans* quaterthiophene and all-*trans* pentathiophene are overall similar to that of the terthiophene and can be found in Sections S3 and S4 in the ESI†. Also here, we can discard vibrational modes due to the more refined resolution measure and observe similar convergence behaviour.

**4.1.4 Assessment of the anharmonicity measure.** In the above discussions, we assessed one-mode anharmonicity measure comparing the obtained line shapes for one-mode calculations. In the respective tables, we however also list an anharmonicity measure obtained by comparing the IMDHO displacements to the anharmonic ones obtained by projecting the coordinate difference of the ground and excited state onto



**Fig. 5** Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* terthiophene for the different sets of vibrational coordinates treated by VCI (black) with a HWHM of 0.02 eV compared to a corrected experimental spectrum of terthiophene at 77 K in ethanol<sup>13</sup> (green). For the definition of the reduced spaces, see the main text. The vertical transition energy is given as a vertical grey line. The experimental spectrum is shifted by 0.000 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eqn (5) is quantified below the respective graphs.

the respective vibrational coordinates  $\left(\frac{A_i^a}{A_i}\right)$ . If the squared ratio for a mode is equal to 1, we expect the IMDHO treatment to be a good approximation and for large deviations, the respective mode is more likely to exhibit anharmonic contributions. For the thiophene test cases, all modes with significant (moderate) effect of one-mode anharmonicity on the one-mode vibrational profiles have an anharmonicity measure of  $\left(\frac{A_i^a}{A_i}\right)^2 < 0.1$ . For pentathiophene, we observe, however, two modes with very small anharmonicity measures below that value that do not show significant one-mode anharmonicities (see Table SII and Fig. S8 in the ESI†). Similar observations are also made for the organic dyes below (see Tables SIII, SIV, and SV as well as Fig. S9, S11, and S13 in the ESI†).

In summary, the correlation of  $\left(\frac{A_i^a}{A_i}\right)^2$  to the observed anharmonicities is less clear than hoped for. This observation might be related to that the Franck-Condon region has a larger influence on the vibrational profile than the region around the equilibrium of the final state, as has been reported earlier.<sup>66</sup> We, hence, speculate that the proposed anharmonicity measure focuses too much on the adiabatic region to give a clear guideline for the effect of the anharmonicity on the vibrational profile. For this reason, we further apply the computationally more expensive comparison of the effect of the one-mode anharmonicity on the one-mode vibronic spectra directly.

We note that among the modes investigated for anharmonicity, those with high harmonic displacements also exhibit the largest anharmonicity effects on the vibrational profiles. An exception is, however, obtained for bithiophene, for which the mode with the second largest displacement exhibits the largest one-mode anharmonicity of the tested modes. We should, however, keep in mind that we have a very limited set of modes here. Furthermore, known highly anharmonic modes like the torsion in oligothiophenes<sup>14</sup> only contribute little to the vibrational profile. The observed correlation between high one-mode



anharmonicity and large harmonic displacements can, hence, not be considered robust to predict the impact of anharmonicity of a given mode.

#### 4.2 Organic biomarkers

In the previous section, we have seen that reduced-space VCI vibrational profiles can be combined with IMDHO-type line shapes to the VCI-in-IMDHO model. This combination can lead to a faster convergence of the vibrational profile. This accompanying reduction in computational cost allows us to apply this methodology also to the somewhat larger organic fluorescent dyes introduced in Fig. 1. In interest of computational cost, we only included two-mode couplings in the PESs for modes, for which  $|A_i| > 0.6$  for these larger examples. For all organic biomarkers, we compare also to experimental spectra. In contrast to the experimental spectra for oligothiophenes above, the broadening in the experimental spectra for the organic biomarkers can be assumed to be dominated by thermal broadening. Fewer details are resolved in these spectra.

**4.2.1 HS84.** For the oligothiophene-based fluorescent dye HS84, we obtained five modes with  $|A_i| > 0.2$  and  $\sigma_i > 322.6 \text{ cm}^{-1}$  (0.04 eV, see Table 3). We note that additional to the listed modes, there are 17 modes, for which  $|A_i| > 0.2$ . All these modes are discarded due to the resolution measure.

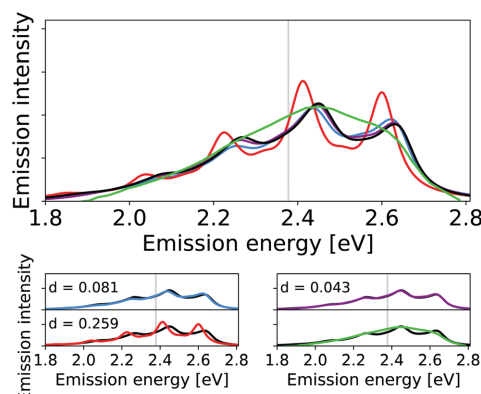
The corresponding calculated vibrational line shapes of the emission spectra for HS84 obtained for reduced-space VCI, IMDHO and VCI-in-IMDHO models are shown in Fig. 6 for a HWHM of 0.04 eV.

The energy differences between the main peaks are similar in all approaches (between 0.180 eV and 0.187 eV). With the reduced-space VCI model, small additional side peaks are obtained. The reduced-space VCI results differ more from the VCI-in-IMDHO spectrum than the IMDHO spectrum. This somewhat better “resolution” of the reduced-space VCI spectrum compared to the VCI-in-IMDHO spectrum, as well as the largest deviation for the reduced-space VCI model, can most likely be attributed to the neglect of the broadening effect of the remaining vibrational degrees of freedom in the reduced-space model.

The differences between the VCI-in-IMDHO and pure IMDHO spectra mainly concern the intensity rather than the position of the peaks. For none of the selected modes, we observe a strong effect of one-mode anharmonicity (see Fig. S9 in the ESI†). Further, we obtain a slight deviation for the

**Table 3** Mode ranking for HS84 with CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|A_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$  and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH)

Mode	$\nu_i/\text{cm}^{-1}$	$ A_i $	$\sigma_i/\text{cm}^{-1}$	1-Mode AH
133	1576	1.120	2044	Weak
131	1547	0.699	879	Weak
68	735	0.697	416	Weak
107	1264	0.420	393	Weak
126	1488	0.313	338	Weak



**Fig. 6** Calculated harmonic (blue), reduced-space VCI (red), VCI-in-IMDHO without mode–mode couplings in the PES (magenta) and VCI-in-IMDHO (black) broadened emission spectra of HS84 molecule with a HWHM of 0.04 eV compared to a corrected experimental spectrum in PBS<sup>8</sup> (green). Vertical electronic emission energy with grey lines. The VCI spectrum is obtained by conventional Lorentzian broadening. In the bottom part, the calculated spectra are directly compared to the best approximation (VCI-in-IMDHO) and the experimental spectrum, respectively. Additionally, the deviation between the calculated spectra according to eqn (5) is quantified in the respective graphs. The modes treated by VCI in the reduced-space VCI are 131, 133, 68, 107, 126, out of which only the modes 133, 131, and 68 were coupled in the underlying PESs. The experimental spectra are shifted by 0.144 eV to get the same position of maximum as VCI-in-IMDHO spectra.

VCI-in-IMDHO spectrum with all the modes from Table 3 accounted for anharmonically including two-mode couplings for selected modes in the PES compared to that with only one-mode anharmonicities in the reduced-space PES (see Fig. 6). This suggests that mode–mode coupling effects, similar to one-mode anharmonicities, play a small role here.

Fig. 6 additionally shows an experimental reference for the emission spectra in solution. This experimental spectrum is however very broad so that the vibrational progression is not resolved. For the HS84, we further observe a pronounced sensitivity of the pure IMDHO spectrum on the inclusion of solvent effects by a continuum solvation model (see Fig. S10 in the ESI†). Hence, solvent effects are expected to be not negligible in this case.

**4.2.2 1,4-Diphenylbutadiene.** Next, we investigate the performance of the VCI-in-IMDHO scheme for 1,4-diphenylbutadiene (DPB). For this molecule, an earlier theoretical study concluded that anharmonic ground state effects are significant.<sup>65,67</sup> Table 4 shows the obtained ranking parameter for the highest ranked modes.

In total, we identify seven modes with a harmonic displacement larger than 0.2 and a  $\sigma_i$  larger than  $200 \text{ cm}^{-1}$  (0.025 eV), out of which two modes have harmonic displacements larger than 0.6 and are therefore coupled in the PES. Again, the mode with the largest harmonic displacement (mode 70) also has the largest  $\sigma_i$  and strongest one-mode anharmonicity among the



**Table 4** Mode ranking for 1,4-diphenylbutadiene with CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|d_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$  and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH)

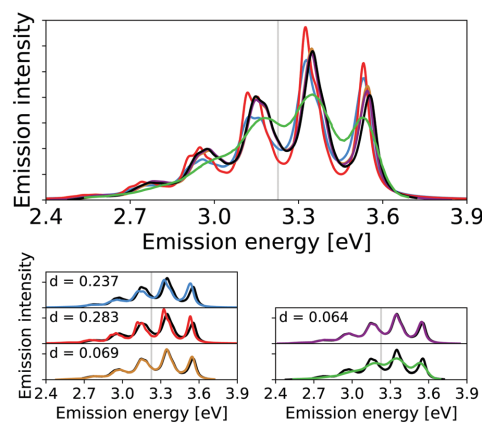
Mode	$\nu_i/\text{cm}^{-1}$	$ d_i $	$\sigma_i/\text{cm}^{-1}$	1-Mode AH
70	1691	1.270	3452	Moderate
53	1285	0.849	964	Weak
69	1653	0.527	663	Weak
41	1016	0.372	277	Weak
60	1435	0.303	314	Weak
57	1363	0.290	286	Weak
64	1544	0.202	223	Weak

ranked modes. All other ranked modes show only weak one-mode anharmonicity.

The vibrational line shapes of the emission spectra of 1,4-diphenylbutadiene for harmonic, reduced-space VCI, and VCI-in-IMDHO treatment of different reduced spaces are shown in Fig. 7 along with an experimental spectrum.<sup>65,67</sup> Again, the treatment of reduced-space anharmonicity has a significant impact. This holds in particular for the intensities, but also the energy differences are shifted. This shift in intensities mostly concerns an increase of the peak around 3.1 eV and a decreased intensity, of the peak around 3.5 eV when considering reduced-space anharmonicity. This shift leads to a better qualitative agreement with the experimental spectrum, similar to the conclusions drawn on the effect of full-space ground-state anharmonicity in ref. 65. We also note that accounting only for the anharmonicity of mode 70, leads already to a rather similar result as the VCI-in-IMDHO result with the seven highest ranked modes in this case (see Fig. 7). This means that the impact of one-mode anharmonicity of mode 70 on the vibrational profile is larger than the mode–mode coupling contributions in this case. Further, we find this effect to be more significant than the impact of the dichloromethane solution treated by continuum solvation model on the IMDHO spectrum (see Fig. S12 in the ESI†).

**4.2.3 Anthracene diimide.** As the last test system, we have chosen the anthracene diimide (ADI) depicted in Fig. 1. We note that the experimental spectrum we compare to is that of an anthracene diimide derivative with a longer linear alkyl chain in the imide substituents ( $-\text{C}_7\text{H}_{13}$ ),<sup>68</sup> which is in line with ref. 7. This can be rationalized by the observation that imide substituents in this compound series have no significant photophysical characteristics *e.g.* spectral band shape influence, which is also confirmed for other perylene diimides.<sup>69–71</sup>

The vibrational profile of this molecule has previously been assessed by means of harmonic calculations accounting for Duschinsky rotations and the different harmonic frequencies in the two involved electronic states as well as including ground-state anharmonicity effects.<sup>7</sup> The authors of this study found significant deviations from the experimental reference with all applied approaches as well a strong dependence on ground state anharmonicity, which, however, was highly dependent on the applied density functional.



**Fig. 7** Calculated harmonic (blue), reduced-space VCI (red), one-mode VCI-in-IMDHO (orange), VCI-in-IMDHO without mode–mode couplings (magenta) and VCI-in-IMDHO (black) broadened emission spectra of 1,4-diphenylbutadiene with a HWHM of 0.025 eV compared to a corrected experimental spectrum in hexane<sup>65,67</sup> (green). Vertical electronic emission energy with grey lines. The VCI spectrum is obtained by conventional Lorentzian broadening. In the bottom part, the calculated spectra are directly compared to the best approximation (VCI-in-IMDHO) and the experimental spectrum, respectively. Additionally, the deviation between the calculated spectra according to eqn (5) is quantified in the respective graphs. The modes treated by VCI in the reduced-space VCI are 70, 53, 69, 41, 60, 57, and 64, out of which only the modes 70 and 53 were coupled in the underlying PESs. Only mode 70 is VCI treated in the one-mode VCI-in-IMDHO treatment. The experimental spectra are shifted by 1.549 eV to get the same position of maximum as VCI-in-IMDHO spectra.

The mode ranking (see Table 5) leads to nine modes with a harmonic displacement larger than 0.2 and a resolution measure larger than  $200 \text{ cm}^{-1}$  out of which three have a harmonic displacement larger than 0.6. All of these modes exhibit only weak one-mode anharmonicity.

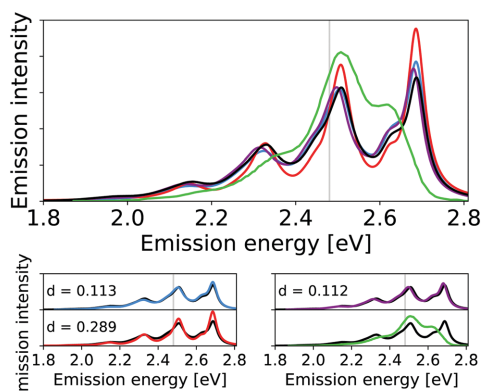
We find a moderate effect of the anharmonicity on the emission spectrum of the ADI (see Fig. 8). Again, this effect is more pronounced for the intensities than for the energy differences. For the two highest-energy peaks (at about 2.5 eV and about 2.7 eV, respectively) we find for the IMDHO model and pure anharmonic model that the peak at 2.7 eV has the highest intensity. This difference is reduced in the VCI-in-IMDHO treatment. In this case, we obtain a negligible effect of the one-mode anharmonicities (*cf.*  $d$  values for the pure IMDHO model and the VCI-in-IMDHO model without mode–mode couplings in the PES from the VCI-in-IMDHO result with mode–mode couplings), but a larger effect of the mode–mode couplings compared to the above examples. Overall, the effect of anharmonicities on the vibrational line shape is similar to that of the continuum solvent (see Fig. S14 in the ESI†). Still, all simulated spectra exhibit large deviations from the experimental reference.

The here calculated vibrational line shapes for the ADI applying CAM-B3LYP/def2-TZVP remind of those obtained both harmonically and with ground-state anharmonicity with



**Table 5** Mode ranking for the anthracene diimide with CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|A_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$  and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH)

Mode	$\nu_i/\text{cm}^{-1}$	$ A_i $	$\sigma_i/\text{cm}^{-1}$	1-Mode AH
79	1447	0.706	833	Weak
75	1398	0.642	712	Weak
92	1652	0.607	785	Weak
29	512	0.566	223	Weak
65	1215	0.352	312	Weak
81	1473	0.332	356	Weak
88	1546	0.318	367	Weak
89	1574	0.252	285	Weak
82	1502	0.227	244	Weak



**Fig. 8** Calculated harmonic (blue), reduced-space VCI (red), VCI-in-IMDHO without mode–mode couplings in the PES (magenta) and VCI-in-IMDHO broadened emission spectra of the anthracene diimide with a HWHM of 0.025 eV compared to a corrected experimental spectrum in dichloromethane<sup>7,68</sup> (green). Vertical electronic emission energy with grey lines. The VCI spectrum is obtained by conventional Lorentzian broadening. In the bottom part, the calculated spectra are directly compared to the best approximation (VCI-in-IMDHO) and the experimental spectrum, respectively. Additionally, the deviation between the calculated spectra according to eqn (5) is quantified in the respective graphs. The modes treated by VCI in the reduced-space VCI are 79, 75, 92, 29, 65, 81, 88, 89, 82 out of which only the modes 79, 75, and 92 were coupled in the underlying PESs. The experimental spectra are shifted by 1.124 eV to get the same position of maximum as VCI-in-IMDHO spectra.

M06-2X in ref. 7. In view of the pronounced functional dependence on the vibrational line shapes for the anthracene diimide, it can be speculated that the deviations obtained from experiments may be more likely dominated by a challenging electronic structure of the involved electronic states rather than the applied vibrational treatment.

## 5 Summary and conclusions

With the VCI-in-IMDHO model, we introduce a hybrid treatment for vibrational line shapes of optical spectra. This model

allows us to treat selected modes by highly accurate and anharmonic vibrational wave function methods (VCI) and the remaining modes by the very approximate IMDHO model. No mode–mode coupling is considered within the modes treated by the IMDHO model and in between the two sets of modes. The Franck–Condon factors for the reduced vibrational space obtained by VCI are broadened by a system-specific line shape obtained from the IMDHO autocorrelation function for a given half-width at half maximum considering all modes but the ones treated by VCI.

For the separation of the modes into these two sets, we refined and extended the criteria introduced in ref. 16. One criterion is the impact on the vibrational progression estimated by the displacement of the minima in the two involved electronic states within the IMDHO model ( $|A_i|$ ), which has also been applied in earlier work.<sup>16,47</sup> We refined the resolution criterion  $\sigma_i$ , which excludes modes due to the expected poor resolution of the vibrational modes rather than a simple frequency criterion.<sup>16,47</sup> We show for oligothiophene examples that the harmonic treatment of the modes with a large  $A_i$  but a  $\sigma_i$  below the chosen half-width at half maximum leads to negligible deviations in the resulting VCI-in-IMDHO spectra compared to the case, where these modes are included in the VCI treatment. We also attempted to assess the impact of the anharmonicity *a priori* by comparing the displacement within the IMDHO model  $|A_i|$  to the displacement between the optimized minima of the two electronic states projected on the respective mode  $|A_i^g|$ .

We, however, could not find a clear correlation of  $\left(\frac{A_i^g}{A_i}\right)^2$  with the impact of anharmonicity on the one-mode vibrational profiles and therefore resorted to the comparison of calculated harmonic and anharmonic one-mode vibrational profiles: if the one-mode vibrational spectra are affected by anharmonicity, the multi-mode anharmonic spectrum will also be. Unfortunately, this cannot be turned around. That means, weak one-mode anharmonicity is no guarantee for negligible overall anharmonicity, as we saw, for the anthracene diimide.

For the three investigated dyes, *i.e.*, the oligothiophene-based dye HS84, 1,4-diphenylbutadiene, and an anthracene diimide, we found between five and nine modes above the thresholds to be included in the VCI treatment. For two to three modes of these, we also accounted for mode–mode couplings in the underlying potential energy surfaces. For all examples, we find some impact on anharmonicity, which is larger for the intensity of the peaks than on the peak positions. This means Franck–Condon factors are more sensitive toward anharmonicity than the vibrational energies. The cases, however, differ in the impact of one-mode anharmonicity and/or mode–mode coupling: for HS84 both anharmonic effects are small but visible. The anharmonic effect in the vibrational line shape of 1,4-diphenylbutadiene is dominated by one-mode anharmonicity, with small contribution from the mode–mode coupling. And for the anthracene diimide, we find negligible effect of one-mode anharmonicity, but a somewhat larger effect of the mode–mode couplings. In case of 1,4-diphenylbutadiene, the VCI-in-IMDHO model leads to an improved agreement with



experiment compared to the pure IMDHO results. For HS84, the agreement is unclear due to a not very well resolved experimental reference, which is furthermore affected by the solvent. For the anthracene diimide, the experimental spectrum could not be recovered with any of the models. The strong functional dependence of the harmonic vibrational line shape<sup>7</sup> suggests that this failure might be due to deficiencies in the applied electronic-energy methods for this molecule.

With the presented VCI-in-IMDHO model, we hence reduce the computational cost compared to the reduced-space VCI approach.<sup>16</sup> It allows computations of emission line shapes of organic dyes accounting for anharmonicity in both involved electronic states, where required. We systematically converge the vibrational profile with increasing vibrational space described by accurate VCI methods. For this, we assess both the overall impact of the mode on the vibrational progression as well as the resolution of the resulting vibronic peaks in a spectrum with a given line width. What remains a challenge is to estimate the anharmonic effect and particularly the impact of mode–mode couplings *a priori*. Still, the hybrid nature of the VCI-in-IMDHO model combined with possibility to systematically increase the vibrational space treated by VCI provides reliable, yet affordable computational setup for vibrational profiles for typical fluorescent dyes. For the organic dyes, we additionally calculated the IMDHO vibrational profiles in continuum solvation models. In particular for the HS84, this effect was significant and larger than that of the anharmonicity. For the other molecules, the impact of anharmonicity and solvation effects are more similar. This suggests that both effects have to be included for accurate computations. Furthermore, for the calculation of vibronic profiles in heterogeneous environments, such as HS84 at an amyloid fibril, more multi-level aspects, *e.g.*, on the electronic-structure calculations and dynamics of the environment are required. The combination of accurate vibrational profiles with different environmental description will be considered in future work.

## Author contributions

NNTM: data curation, software, formal analysis, validation, visualization, writing. CK: conceptualization, methodology, writing.

## Conflicts of interest

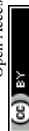
There are no conflicts to declare.

## Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) through the Emmy Noether Young Group Leader Programme (project KO 5423/1-1). We are further grateful to Diana Madsen for the potential energy surfaces for the oligothiophenes.

## Notes and references

- J. Neugebauer, E. Jan Baerends, M. Nooijen and J. Autschbach, *J. Chem. Phys.*, 2005, **122**, 234305.
- Y. Nakai, T. Mori and Y. Inoue, *J. Phys. Chem. A*, 2013, **117**, 83–93.
- D. Padula, J. Cerezo, G. Pescitelli and F. Santoro, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32349–32360.
- S. Höfener, B. A. Günther, M. E. Harding and L. H. Gade, *J. Phys. Chem. A*, 2019, **123**, 3160–3169.
- D. Jacquemin, E. Brémond, I. Ciofini and C. Adamo, *J. Phys. Chem. Lett.*, 2012, **3**, 468–471.
- S. Di Tommaso, D. Bousquet, D. Moulin, F. Baltenneck, P. Riva, H. David, A. Fadli, J. Gomar, I. Ciofini and C. Adamo, *J. Comput. Chem.*, 2017, **38**, 998–1004.
- Y. Houari, A. D. Laurent and D. Jacquemin, *J. Phys. Chem. C*, 2013, **117**, 21682–21691.
- H. Shirani, M. Linares, C. J. Sigurdson, M. Lindgren, P. Norman and K. P. R. Nilsson, *Chem. – Eur. J.*, 2015, **21**, 15133–15137.
- J. L. Price and J. C. Morris, *Ann. Neurol.*, 1999, **45**, 358–368.
- A. Åslund, C. J. Sigurdson, T. Klingstedt, S. Grathwohl, T. Bolmont, D. L. Dickstein, E. Glimsdal, S. Prokop, M. Lindgren and P. Konradsson, *et al.*, *ACS Chem. Biol.*, 2009, **4**, 673–684.
- I. Berg, K. P. R. Nilsson, S. Thor and P. Hammarström, *Nat. Protoc.*, 2010, **5**, 935–944.
- V. Mahajan, T. Klingstedt, R. Simon, K. P. R. Nilsson, A. Thueringer, K. Kashofer, J. Haybaeck, H. Denk, P. M. Abuja and K. Zatloukal, *Gastroenterology*, 2011, **141**, 1080–1090.
- R. S. Becker, J. Seixas De Melo, A. L. Maçanita and F. Elisei, *J. Phys. Chem.*, 1996, **100**, 18683–18695.
- R. Improta, F. J. Ferrer, E. Stendardo and F. Santoro, *ChemPhysChem*, 2014, **15**, 3320–3333.
- M. Wehrle, M. Šulc and J. Vaníček, *J. Chem. Phys.*, 2014, **140**, 244114.
- D. Madsen, O. Christiansen, P. Norman and C. König, *Phys. Chem. Chem. Phys.*, 2019, **21**, 17410–17422.
- C. Gustafsson, M. Linares and P. Norman, *J. Phys. Chem. A*, 2020, **124**, 875–888.
- J. M. Bowman, K. Christoffel and F. Tobin, *J. Phys. Chem.*, 1979, **83**, 905–912.
- O. Christiansen, *J. Chem. Phys.*, 2004, **120**, 2149–2159.
- O. Christiansen, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2942–2953.
- O. Christiansen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6672–6687.
- P. M. Champion and A. C. Albrecht, *J. Chem. Phys.*, 1980, **72**, 6498–6506.
- A. B. Myers, R. A. Mathies, D. J. Tannor and E. J. Heller, *J. Chem. Phys.*, 1982, **77**, 3857–3866.
- J. Cerezo, F. J. Avila Ferrer, G. Prampolini and F. Santoro, *J. Chem. Theory Comput.*, 2015, **11**, 5810–5825.
- R. Zalesny, N. A. Murugan, F. Gel'mukhanov, Z. Rinkevicius, B. Osmialowski, W. Bartkowiak and H. Ågren, *J. Phys. Chem. A*, 2015, **119**, 5145–5152.



- 26 J. Cerezo, D. Aranda, F. J. Avila Ferrer, G. Prampolini and F. Santoro, *J. Chem. Theory Comput.*, 2020, **16**, 1215–1231.
- 27 A. Segalina, J. Cerezo, G. Prampolini, F. Santoro and M. Pastore, *J. Chem. Theory Comput.*, 2020, **16**, 7061–7077.
- 28 T. J. Zuehlsdorff, S. V. Shedge, S. Y. Lu, H. Hong, V. P. Aguirre, L. Shi and C. M. Isborn, *Annu. Rev. Phys. Chem.*, 2021, **72**, 165–188.
- 29 P. P. Fehér, Á. Madarász and A. Stirling, *J. Chem. Theory Comput.*, 2021, **17**, 6340–6352.
- 30 R. A. Simon, H. Shirani, K. O. A. Åslund, M. Bäck, V. Haroutunian, S. Gandy and K. P. R. Nilsson, *Chem. – Eur. J.*, 2014, **20**, 12537–12543.
- 31 S. E. Wallace-Williams, S. Moeller, R. A. Goldbeck, K. M. Hanson, J. W. Lewis, W. A. Lee and D. S. Kliger, *J. Phys. Chem.*, 1993, **97**, 9587–9592.
- 32 Y.-P. Sun, C. E. Bunker, P. L. Wickremesinghe, H. W. Rollins and G. E. Lawson, *J. Phys. Chem.*, 1995, **99**, 3423–3429.
- 33 H. Langhals, *Heterocycles*, 1995, **40**, 477–500.
- 34 X. Zhan, A. Facchetti, S. Barlow, T. J. Marks, M. A. Ratner, M. R. Wasielewski and S. R. Marder, *Adv. Mater.*, 2011, **23**, 268–284.
- 35 C. Li and H. Wonneberger, *Adv. Mater.*, 2012, **24**, 613–636.
- 36 J. Li, F. Dierschke, J. Wu, A. C. Grimsdale and K. Müllen, *J. Mater. Chem.*, 2006, **16**, 96–100.
- 37 H. Cao, D. I. Diaz, N. DiCesare, J. R. Lakowicz and M. D. Heagy, *Org. Lett.*, 2002, **4**, 1503–1505.
- 38 Y. Che, X. Yang, S. Loser and L. Zang, *Nano Lett.*, 2008, **8**, 2219–2223.
- 39 L. Zhu, W. Wu, M.-Q. Zhu, J. J. Han, J. K. Hurst and A. D. Li, *J. Am. Chem. Soc.*, 2007, **129**, 3524–3526.
- 40 F. Duschinsky, *Acta Physicochim. U.R.S.S.*, 1937, **7**, 551–566.
- 41 A. B. Myers and R. A. Mathies, *Biol. Appl. Raman Spectrosc.*, 1987, **2**, 1–58.
- 42 E. U. Condon, *Phys. Rev.*, 1928, **32**, 858–872.
- 43 T. Sharp and H. Rosenstock, *J. Chem. Phys.*, 1964, **41**, 3453–3463.
- 44 D. J. Tannor and E. J. Heller, *J. Chem. Phys.*, 1982, **77**, 202–218.
- 45 F. Neese, T. Petrenko, D. Ganyushin and G. Olbrich, *Coord. Chem. Rev.*, 2007, **251**, 288–327.
- 46 T. Petrenko and F. Neese, *J. Chem. Phys.*, 2012, **137**, 234107.
- 47 D. Barton, C. König and J. Neugebauer, *J. Chem. Phys.*, 2014, **141**, 164115.
- 48 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision C.01*, Gaussian Inc., Wallingford CT, 2016.
- 49 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 50 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 51 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
- 52 E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032–3041.
- 53 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.
- 54 O. Christiansen, I. H. Godtlibsen, E. M. Gras, W. Györfly, M. B. Hansen, M. B. Hansen, E. L. Klinting, J. Kongsted, C. König, S. A. Losilla, D. Madsen, N. K. Madsen, P. Seidler, K. Sneskov, M. Sparta, B. Thomsen, D. Toffoli and A. Zocante, *Molecular Interactions Dynamics And Simulation C++ (MidasCPP) package*.
- 55 D. Madsen, PhD thesis, Aarhus University, 2019.
- 56 T. Moitra, D. Madsen, O. Christiansen and S. Coriani, *J. Chem. Phys.*, 2020, **153**, 234111.
- 57 M. Sparta, D. Toffoli and O. Christiansen, *Theor. Chem. Acc.*, 2009, **123**, 413–429.
- 58 D. Toffoli, M. Sparta and O. Christiansen, *Mol. Phys.*, 2011, **109**, 673–685.
- 59 J. M. Bowman, X. Huang, L. B. Harding and S. Carter, *Mol. Phys.*, 2007, **104**, 33–45.
- 60 J. M. Luis, B. Kirtman and O. Christiansen, *J. Chem. Phys.*, 2006, **125**, 154114.
- 61 V. Rodriguez-Garcia, K. Yagi, K. Hirao, S. Iwata and S. Hirata, *J. Chem. Phys.*, 2006, **125**, 014109.
- 62 G. Rauhut, *J. Phys. Chem. A*, 2015, **119**, 10264–10271.
- 63 R. Piessens, E. de Doncker-Kapenga, C. W. Überhuber and D. K. Kahaner, *Quadpack: A subroutine package for automatic integration*, Springer-Verlag, Berlin Heidelberg, 1983.
- 64 F. J. Avila Ferrer, J. Cerezo, E. Stendardo, R. Impropa and F. Santoro, *J. Chem. Theory Comput.*, 2013, **9**, 2072–2082.
- 65 A. Charaf-Eddin, A. Planchat, B. Mennucci, C. Adamo and D. Jacquemin, *J. Chem. Theory Comput.*, 2013, **9**, 2749–2760.
- 66 W. Domcke, L. Cederbaum, H. Köppel and W. von Niessen, *Mol. Phys.*, 1977, **34**, 1759–1770.
- 67 H. Du, R.-C. A. Fuh, J. Li, L. A. Corkan and J. S. Lindsey, *Photochem. Photobiol.*, 1998, **68**, 141–142.
- 68 A. R. Mohebbi, C. Munoz and F. Wudl, *Org. Lett.*, 2011, **13**, 2560–2563.
- 69 L. Chen, C. Li and K. Müllen, *J. Mater. Chem. C*, 2014, **2**, 1938–1956.
- 70 B. Zhang, H. Soleimanejad, D. J. Jones, J. M. White, K. P. Ghiggino, T. A. Smith and W. W. Wong, *Chem. Mater.*, 2017, **29**, 8395–8403.
- 71 R. Matthews, J. Swisher, K. M. Hutchins and E. B. Pentzer, *Chem. Mater.*, 2018, **30**, 3571–3577.

Electronic Supplementary Information to  
Tailored anharmonic-harmonic vibrational profiles for fluorescent  
biomarkers

Nghia Nguyen Thi Minh and Carolin König

**S1** *Trans* bithiophene

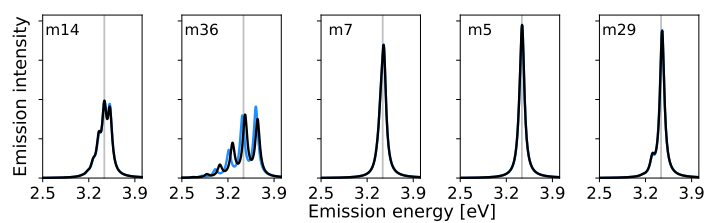


Figure S1: IMDHO-treated (blue) and VCI-treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of *trans* bithiophene employing PESs from ref. [1]. The vertical transition energy is given as a vertical grey line.

## S2 All-*trans* terthiophene

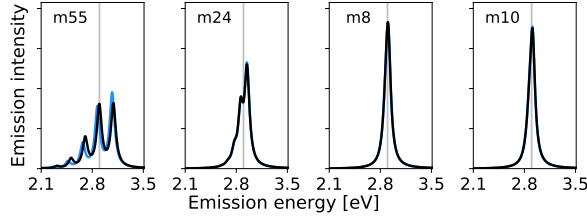


Figure S2: IMDHO-treated (blue) and VCI-treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of all-*trans* terthiophene employing PESs from ref. [1]. The vertical transition energy is given as a vertical grey line.

## S3 All-*trans* quaterthiophene

Table SI: Mode ranking for all-*trans* quaterthiophene using CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$  and anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode (i)	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
74	1616	1.280	0.280	0.048	3440	moderate
9	162	1.02	0.637	0.390	168	weak
32	705	0.894	1.034	1.337	575	weak
15	333	0.693	1.439	4.315	187	weak

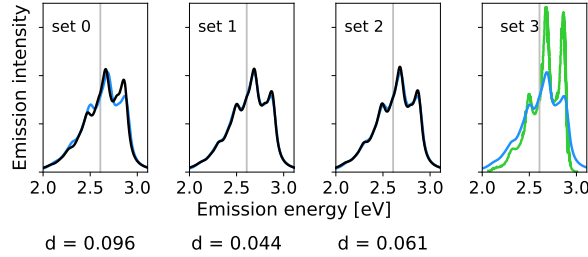


Figure S3: Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* quaterthiophene for the different sets of vibrational coordinates treated by VCI (black) and that for the largest set of modes set 3 (blue) with a HWHM of  $\gamma = 0.04$  eV compared to a corrected experimental spectrum of quaterthiophene at 77 K in ethanol [2] (green). The vertical transition energy is given as a vertical grey line. Within the anharmonic and hybrid models, (set 0) including 0 anharmonic mode; (set 1) including the one with most pronounced one mode anharmonicity: 74; (set 2) including two highest ranked modes: 74, 32 and (set 3) including 4 modes: 74, 9, 32, 15. The experimental spectra is shifted by 0.086 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eq. (5) in the main text is quantified below the respective graphs.



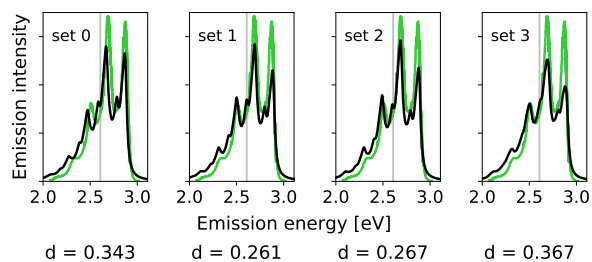


Figure S4: Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* quaterthiophene for the different sets of vibrational coordinates treated by VCI (black) with a HWHM of  $\gamma = 0.02$  eV compared to a corrected experimental spectrum of quaterthiophene at 77 K in ethanol [2] (green). The vertical transition energy is given as a vertical grey line. Within the anharmonic and hybrid models, (set 0) including 0 anharmonic modes; (set 1) including the one with most pronounced one mode anharmonicity: 74; (set 2) including two highest ranked modes: 74, 32 and (set 3) including 4 modes: 74, 9, 32, 15. The experimental spectra is shifted by 0.010 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eq. (5) in the main text is quantified below the respective graphs.

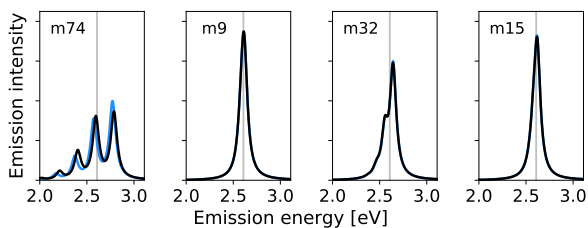


Figure S5: IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of all-*trans* quaterthiophene employing PESs obtained from [1]. The vertical transition energy is given as a vertical grey line.

## S4 All-*trans* pentathiophene

Table SII: Mode ranking for all-*trans* pentathiophene using CAM-B3LYP/6-31+g(d). The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$  and anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
92	1600	0.960	0.164	0.029	1479	weak
91	1592	0.835	0.191	0.052	1164	weak
41	711	0.795	0.989	1.549	483	weak
18	323	0.690	1.429	4.293	180	weak

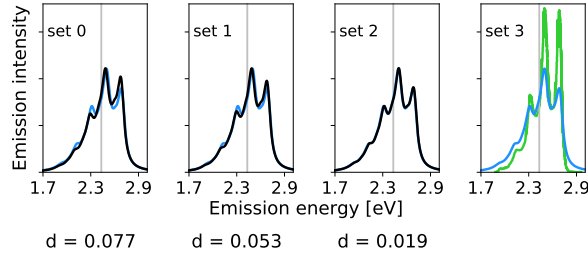


Figure S6: Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* pentathiophene for the different sets of vibrational coordinates treated by VCI (black) and that for the largest set of modes set 3 (blue) with a HWHM of  $\gamma = 0.04$  eV compared to a corrected experimental spectrum of pentathiophene at 77 K in ethanol [2] (green). The vertical transition energy is given as a vertical grey line. Within the anharmonic and hybrid models, (set 0) including 0 anharmonic mode, (set 1) including the highest ranked mode: 92; (set 2) including two anharmonic modes 92, 91 and (set 3) including all important modes: 92, 91, 41, 18. The experiment spectrum is shifted by 0.101 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eq. (5) in the main text is quantified below the respective graphs.

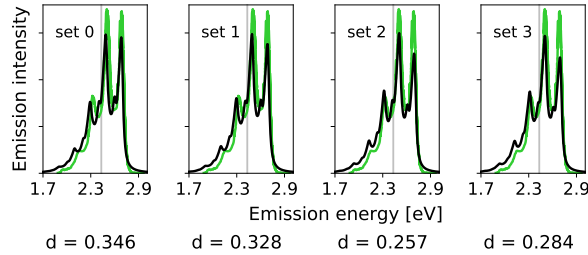


Figure S7: Hybrid VCI-in-IMDHO vibrational profiles for the  $S_0 \leftarrow S_1$  emission for all-*trans* pentathiophene for the different sets of vibrational coordinates treated by VCI (black) with a HWHM of  $\gamma = 0.02$  eV compared to a corrected experimental spectrum of pentathiophene at 77 K in ethanol [2] (green). The vertical transition energy is given as a vertical grey line. Within the anharmonic and hybrid models, (set 0) including 0 anharmonic mode, (set 1) including the highest ranked mode: 92; (set 2) including two anharmonic modes 92, 91 and (set 3) including all important modes: 92, 91, 41, 18. The experiment spectrum is shifted 0.005 eV to get the same position of maximum as calculated spectrum. Additionally, the deviation between the shown spectra according to eq. (5) in the main text is quantified below the respective graphs.

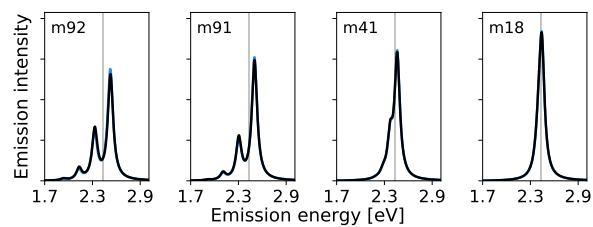


Figure S8: IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of all-*trans* pentathiophene using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S5 HS84

Table SIII: Mode ranking for HS84 using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$  and anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
133	1576	1.120	0.225	0.040	2044	weak
131	1547	0.699	0.529	0.572	879	weak
68	735	0.697	1.584	5.166	416	weak
107	1264	0.420	0.063	0.023	393	weak
126	1488	0.313	0.082	0.068	338	weak

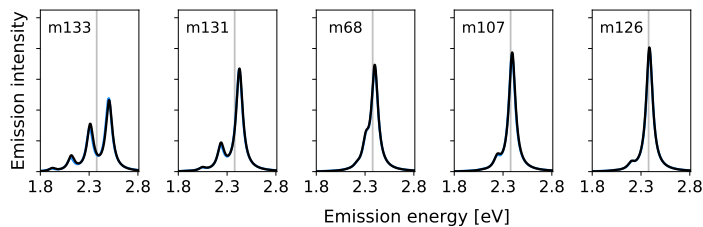


Figure S9: IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.04$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of HS84 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

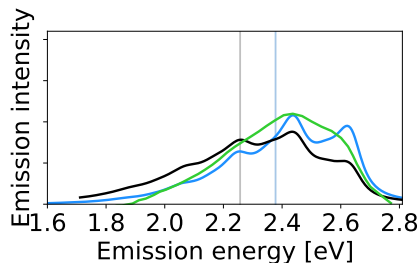


Figure S10: Calculated IMDHO harmonic emission spectra of HS84 in vacuum (blue), in water estimated with the PCM model (black) and a HWHM of  $\gamma = 0.04$  eV using CAM-B3LYP/def2-TZVP compared to a corrected experimental spectrum in PBS[3] (green). The vertical transition energy in vacuum is given as a vertical grey line. The experimental and the IMDHO spectrum in solvent as well as its corresponding excitation energy (vertical blue light line) are shifted by 0.130 eV and 0.211 eV, respectively, to get the same position of maximum as IMDHO harmonic spectra in vacuum. Additionally, the deviation between the calculated spectrum in vacuum and in solvent according to eq. (5) in the main text is 0.211 and 0.343, respectively.

## S6 1,4-diphenylbutadiene

Table SIV: Mode ranking for 1,4-diphenylbutadiene using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$  and anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
70	1691	1.270	0.287	0.051	3452	moderate
53	1285	0.849	0.237	0.078	964	weak
69	1653	0.527	0.117	0.049	663	weak
41	1016	0.372	0.141	0.144	277	weak
60	1435	0.303	0.084	0.077	314	weak
57	1363	0.290	0.091	0.099	286	weak
64	1544	0.202	0.053	0.069	223	weak

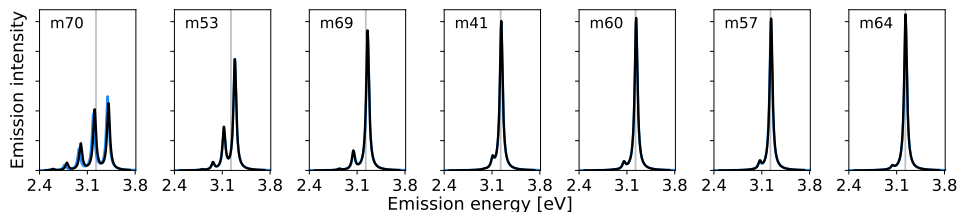


Figure S11: IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of 1,4-diphenylbutadiene using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

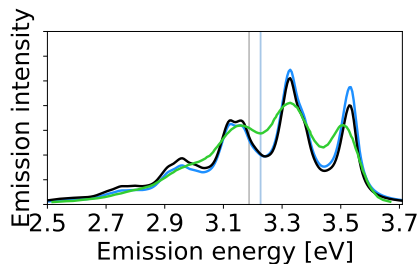


Figure S12: Calculated IMDHO harmonic emission spectra of 1,4-diphenylbutadiene in vacuum (blue), in hexane estimated with the PCM model (black) and a HWHM of  $\gamma = 0.025$  eV using CAM-B3LYP/def2-TZVP compared to a corrected experimental spectrum in hexane[4, 5] (green). The vertical transition energy in vacuum is given as a vertical grey line. The experimental and the IMDHO spectrum in solvent as well as its corresponding excitation energy (vertical blue light line) are shifted by 1.527 eV and 0.196 eV, respectively, to get the same position of maximum as IMDHO harmonic spectra in vacuum. Additionally, the deviation between the calculated spectrum in vacuum and in solvent according to eq. (5) in the main text is 0.289 and 0.253, respectively.

## S7 Anthracene diimide

Table SV: Mode ranking for the anthracene diimide using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic  $|\Delta_i|$ , anharmonic  $|\Delta_i^a|$  dimensionless displacement, the relative difference between harmonic and anharmonic displacement  $\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode	$\nu_i/\text{cm}^{-1}$	$ \Delta_i $	$ \Delta_i^a $	$\left(\frac{\Delta_i^a}{\Delta_i}\right)^2$	$\sigma_i/\text{cm}^{-1}$	1-mode AH
79	1447	0.706	0.149	0.045	833	weak
75	1398	0.642	0.133	0.043	712	weak
92	1652	0.607	0.107	0.031	785	weak
29	512	0.566	0.336	0.350	223	weak
65	1215	0.352	0.077	0.048	312	weak
81	1473	0.332	0.061	0.034	356	weak
88	1546	0.318	0.059	0.035	367	weak
89	1574	0.252	0.045	0.032	285	weak
82	1502	0.227	0.039	0.030	244	weak

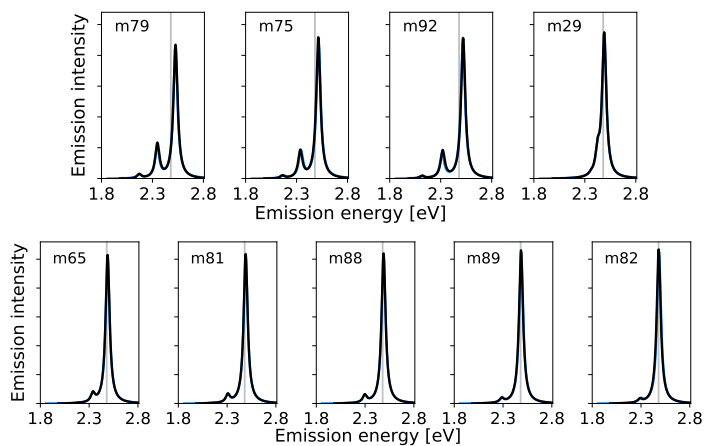


Figure S13: IMDHO treated (blue) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of anthracene diimide using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

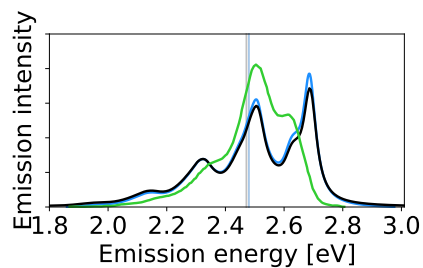


Figure S14: Calculated IMDHO harmonic emission spectra of anthracene diimide in vacuum (blue), in dichloromethane estimated with the PCM model (black) and a HWHM of  $\gamma = 0.025$  eV using CAM-B3LYP/def2-TZVP compared to a corrected experimental spectrum in dichloromethane[6, 7] (green). The vertical transition energy in vacuum is given as a vertical grey line. The experimental and the IMDHO spectrum in solvent as well as its corresponding excitation energy (vertical blue light line) are shifted by 1.121 eV and 0.169 eV, respectively, to get the same position of maximum as IMDHO harmonic spectra in vacuum. Additionally, the deviation between the calculated spectrum in vacuum and in solvent according to eq. (5) in the main text is 0.489 and 0.522, respectively.

## References

- [1] D. Madsen, O. Christiansen, P. Norman and C. König, Phys. Chem. Chem. Phys., 2019, **21**, 17410–17422.
- [2] R. S. Becker, J. Seixas De Melo, A. L. Maçanita and F. Elisei, J. Phys. Chem., 1996, **100**, 18683–18695.
- [3] H. Shirani, M. Linares, C. J. Sigurdson, M. Lindgren, P. Norman and K. P. R. Nilsson, Chem. - A Eur. J., 2015, **21**, 15133–15137.
- [4] A. Charaf-Eddin, A. Plachat, B. Mennucci, C. Adamo and D. Jacquemin, J. Chem. Theory Comput., 2013, **9**, 2749–2760.
- [5] H. Du, R.-C. A. Fuh, J. Li, L. A. Corkan and J. S. Lindsey, Photochem. Photobiol., 1998, **68**, 141–142.
- [6] Y. Houari, A. D. Laurent and D. Jacquemin, J. Phys. Chem. C, 2013, **117**, 21682–21691.
- [7] A. R. Mohebbi, C. Munoz and F. Wudl, Org. Lett., 2011, **13**, 2560–2563.



A.2 TAU PROTEIN BINDING MODES IN ALZHEIMER'S DISEASE FOR CATIONIC LUMINESCENT LIGANDS

The publication:

**Tau protein binding modes in Alzheimer's disease for cationic luminescent ligands**

Y. Todarwal, C. Gustafsson, T.M.N. Nguyen, I. Ertzgaard, T. Klingstedt, B. Ghetti, R. Vidal, C. König, M. Lindgren, K P. R Nilsson, M. Linares and P. Norman,

*J. Phys. Chem. B*, 2021, 125(42):11628–11636,

DOI: 10.1021/acs.jpcc.1c06019,

is included in this thesis under the terms of the Attribution 4.0 International (CC BY 4.0) license.

## Tau Protein Binding Modes in Alzheimer's Disease for Cationic Luminescent Ligands

Yogesh Todarwal, Camilla Gustafsson, Nghia Nguyen Thi Minh, Ingrid Ertzgaard, Thérèse Klingstedt, Bernardino Ghetti, Ruben Vidal, Carolin König, Mikael Lindgren, K. Peter R. Nilsson, Mathieu Linares, and Patrick Norman\*

Cite This: *J. Phys. Chem. B* 2021, 125, 11628–11636

Read Online

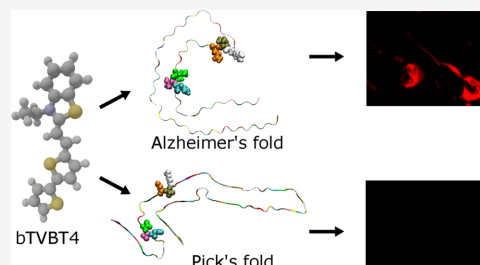
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** The bi-thiophene-vinylene-benzothiazole (bTVBT4) ligand developed for Alzheimer's disease (AD)-specific detection of amyloid tau has been studied by a combination of several theoretical methods and experimental spectroscopies. With reference to the cryo-EM tau structure of the tau protofilament (*Nature* 2017, 547, 185), a periodic model system of the fibril was created, and the interactions between this fibril and bTVBT4 were studied with nonbiased molecular dynamics simulations. Several binding sites and binding modes were identified and analyzed, and the results for the most prevailing fibril site and ligand modes are presented. A key validation of the simulation work is provided by the favorable comparison of the theoretical and experimental absorption spectra of bTVBT4 in solution and bound to the protein. It is conclusively shown that the ligand–protein binding occurs at the hydrophobic pocket defined by the residues Ile360, Thr361, and His362. This binding site is not accessible in the Pick's disease (PiD) fold, and fluorescence imaging of bTVBT4-stained brain tissue samples from patients diagnosed with AD and PiD provides strong support for the proposed tau binding site.



### INTRODUCTION

Several intrinsically disordered proteins are known to self-assemble into  $\beta$ -sheet filament structures (cross- $\beta$ ), or amyloid fibrils, associated with neurodegenerative diseases such as Alzheimer's (AD), Parkinson's, and Pick's (PiD) diseases. Our current knowledge of the biogenesis and aggregation steps of amyloids has recently been comprehensively summarized, including a review of experimental as well as computational work.<sup>1</sup> In AD, notably, the key proteins are amyloid  $\beta$  ( $A\beta$ ) with 40 or 42 residues and tau with 352 to 421 residues. The domain organization of tau is complex, and disease filaments show six tau isoforms and distinct morphologies, but paired helical filaments (PHFs) in neurofibrillary tangles are central to the development of AD. These filaments are composed of a rigid and structurally ordered core and a flexible and structurally disordered coat. The molecular structure of the PHF core with protofilaments comprising residues 306–378 in tau has been determined by means of cryogenic electron microscopy (cryo-EM), revealing two C-shaped protofilaments related by helical symmetry and stacked with a rise of 4.7 Å and a twist of *ca.* 1°.<sup>2</sup> Compared to AD, PiD is less common. Patients suffering from PiD have been found to have intraneuronal inclusions of hyperphosphorylated tau aggregates. However, while the symptoms of PiD may be similar to the symptoms of AD, the tau inclusion bodies associated with

PiD pathology are both biochemically and histologically distinct from the aggregates of tau identified in patients with AD pathology.<sup>3</sup> Tau in AD is also known to be highly phosphorylated. However, it has been shown that the phosphorylation sites predominantly reside outside the core region,<sup>4</sup> and our study is consequently concerned with the pristine structure from ref 2.

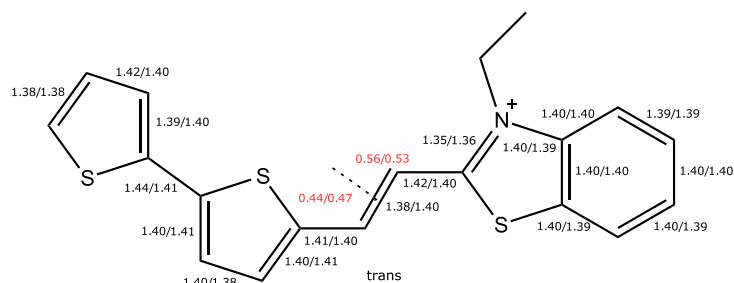
Several methods exist to study the molecular and functional aspects of tau physiopathology.<sup>5</sup> Noninvasive detection and imaging of  $A\beta$  and tau fibril deposits can be achieved by means of positron emission tomography<sup>6–15</sup> and fluorescence spectroscopy. For the latter, small hydrophobic and environment-sensitive ligands have been developed most commonly as derivatives of Thioflavin T<sup>16</sup> and Congo red<sup>17,17</sup> and therefore not suited for clinical studies due to their toxic character and inability to pass the blood–brain barrier (BBB). As an alternative, Nilsson and co-workers have proposed a class of ligands known as luminescent conjugated oligothiophenes

Received: July 6, 2021

Revised: September 30, 2021

Published: October 13, 2021





**Figure 1.** Molecular structure of the bTVBT4 ligand. Selected ground and excited state ( $S_0/S_1$ ) bond length (Å) parameters of the  $\pi$ -conjugated backbone are given. Natural population analysis (NPA) charges for the bi-thiophene and benzothiazole moieties (as separated by the dashed line) in the  $S_0$  and  $S_1$  states are given in red. The  $S_0$  and  $S_1$  states are described at the levels of DFT/B3LYP and TDDFT/CAM-B3LYP, respectively.

(LCOs) that show aggregate-specific strong fluorescent signals upon binding to a wide range of protein aggregate morphotypes.<sup>18</sup> Such LCOs have the ability to pass through the BBB<sup>19,20</sup> and facilitate early-stage detection of the buildup of misfolded protein aggregates,<sup>21–24</sup> and based on the bi-thiophene-vinylene (bTV) scaffold, ligands demonstrating tau-specific binding have been synthesized.<sup>25</sup> On the theoretical side, we have developed a methodology to provide a microscopic understanding of the ligand–protein interactions by means of unbiased molecular dynamics (MD) and subsequent spectroscopy simulations,<sup>26,27</sup> adopting results in aqueous solution as reference.<sup>28–30</sup> We have applied this protocol to study the binding of the anionic p-FTAA ligand to amyloid- $\beta$  and showed that the fingerprinting optical responses are associated with the planarity of the  $\pi$ -conjugated system.<sup>27</sup> The ligand subject of the present study, bTVBT4, belongs to the bTV category and features a cationic benzothiazole (BT) moiety, see Figure 1. In contrast to p-FTAA, it demonstrates tau-specific binding, and a prime objective of the present study is to see whether or not this protein-specificity is possible to rationalize within the realm of our simple fibrillar models of the disease aggregates. In the next section, we will give an overview of the adopted methodology with guiding references to the Supporting Information (SI), where a rich amount of underlying details can be found. Thereafter follows a section discussing the key results obtained from our theoretical and experimental studies of this ligand, and we provide also here references to SI for additional information with the intent to keep the main article to the point and focused on the main message.

## METHODS

The binding of bTVBT4 to the tau fibril was studied with molecular dynamics (MD) simulations, with general computational details found in SI section 1.1.4. A periodic model system representing the tau fibril was created by repeating units of tau oligomers originating from the cryo-EM structure (PDB ID: 5O3I), as described in SI sections 1.2.2 and 1.2.3. Force field parameters specific to the bTVBT4 ligand were developed with reference to quantum chemical calculations at the level of B3LYP/6-31+G(d,p), as described in SI section 1.1.2.

A detailed description of the MD simulations conducted to identify the ligand binding sites on the tau fibril is provided in SI section 1.2.4. These calculations were based on the full

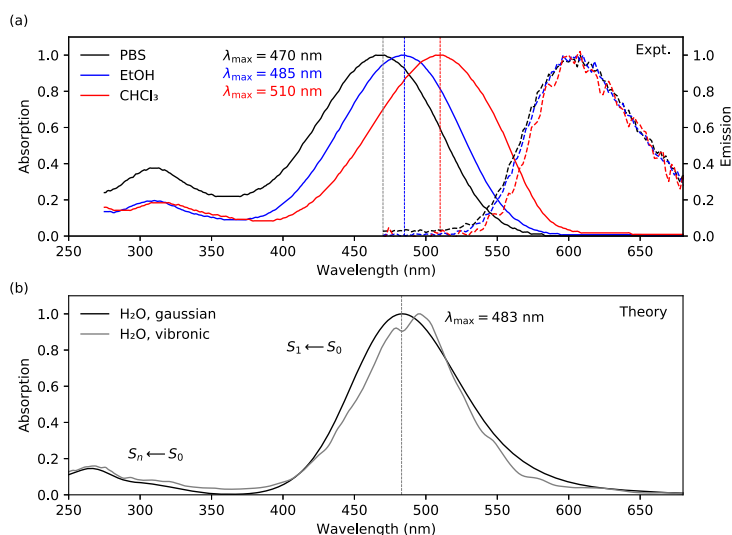
periodic model system for tau together with 60 ligands. To further analyze the binding modes of the main binding site, a reduced model system was created as described in SI section 1.2.5. This reduced system, consisting of 10 tau oligomer chains with position constraints imposed on the outermost chains, was also used to determine free energy profiles for the binding of bTVBT4 to the tau fibril by means of umbrella sampling. The calculations of the potentials of mean force (PMFs) for the different binding modes are described in SI section 1.2.6.

Spectrum calculations were carried out at the level of CAM-B3LYP/aug-cc-pVDZ with varying degrees of exact Hartree–Fock exchange applied in the long-range limit (100% and the standard setting of 65%). We adopted the polarizable embedding (PE) model to describe the ligand environment, and details on the spectrum convergence with respect to PE parameters are presented in SI section 1.1.7 for the case of water solution. For the spectrum calculations of the bTVBT4 ligand in the binding site, there is an additional aspect of how to properly sample all binding modes. Details on our approach to addressing this issue are provided in SI section 1.2.7.

Experimental fluorescence spectra were obtained from frozen frontal cortical brain sections from patients with either Alzheimer's or Pick's disease pathology, labeled with a fluorescent dye conjugated to antibodies against tau fibrils. Details on these experimental procedures are presented in SI sections 2.1–2.2. Furthermore, an experimental absorption and emission study of the bTVBT4 ligand in various solvents was performed, and details about this are provided in SI section 2.3.

## RESULTS AND DISCUSSION

As demonstrated in a study of a flexible anionic pentameric oligothiophene (p-FTAA) targeting  $A\beta$ , binding modes of LCOs interacting with amyloids can be revealed by means of unbiased MD simulations on a time scale of a few hundred nanoseconds.<sup>26</sup> It was later conclusively shown that the fingerprinting optical signal responses of this particular ligand binding are primarily due to an increased planarity in the  $\pi$ -conjugated system of the ligand.<sup>27,29</sup> There are good reasons, however, to believe that the underlying microscopic mechanisms are different in the present case, as the bTVBT4 ligand differs in several important ways—(i) it is cationic instead of anionic with a charge that is largely delocalized over the  $\pi$ -conjugated system instead of being localized to carboxylate



**Figure 2.** (a) Experimental absorption and emission spectra of bTVBT4 in different solvents (PBS is phosphate-buffered saline) obtained at room temperature. (b) Theoretical absorption spectra for bTVBT4 in water solution using Gaussian and vibronic line broadenings. The 10 lowest states are included in the calculations performed at the level of TDDFT/CAM-B3LYP(100%).

side chains and (ii) the molecular structure is not as flexible due to the vinylenic double bond. It is reasonable to assume that the relevant binding modes of the ligands targeting tau are found in the structurally ordered core, and we therefore adopted the cryo-EM structure as a starting point in an attempt to reveal the most important binding modes of bTVBT4.

**Electronic Structure of bTVBT4.** We first performed a complete conformer study of bTVBT4 in the electronic ground state,  $S_0$ , and found the *trans*-conformer shown in Figure 1 to be energetically more stable by 2.4 kJ/mol as compared with the corresponding bi-thiophene *cis*-conformer. The conformers associated with variations of the three vinylenic dihedral angles were also considered but resulted in increases of the energy by 6.8–58.4 kJ/mol, indicating that these conformers are not as relevant for ligand binding.

The positive charge of the cationic ligand is delocalized, and with a split separation made in the vinylenic double bond, the bi-thiophene and benzothiazole units accumulate charges of 0.44e and 0.56e, respectively, in the ground state and 0.47e and 0.53e in the excited state, which suggests a small charge transfer associated with the  $S_1 \leftarrow S_0$  transition. With a gauge-origin chosen as the center of nuclear charge, the dipole moment of the cationic ligand becomes well defined and point in the direction from the bi-thiophene toward the benzothiazole moiety. Our charge analysis predicts a lower dipole moment in the  $S_1$  state as compared to the  $S_0$  state, and this finding is corroborated by a quadratic response theory<sup>31</sup> calculation of the (always) well-defined dipole moment difference that, here, predicts a reduction of the dipole moment from 5.5 to 2.1 D; see Figure S2 in the SI. It also agrees with the observed negative and zero solvatochromic shifts in the experimental absorption and emission spectroscopies, respectively, as presented in Figure 2a. The reasoning behind this is that, in absorption, the ground state (with its

larger dipole moment and being relaxed also with respect to the slow degrees of freedom of the solvent) becomes more stabilized by the binding reaction field of the solvent as compared to the excited state, and hence, the transition energy increases with solvent polarity. In emission, on the other hand, the small dipole moment of the excited (initial) state causes the energy to be weakly dependent on the solvent polarity, and the ground state will not be in equilibrium with the slow degrees of freedom of the solvent, resulting in a weak dependence also of the final state energy with respect to solvent polarity. Arguments combined explain the observed small variation of the transition energy with solvent polarity in emission spectroscopy.

An indirect probe of changes in the electronic structure between the ground and excited states is provided by the comparison of equilibrium molecular structures. For pentameric oligothiophene ligands, these changes have been shown to be localized to the three central rings and interpreted as a quite localized exciton. For these central units, the main changes amount to single–double bond inversion, resulting in a rigid planar system in the excited state.<sup>28</sup> For the cationic bTVBT4, this effect is much less pronounced, and bond length parameters in the  $\pi$ -conjugated backbone are largely unaffected by the electronic excitation; see Figure 1. As a consequence, and as we shall see, this leads to intense 0–0 transitions in the vibronic absorption spectra.

At the Franck–Condon points, the transition state properties are determined with the use of the time-dependent density functional theory (TDDFT) method in conjunction with a range-separated hybrid exchange–correlation functional (CAM-B3LYP) to account for the charge-transfer character in the  $S_1 \leftarrow S_0$  transition. There is no significant difference in the properties for the *trans*- and *cis*-conformer, as seen in Table 1. Varying the amount of exact exchange in the long-range

**Table 1.** Vertical Absorption Energies ( $\Delta E$  in eV), Transition Wavelengths ( $\lambda$  in nm), and Oscillator Strengths ( $f$ ) for the Two Lowest Conformations of bTVBT4, Where the Trans-Conformation is Shown in Figure 1<sup>a</sup>

conformer	CAM-B3LYP(100%)			CAM-B3LYP(65%)		
	$\Delta E$	$\lambda$	$f$	$\Delta E$	$\lambda$	$f$
<i>trans</i>	2.619	473.5	1.45	2.541	487.9	1.45
<i>cis</i>	2.621	473.0	1.43	2.547	486.8	1.44

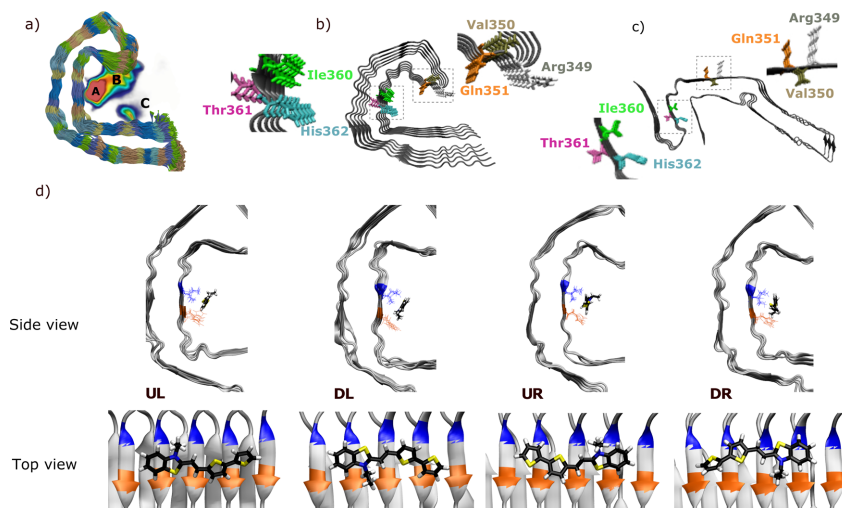
<sup>a</sup>A varying percentage degree of exact exchange is used in the long-range limit (100% and the standard setting of 65%) in both cases combined with the aug-cc-pVDZ basis set.

limit from the standard setting of 65 to 100% (fully correct asymptotic behavior in charge-transfer excitations) gives rise to a small increase in the excitation energy of 0.08 eV, in line with the transition being predominantly described by a single-electron excitation from the highest occupied molecular orbital ( $\pi$ ) to lowest unoccupied molecular orbital ( $\pi^*$ ) and where the former is somewhat more localized to the bi-thiophene group.

**Spectra of bTVBT4 in Water.** Experimental emission spectra of LCO ligands often display clear vibrational progressions that to a high degree of precision can be simulated with the use of anharmonic vibrational configuration interaction wave function calculations, as demonstrated in the case of oligothiophenes with an attribution of the observed vibrational progressions with a separation of some 0.18–0.19 eV to inter-ring carbon–carbon stretching and ring breathing motions.<sup>32</sup> The relatively large force constants associated with ring rotations in the excited state promote such a vibrational resolution in the experiment. In the ground state, this is not the case, and therefore, in calculations of absorption spectra, it becomes necessary to also account for the slow dihedral motions that are not well described by rectilinear coordinates.

As a remedy, we propose to, in a first step, perform an anharmonic vibrational calculation on the ligand in isolation in the low-temperature limit, such that small oscillations from a well-defined molecular structure minimum provide an accurate description of the nuclear motions. For bTVBT4, this results in the identification of three vibrational stretching and angle-bending modes in the region of 1300–1600  $\text{cm}^{-1}$  to be of prime importance for the absorption spectrum profile. The 0–0 transition dominates the absorption spectrum, leading to a strongly inhomogeneous broadening; see Figure S3 in the SI. In a second step, we adopt this low-temperature spectrum profile in combination with room-temperature MD simulations to account for the slow degrees of motion by means of the technique of snapshot averaging.<sup>28,33</sup> This removes the need to introduce an *ad hoc* line broadening (typically Gaussian or Lorentzian) in the simulations and instead relies on the arguably reasonable approximation that the fast and slow degrees of vibrational motions can be decoupled.

For bTVBT4, we derived the set of molecular mechanics force field parameters based on the general AMBER force field (GAFF) with force constants describing the four dihedral rotations subsequently fitted to relaxed-scan potential energy curves obtained at the level of DFT/B3LYP, see Section 1.1.2 in SI for a detailed description. Together with the standard TIP3P force field for water, we performed room-temperature MD simulations of the ligand in an aqueous solution with a chlorine ion introduced to neutralize the system. Only the bi-thiophene *cis/trans*-isomerization can be observed in the dynamics, and it resulted in a statistical 50/50 conformer population in solution as compared to 35/65 in vacuum. The relative stabilization of the *cis*-conformer in solution is connected with the larger *local* dipole moment in the bi-thiophene moiety as compared to that in the *trans*-conformer.



**Figure 3.** Summary of the MD simulations of the interactions between bTVBT4 and the amyloid fibrillar structure of tau. (a) Density map of the bTVBT4 interactions with the tau Alzheimer's fold, where red indicates the highest, blue a lower, and white a zero ligand density, (b) Alzheimer's fold (PDB ID: 5O3L), (c) Pick's fold (PDB ID: 6GX5), and (d) identified binding modes for bTVBT4 in site A during MD simulation, with separation of down (D), up (U), left (L), and right (R) orientation of the ethyl group with respect to His362 (orange) and Ile360 (blue).

**Table 2. Lennard-Jones (LJ) and Coulombic Interaction Energies (in kJ/mol) between bTVBT4 and Residues Ile360, Thr361, and His362 That Make up Binding Site A; and Arg349, Val350, and Gln351 That Make up Binding Site B<sup>a</sup>**

site	type				total
A		Ile360	Thr361	His362	
	LJ	-50.5 ± 7	-4.8 ± 2	-36.7 ± 8	-91.9 ± 9
B	Coulomb	4.8 ± 1	-1.8 ± 2	-24.9 ± 14	-21.9 ± 14
		Arg349	Val350	Gln351	
	LJ	-73.2 ± 13	-13.0 ± 4.0	-32.0 ± 11	-118.2 ± 14
	Coulomb	5.6 ± 12	-7.8 ± 4	1.8 ± 9	-0.5 ± 16

<sup>a</sup>Residue Arg349 has a charge of +e, whereas others are charge-neutral.

The subsequent spectrum calculations of bTVBT4 in water were performed with the employment of the PE model using the standard Ahlstrom isotropic polarizabilities and atomic charges for a polarizable shell of thickness 15 Å and the TIP3P charges for an exterior nonpolarizable shell of thickness 5 Å—the calculations are well converged with respect to shell thicknesses, see Figure S8 in the SI. We performed 200 snapshot calculations (100 from each of the *cis/trans*-conformers), and the resulting spectra are depicted in Figure 2b, both using the vibronic line broadening as described above, as well as using a conventional Gaussian line profile, with a standard deviation of 0.15 eV. The two spectra are found to be in close agreement, indicating that the number of snapshots is sufficiently large to almost fully smear out the inhomogeneity in the vibronic line profile. Taken separately, the *cis*- and *trans*-conformations yield averaged  $\lambda_{\text{max}}$ -values of 485 and 480 nm, respectively, and taken together, the value for  $\lambda_{\text{max}}$  becomes equal to 483 nm. This theoretical result should be compared to the experimental value of 470 nm in PBS solution, see Figure 2a, which amounts to a discrepancy of 0.07 eV. We retrieved a representative *trans*-conformer snapshot and incorporated the first solvation shell of 14 water molecules into the quantum mechanical region, and this, in effect, changed the transition wavelength from 480 to 475 nm. Based on these results, we deem our theoretical ligand spectrum in solution to be highly accurate and suitable as a reference for the assessment of spectral changes due to protein binding.

**Binding of bTVBT4 to Tau.** Based on the experimental protofilament structure,<sup>2</sup> we built a series of forced periodic, right-handed helical, tau fibril systems, all with a 2<sub>1</sub> screw axis symmetry but with slightly varying twist angles. During the NPT-relaxations of these systems, the forced periodicity induced stress that at times resulted in structural kinks. Our smoothest system (with least stress) was found using 185 protofilament layers, corresponding to a twist angle of 0.97°, with a rise of 4.8 Å at 300 K, which is in close agreement with the experimental cryo-EM structure.<sup>2</sup>

Separate unbiased MD simulations were run with 60 ligands inside and outside the cavity formed by the C-shaped protofilaments along the fibril axis. In Figure 3a, we present the combined ligand density for these simulations, where the density is averaged over time (200 ns) as well as summed along the fibril axis by overlaying all protofilaments on top of one another. From this density plot with maximum values color-coded in red, it is clear that the strongest ligand–protein interactions occur at two separate sites inside the cavity labeled A (higher density) and B (lower density).

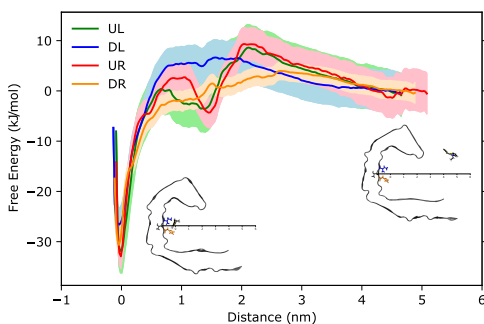
In binding site A, bTVBT4 primarily interacts with residues Ile360, Thr361, and His362, and in site B, the ligand primarily interacts with Arg349, Val350, and Gln351; see Figure 3b. From the interaction energies presented in Table 2, it is clear

that the ligand binding at site B is driven by Lennard-Jones interactions, whereas that at site A also has significant contributions from electrostatic interactions. In a recent experimental study, the tau binding sites for a ligand denoted APN-1607, were revealed with the cryo-EM technique.<sup>34</sup> Common for the ligand structures of APN-1607 and bTVBT4 are the benzothiazole scaffold and vinylene moiety, and we note that the binding sites A and B found in our unbiased MD simulations are also identified in the cryo-EM maps but with reversed probabilities, *i.e.*, for APN-1607, site B is the major binding site, whereas site A is one out of several minor binding sites (see Figure 1e in ref 34). This suggests that the bi-thiophene moiety steers the binding toward site A and exemplifies an exciting prospect of ligand binding control by chemical design. A theoretical docking study of several PET tau tracers belonging to different chemical families has been conducted,<sup>35</sup> showing binding to site B but not site A and thus suggesting that bTVBT4 stands out in comparison to other ligands.

Focusing on the strongest binding site A, there are four major binding modes depending on the up/down (U/D) and left/right (L/R) orientation of the ethyl group with respect to the reference frame defined in Figure 3d. In addition, for each major mode, there are four minor modes associated with the dihedral rotations around the thiazole–ethyl bond and the bi-thiophene inter-ring bond, but the barriers in between these conformations are low, and interconversions occur frequently during the course of the dynamics; see Figures S6 and S12 in the SI.

To obtain the free energy profiles for the bTVBT4 binding in the four major modes of site A, we adopted the PMF approach in conjunction with umbrella sampling.<sup>36</sup> The ligand was pulled from each of the major modes some 5 nm in the direction of the entrance of the cavity, as depicted in the two insets of Figure 4. At a pulling distance of about 1 nm, some of the trajectories resulted in interactions with site B, which is seen as shallow local minima on the potential. As exemplified by the PMF for mode DR, barrierless access to site A inside the cavity is available for bTVBT4. Further, we note that the binding energies are about 26–33 kJ/mol for the different modes with a mode ranking in terms of binding strength according to UR > UL > DR > DL, *i.e.*, up is favored over down and right is favored over left. Based on the binding energies calculated from this PMF study, the Boltzmann distribution of the major modes at site A are estimated to be 47, 30, 19, and 4%, respectively.

**Absorption Spectra of bTVBT4 in the Tau Binding Site.** A correct spectrum averaging also requires the minor mode populations, and we obtained these by performing separate MD simulation runs for each of the four major modes (500 ns in each mode) and determining the times spent in the



**Figure 4.** Free energy profiles for bTVBT4 in the UL (green), DL (blue), UR (red), and DR (orange) modes, obtained by the potential of mean force approach by pulling the ligand from binding site A to become free in solution.

minor modes—a smaller model system based on 10 protofilament layers was used in these time-extended simulations. In the end, 300 uncorrelated snapshots with at least 20 ps separations were extracted from the simulations in accordance with the major and minor mode population statistics.

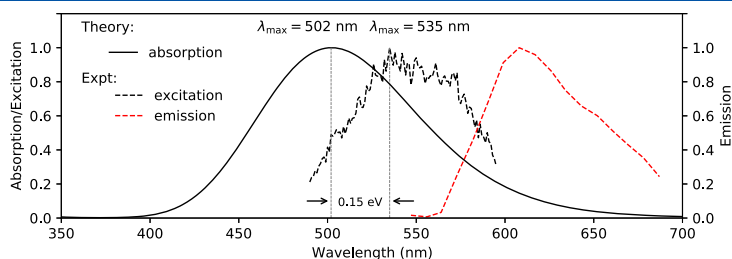
We determined the PE parameters needed for the spectrum calculations by means of the procedure of molecular fractionation with conjugate caps.<sup>37</sup> For the  $A\beta(1-42)$  fibril periodic model system, it has been shown that the PE parameters determined in this manner can be considered to be time-independent, *i.e.*, not changing over the course of the dynamics as the fibril is in fact quite rigid.<sup>27</sup> But, the ligand p-FTAA exhibits a strong Coulombic-driven binding to  $A\beta$  at fixed loci, while the bTVBT4 ligand shows movements along the fibril axis while remaining in binding site A (see Figure S13). We therefore approximated the PE parameters for tau not only to be time-independent but also unresponsive to the ligand interaction, *i.e.*, we determined a single set of protofilament parameters by averaging over 10 layers in the absence of a ligand and then adopted this set to all protofilaments and all snapshots—the alternative to derive snapshot-specific PE parameters was not pursued due to reasons of computational cost.

With a polarizable embedding of 20 Å and based on 300 snapshots to reach model convergence, the absorption

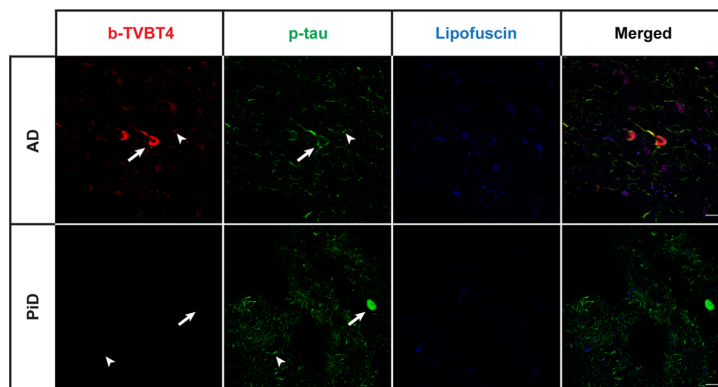
spectrum of bTVBT4 in binding site A was obtained and is presented in Figure 5. The theoretical absorption spectrum has a  $\lambda_{\max}$  of 502 nm, which represents a blue shift of 33 nm compared to the corresponding experimental excitation spectrum. Expressed in terms of energy, the discrepancy between theory and experiment amounts to 0.15 eV, and we deem this to be reasonable, given the nonspecific and fully classical description that had to be adopted for the protein environment.

The planarity parameter defined in ref 29 was shown to correlate well to the transition energy for anionic LCOs and be the key descriptor for the changes of optical responses of p-FTAA upon binding to  $A\beta$ .<sup>27</sup> For bTVBT4, this parameter changes insignificantly from  $3.43 \pm 0.20$  in solution to  $3.49 \pm 0.18$  in site A of tau, suggesting that the associated red shift in the absorption spectrum is primarily due to changes in the electronic structure and not molecular structure. When comparing experimental decay times of bTVBT4 in various solutions and when bound to tau aggregates in AD, it is noted that it displays strikingly longer decay times in the latter case, ranging from 1.7 to 2.4 ns; see Figure S19 in the SI. Such increases in LCO decay times and also fluorescence intensities are believed to be associated with constrained vibrational motions,<sup>27</sup> so although the difference in the theoretical planarity parameter is small for bTVBT4 in between solution and protein, there is experimental evidence that bTVBT4 adopts a more distinct conformation when bound to the aggregates.

When comparing the cryo-EM structure for the tau fold in AD<sup>2</sup> to that in Pick's disease (PiD) fold,<sup>38</sup> see Figure 3c, it is seen that the binding site A in tau is ligand-accessible in AD but not in PiD. Therefore, we next stained brain tissue samples from patients diagnosed with AD or PiD with bTVBT4 as well as an antibody reference marker AT8 that detects abnormally phosphorylated tau present in both AD and PiD;<sup>39</sup> see column two in Figure 6. When stained with 100 nM bTVBT4, the AT8-immunopositive aggregates in AD, such as neurofibrillary tangles, neuropil threads, and dystrophic neurites, become visible through bTVBT4 fluorescence, whereas the AT8-immunopositive aggregates in PiD, such as Pick bodies, do not; see column 1 in Figure 6. The staining of tau deposits in PiD was also absent when using 10 times higher (1  $\mu$ M) ligand concentration (data not shown). Hence, these tissue-staining experiments support our proposal of site A as the main tau binding site for the cationic bTVBT4 ligand because the



**Figure 5.** Theoretical absorption and experimental excitation and emission spectra for bTVBT4 bound to the tau protein in the Alzheimer fold. The 10 lowest states are included in the calculations performed at the level of TDDFT/CAM-B3LYP(100%). The experimental excitation and emission spectra were recorded at room temperature from tau deposits in a bTVBT4-stained AD brain tissue section washed with PBS (see SI section 2.2 for further details).



**Figure 6.** Fluorescence images of the brain section from a patient diagnosed with Alzheimer's disease (AD, top panel) or Pick's disease (PiD, bottom panel) stained with 100 nM bTVBT4 (red) and phospho-tau antibody AT8 (p-tau, green). Arrow: neurofibrillary tangle (AD), Pick body (PiD). Arrowhead: neuropil thread. As the autofluorescence from lipofuscin granules can overlap with bTVBT4 emission, an additional channel in which the settings only allowed excitation of lipofuscin (blue) is also shown. Scale bars represent 20  $\mu\text{m}$ .

disease-specificity of the ligand is thereby explained by the differences in protofilament folds in AD and PiD.

## SUMMARY AND CONCLUSIONS

A comprehensive theoretical–experimental study of the cationic luminescent ligand bTVBT4 has been carried out in the context of tau protein detection in the brain section from patients diagnosed with Alzheimer's disease. The theoretical work includes (i) the development of a tau fibril model system based on the published cryo-EM structure,<sup>2</sup> (ii) the methodological advancement for simulations of inhomogeneous vibronic absorption spectra of ligands in solution by a combination of an anharmonic vibrational theory and snapshot averaging, (iii) the identification of tau binding sites for bTVBT4 by means of unbiased atomistic molecular dynamics simulations, and (iv) the characterization of the strongest binding site by means of potentials of mean force and absorption spectra. The experimental work included (i) the determination of solvatochromic shifts in absorption and emission spectra, (ii) the determination of the excitation and emission spectra of bTVBT4 bound to tau, and (iii) the presentation of fluorescence images of the brain section from patients diagnosed with Alzheimer's and Pick's diseases as to address the specificity in the bTVBT4 binding. All things considered, there is strong evidence for the strongest interactions between bTVBT4 and tau to occur at the site involving residues Ile360, Thr361, and His362, and for the binding at this site to be predominantly driven by Lennard-Jones interactions.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.1c06019>.

Vibronic UV/vis spectra; details on the force field parametrization as well as MD, PMF, and spectrum simulations; ligand planarity analysis; and experimental details (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Patrick Norman – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden; [orcid.org/0000-0002-1191-4954](https://orcid.org/0000-0002-1191-4954); Email: [panor@kth.se](mailto:panor@kth.se)

### Authors

Yogesh Todarwal – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden

Camilla Gustafsson – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden; [orcid.org/0000-0002-7624-4867](https://orcid.org/0000-0002-7624-4867)

Nghia Nguyen Thi Minh – Leibniz University Hannover, Institute of Physical Chemistry and Electrochemistry, 30167 Hannover, Germany

Ingrid Ertzgaard – Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

Therése Klingstedt – Department of Physics, Chemistry and Biology, Linköping University, SE 581 83 Linköping, Sweden

Bernardino Ghetti – Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States

Ruben Vidal – Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States

Carolin König – Leibniz University Hannover, Institute of Physical Chemistry and Electrochemistry, 30167 Hannover, Germany; [orcid.org/0000-0001-8931-4337](https://orcid.org/0000-0001-8931-4337)

Mikael Lindgren – Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim, Norway; [orcid.org/0000-0001-6649-7871](https://orcid.org/0000-0001-6649-7871)

K. Peter R. Nilsson – Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden



Mathieu Linares – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden; Laboratory of Organic Electronics, ITN and Scientific Visualization Group, ITN, Linköping University, SE-581 83 Linköping, Sweden; [orcid.org/0000-0002-9720-5429](https://orcid.org/0000-0002-9720-5429)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jpcb.1c06019>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Financial support is acknowledged from the European Commission (Grant No. 765739), the Swedish Research Council (Grant Nos. 2018-4343 and 2016-07213), the Swedish e-Science Research Centre (SeRC), the German Research Foundation (Project No. KO 5423/1-1), and the U.S. National Institutes of Health (Grant No. UO1NS110437) as well as computational resources provided by the Swedish National Infrastructure for Computing (SNIC).

## REFERENCES

- (1) Nguyen, P. H.; Ramamoorthy, A.; Sahoo, B. R.; Zheng, J.; Faller, P.; Straub, J. E.; Dominguez, L.; Shea, J.-E.; Dokholyan, N. V.; De Simone, A.; et al. Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chem. Rev.* **2021**, *121*, 2545–2647.
- (2) Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W.; et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017**, *547*, 185–190.
- (3) van Eersel, J.; Bi, M.; Ke, Y. D.; Hodges, J. R.; Xuereb, J. H.; Gregory, G. C.; Halliday, G. M.; Götz, J.; Kril, J. J.; Ittner, L. M. Phosphorylation of soluble tau differs in Pick's disease and Alzheimer's disease brains. *J. Neural. Transm.* **2009**, *116*, 1243–1251.
- (4) Hanger, D. P.; Byers, H. L.; Wray, S.; Leung, K.-Y.; Saxton, M. J.; Seereeram, A.; Reynolds, C. H.; Ward, M. A.; Anderton, B. H. Novel Phosphorylation Sites in Tau from Alzheimer Brain Support a Role for Casein Kinase 1 in Disease Pathogenesis. *J. Biol. Chem.* **2007**, *282*, 23645–23654.
- (5) Smet-Nocca, C. *Tau Protein: Methods and Protocols*; Smet-Nocca, C., Ed.; Methods in Molecular Biology; Humana Press, 2017; Vol. 1523. <https://www.springer.com/gp/book/9781493965960>
- (6) Klunk, W. E.; Engler, H.; Nordberg, A.; Wang, Y.; Blomqvist, G.; Holt, D. P.; Bergström, M.; Savitcheva, I.; Huang, G.-F.; Estrada, S.; et al. Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann. Neurol.* **2004**, *55*, 306–319.
- (7) Small, G. W.; Kepe, V.; Ercoli, L. M.; Siddarth, P.; Bookheimer, S. Y.; Miller, K. J.; Lavretsky, H.; Burggren, A. C.; Cole, G. M.; Vinters, H. V.; et al. PET of Brain Amyloid and Tau in Mild Cognitive Impairment. *N. Engl. J. Med.* **2006**, *355*, 2652–2663.
- (8) Kudo, Y.; Okamura, N.; Furumoto, S.; Tashiro, M.; Furukawa, K.; Maruyama, M.; Itoh, M.; Iwata, R.; Yanai, K.; Arai, H. 2-(2-[2-Dimethylaminothiazol-5-yl]Ethenyl)-6-(2-[Fluoro]Ethoxy)-Benzoxazole: A Novel PET Agent for In Vivo Detection of Dense Amyloid Plaques in Alzheimer's Disease Patients. *J. Nucl. Med.* **2007**, *48*, 553–561.
- (9) Taghavi, A.; Nasir, S.; Pickhardt, M.; Haußen, R. H.-v.; Mall, G.; Mandelkow, E.; Mandelkow, E.-M.; Schmidt, B. N'-Benzylidene-Benzohydrazides as Novel and Selective Tau-PHF Ligands. *J. Alzheimer's Dis.* **2011**, *27*, 835–843.
- (10) Fodero-Tavoletti, M. T.; Okamura, N.; Furumoto, S.; Mulligan, R. S.; Connor, A. R.; McLean, C. A.; Cao, D.; Rigopoulos, A.; Cartwright, G. A.; O'Keefe, G.; et al. 18F-THK523: a novel in vivo tau imaging ligand for Alzheimer's disease. *Brain* **2011**, *134*, 1089–1100.
- (11) Yang, L.; Rieves, D.; Ganley, C. Brain Amyloid Imaging - FDA Approval of Florbetapir F18 Injection. *N. Engl. J. Med.* **2012**, *367*, 885–887.
- (12) Zhang, W.; Arteaga, J.; Cashion, D. K.; Chen, G.; Gangadharmath, U.; Gomez, L. F.; Kasi, D.; Lam, C.; Liang, Q.; Liu, C.; et al. A Highly Selective and Specific PET Tracer for Imaging of Tau Pathologies. *J. Alzheimer's Dis.* **2012**, *31*, 601–612.
- (13) Maruyama, M.; Shimada, H.; Suhara, T.; Shinotoh, H.; Ji, B.; Maeda, J.; Zhang, M.-R.; Trojanowski, J. Q.; Lee, V. M.-Y.; Ono, M.; et al. Imaging of Tau Pathology in a Tauopathy Mouse Model and in Alzheimer Patients Compared to Normal Controls. *Neuron* **2013**, *79*, 1094–1108.
- (14) Xia, C. F.; Arteaga, J.; Chen, G.; Gangadharmath, U.; Gomez, L. F.; Kasi, D.; Lam, C.; Liang, Q.; Liu, C.; Mocharla, V. P.; et al. [18F]T807, a Novel Tau Positron Emission Tomography Imaging Agent for Alzheimer's Disease. *Alzheimer's Dement.* **2013**, *9*, 666–676.
- (15) Kolb, H. C.; Andrés, J. I. Tau Positron Emission Tomography Imaging. *Cold Spring Harb. Perspect. Biol.* **2017**, *9*, No. a023721.
- (16) Naiki, H.; Higuchi, K.; Hosokawa, M.; Takeda, T. Fluorometric Determination of Amyloid Fibrils in vitro using the Fluorescent dye, Thioflavine T. *Anal. Biochem.* **1989**, *177*, 244–249.
- (17) Puchtler, H.; Sweat, F. Congo Red as a Stain for Fluorescence Microscopy of Amyloid. *J. Histochem. Cytochem.* **1965**, *13*, 693–694.
- (18) Rasmussen, J.; Mahler, J.; Beschoner, N.; Kaeser, S. A.; Häslér, L. M.; Baumann, F.; Nyström, S.; Portelius, E.; Blennow, K.; Lashley, T.; et al. Amyloid Polymorphisms Constitute Distinct Clouds of Conformational Variants in Different Etiological Subtypes of Alzheimer's Disease. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 13018–13023.
- (19) Åslund, A.; Sigurdson, C. J.; Klingstedt, T.; Grathwohl, S.; Bolmont, T.; Dickstein, D. L.; Glimsdal, E.; Prokop, S.; Lindgren, M.; Konradsson, P.; et al. Novel Pentameric Thiophene Derivatives for In Vitro and In Vivo Optical Imaging of a Plethora of Protein Aggregates in Cerebral Amyloidoses. *ACS Chem. Biol.* **2009**, *4*, 673–684.
- (20) Calvo-Rodríguez, M.; Hou, S. S.; Snyder, A. C.; Dujardin, S.; Shirani, H.; Nilsson, K. P. R.; Bacskaï, B. J. In vivo detection of tau brils and amyloid aggregates with luminescent conjugateddithiophenes and multiphoton microscopy. *Acta Neuropathol. Commun.* **2019**, *7*, 171.
- (21) Sigurdson, C. J.; Nilsson, P. R.; Hornemann, S.; Manco, G.; Polymenidou, M.; Schwarz, P.; Leclerc, M.; Hammarström, P.; Wüthrich, K.; Aguzzi, A. Prion Strain Discrimination Using Luminescent Conjugated Polymers. *Nat. Methods* **2007**, *4*, 1023–1030.
- (22) Klingstedt, T.; Åslund, A.; Simon, R. A.; Johansson, L. B.; Mason, J. J.; Nyström, S.; Hammarström, P.; Nilsson, K. P. R. Synthesis of a Library of Oligothiophenes and Their Utilization as Fluorescent Ligands for Spectral Assignment of Protein Aggregates. *Org. Biomol. Chem.* **2011**, *9*, 8356–8370.
- (23) Klingstedt, T.; Shirani, H.; Åslund, K. O. A.; Cairns, N. J.; Sigurdson, C. J.; Goedert, M.; Nilsson, K. P. R. The Structural Basis for Optimal Performance of Oligothiophene-Based Fluorescent Amyloid Ligands: Conformational Flexibility is Essential for Spectral Assignment of a Diversity of Protein Aggregates. *Chem. - Eur. J.* **2013**, *19*, 10179–10192.
- (24) Shirani, H.; Linares, M.; Sigurdson, C. J.; Lindgren, M.; Norman, P.; Nilsson, K. P. R. A Palette of Fluorescent Thiophene-Based Ligands for the Identification of Protein Aggregates. *Chem. - Eur. J.* **2015**, *21*, 15133–15137.
- (25) Shirani, H.; Appelqvist, H.; Bäck, M.; Klingstedt, T.; Cairns, N. J.; Nilsson, K. P. R. Synthesis of Thiophene-Based Optical Ligands That Selectively Detect Tau Pathology in Alzheimer's Disease. *Chemistry* **2017**, *23*, 17127–17135.
- (26) König, C.; Skånberg, R.; Hotz, I.; Ynnerman, A.; Norman, P.; Linares, M. Binding sites for luminescent amyloid biomarkers from non-biased molecular dynamics simulations. *Chem. Commun.* **2018**, *54*, 3030–3033.

- (27) Gustafsson, C.; Linares, M.; Norman, P. Quantum Mechanics/Molecular Mechanics Density Functional Theory Simulations of the Optical Properties Fingerprinting the Ligand-Binding of Pentameric Formyl Thiophene Acetic Acid in Amyloid- $\beta$ (1-42). *J. Phys. Chem. A* **2020**, *124*, 875–888.
- (28) Sjöqvist, J.; Linares, M.; Lindgren, M.; Norman, P. Molecular dynamics effects on luminescence properties of oligothiophene derivatives: A molecular mechanics-response theory study based on the CHARMM force field and density functional theory. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17532–17542.
- (29) Sjöqvist, J.; Maria, J.; Simon, R. A.; Linares, M.; Norman, P.; Nilsson, K. P. R.; Lindgren, M. Toward a molecular understanding of the detection of amyloid proteins with flexible conjugated oligothiophenes. *J. Phys. Chem. A* **2014**, *118*, 9820–9827.
- (30) Sjöqvist, J.; Linares, M.; Mikkelsen, K. V.; Norman, P. QM/MM-MD simulations of conjugated polyelectrolytes: A study of luminescent conjugated oligothiophenes for use as biophysical probes. *J. Phys. Chem. A* **2014**, *118*, 3419–3428.
- (31) Norman, P.; Ruud, K.; Saue, T. *Principles and Practices of Molecular Properties*; John Wiley & Sons, Ltd.: Chichester, UK, 2018.
- (32) Madsen, D.; Christiansen, O.; Norman, P.; König, C. Vibrationally resolved emission spectra of luminescent conjugated oligothiophenes from anharmonic calculations. *Phys. Chem. Chem. Phys.* **2019**, *21*, 17410–17422.
- (33) Sjöqvist, J.; Linares, M.; Norman, P. Platinum(II) and Phosphorus MM3 Force Field Parametrization for Chromophore Absorption Spectra at Room Temperature. *J. Phys. Chem. A* **2010**, *114*, 4981–4987.
- (34) Shi, Y.; Murzin, A. G.; Falcon, B.; Epstein, A.; Machin, J.; Tempest, P.; Newell, K. L.; Vidal, R.; Garringer, H. J.; Sahara, N.; et al. Cryo-EM structures of tau filaments from Alzheimer's disease with PET ligand APN-1607. *Acta Neuropathol.* **2021**, *141*, 697–708.
- (35) Murrigan, N. A.; Nordberg, A.; Ågren, H. Different Positron Emission Tomography Tau Tracers Bind to Multiple Binding Sites on the Tau Fibril: Insight from Computational Modeling. *ACS Chem. Neurosci.* **2018**, *9*, 1757–1767.
- (36) Leach, A. *Molecular Modelling: Principles and Applications*, 2nd ed.; Pearson: Dorchester, Dorset, 2001.
- (37) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (38) Falcon, B.; Zhang, W.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Vidal, R.; Crowther, R. A.; Ghetti, B.; Scheres, S. H. W.; Goedert, M. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **2018**, *561*, 137–140.
- (39) Goedert, M.; Jakes, R.; Vanmechelen, E. Monoclonal antibody AT8 recognises tau protein phosphorylated at both serine 202 and threonine 205. *Neurosci. Lett.* **1995**, *189*, 167–170.

S1

# Tau Protein Binding Modes in Alzheimer's Disease for Cationic Luminescent Ligands

Yogesh Tadarwal,<sup>†</sup> Camilla Gustafsson,<sup>†</sup> Nghia Nguyen Thi Minh,<sup>‡</sup> Ingrid Ertzgaard,<sup>¶</sup> Therése Klingstedt,<sup>§</sup> Bernardino Ghetti,<sup>||</sup> Ruben Vidal,<sup>||</sup> Carolin König,<sup>‡</sup> Mikael Lindgren,<sup>¶</sup> K. Peter R. Nilsson,<sup>⊥</sup> Mathieu Linares,<sup>†,#,@</sup> and Patrick Norman<sup>\*,†</sup>

<sup>†</sup>*Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden*

<sup>‡</sup>*Leibniz University Hannover, Institute of Physical Chemistry and Electrochemistry, Callinstr. 3A, 30167 Hannover, Germany*

<sup>¶</sup>*Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

<sup>§</sup>*Department of Physics, Chemistry and Biology, Linköping University, SE 581 83 Linköping, Sweden*

<sup>||</sup>*Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, IN 46202, USA*

<sup>⊥</sup>*Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden*

<sup>#</sup>*Laboratory of Organic Electronics, ITN, Linköping University, SE-581 83 Linköping, Sweden*

<sup>@</sup>*Scientific Visualization Group, ITN, Linköping University, SE-581 83, Linköping, Sweden*

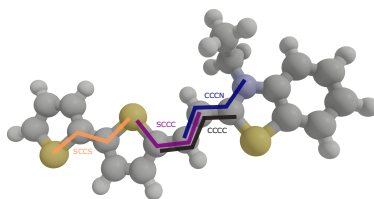
E-mail: panor@kth.se

# Supplementary Information

## 1 Theoretical methodology, computational details, and supplementary results

### 1.1 Modeling bTVBT4 biomarker properties in vacuum and solution

All geometry optimization calculations were performed in the Gaussian (version 16.B.01) program,<sup>1</sup> and all spectra calculations were done in the Dalton (version 2018.alpha) program.<sup>2</sup> The ground state geometry optimization of all the conformers (Fig. S1b) was carried out using B3LYP<sup>3-5</sup> density functional theory in conjunction with the 6-31+G(d,p)<sup>6,7</sup> basis set. A single point absorption spectra calculation was performed on each optimized geometry of the conformer using TDDFT with CAM-B3LYP(100%) [CAM-B3LYP ( $\alpha = 0.19, \beta = 0.81, \mu = 0.33$ )]<sup>8</sup> functional in combination with aug-cc-pVDZ<sup>9</sup> basis set. The choice of basis set 6-31+G(d,p) for the geometry optimization and aug-cc-pVDZ for single point absorption spectra calculation was made to reduce computational cost, and in addition, as shown in our benchmark calculations summarized in Table S3, both basis set yield very similar values for absorption maxima for the  $S_1 \leftarrow S_0$  transition. For the most stable conformers *cis* and *trans* (conformer 1 and 2 in the Fig. S1b), the ground state geometries was re-optimized at the B3LYP/aug-cc-pVDZ and dipole moment were calculated at the CAM-B3LYP(100%)/aug-cc-pVDZ method, keeping gauge-origin at the center of nuclear charge. The Fig. S2 shows the dipole moment vectors of *trans* conformer in the ground state ( $\mu_{S_0}$ ), excited state ( $\mu_{S_1}$ ) and their difference ( $\mu_{\text{diff}}$ ). The magnitudes of  $\mu_{S_0}$ ,  $\mu_{S_1}$ , and  $\mu_{\text{diff}}$  are 5.54 D, 2.08 D, and 4.74 D, respectively. The dipole moment of the first excited state is significantly lower as compared to the ground state. This decrease in dipole moment can also be seen from the NPA charge analysis shown in Fig. 1.



(a)

bTVBT4 geometry	SCCS	SCCC	CCCC	CCCN	$\Delta E$ (kJ/mol)	$\lambda$ nm (eV)
	180	0	180	180	0.0	A: 472 (2.63) B: 488 (2.54)
	0	0	180	180	2.4	A: 472 (2.63) B: 488 (2.54)
	180	180	180	180	6.8	A: 479 (2.59) B: 496 (2.50)
	0	180	180	180	9.5	A: 479 (2.59) B: 494 (2.51)
	180	0	0	180	32.4	A: 477 (2.60)
	0	0	0	180	34.2	A: 473 (2.62)
	180	180	0	180	35.0	A: 490 (2.53)
	0	180	0	180	37.8	A: 488 (2.54)
bTVBT4 geometry	SCCS	SCCC	CCCC	CCCN	$\Delta E$ (kJ/mol)	$\lambda$ nm (eV)
	180	0	180	0	16.1	A: 468 (2.65)
	0	0	180	0	18.8	A: 468 (2.65)
	180	180	180	0	22.4	A: 473 (2.61)
	0	180	180	0	24.8	A: 471 (2.63)
	180	0	0	0	41.3	A: 440 (2.82)
	0	0	0	0	43.0	A: 436 (2.84)
	180	180	0	0	55.2	A: 421 (2.95)
	0	180	0	0	58.4	A: 418 (2.96)

(b)

Figure S1: (a) Naming convention for dihedral angles. (b) Possible conformations of bTVBT4. For simplicity in identifying conformer, dihedral angles between 0–90 degrees are denoted as 0, and angles between 90–180 degrees are denoted as 180. Relative energies are calculated with respect to most stable conformer. All structures were optimized at the B3LYP/6-31+G(d,p), followed by frequency calculation at the same level. All absorption spectra calculations were performed at two different levels: [A] CAM-B3LYP(100%)/aug-cc-pVDZ and [B] CAM-B3LYP/aug-cc-pVDZ, adopting the B3LYP-optimized geometries.

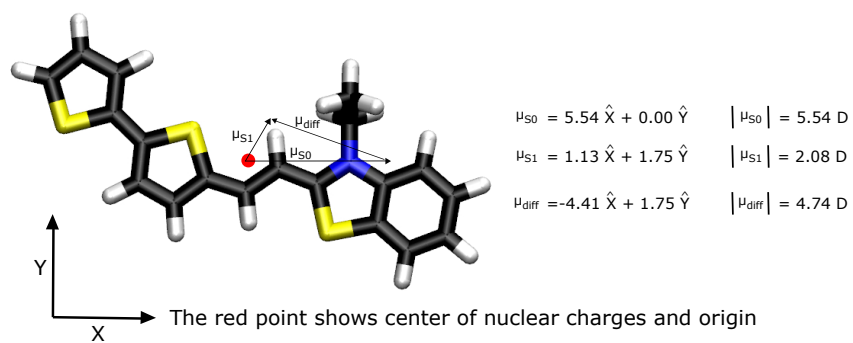


Figure S2: Calculated dipole moments of ground and excited state for *trans* conformer. Method: CAM-B3LYP(100%)/aug-cc-pVDZ (quadratic response theory, double residue) on ground state geometry optimized at B3LYP/aug-cc-pVDZ. Calculations are performed using Dalton. Symbols  $\mu_{S_0}$ ,  $\mu_{S_1}$ , and  $\mu_{\text{diff}}$  represent dipole moment of ground state, first excited state and difference between this two, respectively.

### 1.1.1 Vibrational structure and vibronic spectra

Several vibrational frequency calculations have been performed in order to aid in structural assignments at the harmonic level. The ground and excited state equilibrium structures for the *cis* and *trans* conformers were generated by TDDFT structure optimization in the first excited state using the long-range corrected and range-separated hybrid functional (CAM-B3LYP) and aug-cc-pVDZ basis set<sup>10</sup> in Gaussian package.<sup>1</sup>

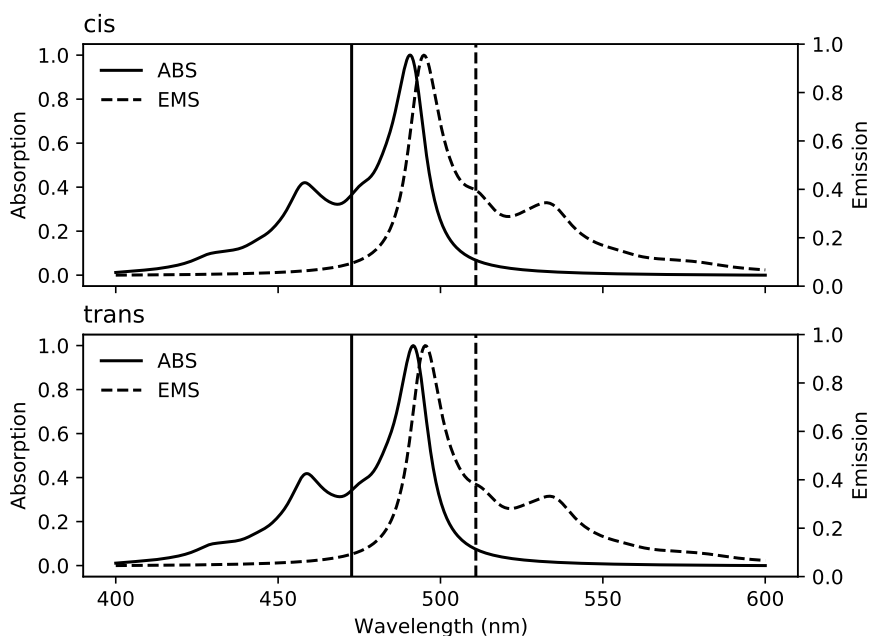


Figure S3: Computational vibrationaly resolved absorption and emission spectra for *cis* and *trans* bTVBT4 in vacuum (CAM-B3LYP/aug-cc-pVDZ). For the *cis* reference structure, we included three modes (with the frequencies 1317  $\text{cm}^{-1}$ , 1469  $\text{cm}^{-1}$ , and 1600  $\text{cm}^{-1}$ ) anharmonically in the absorption calculation and six modes (646  $\text{cm}^{-1}$ , 1246  $\text{cm}^{-1}$ , 1428  $\text{cm}^{-1}$ , 1458  $\text{cm}^{-1}$ , 1466  $\text{cm}^{-1}$ , and 1571  $\text{cm}^{-1}$ ) for the emission. For the *trans* reference structure, we accordingly treated three modes (i.e. 1323  $\text{cm}^{-1}$ , 1468  $\text{cm}^{-1}$ , and 1608  $\text{cm}^{-1}$ ) anharmonically in the absorption and seven modes (i.e., 680  $\text{cm}^{-1}$ , 1215  $\text{cm}^{-1}$ , 1241  $\text{cm}^{-1}$ , 1427  $\text{cm}^{-1}$ , 1459  $\text{cm}^{-1}$ , 1482  $\text{cm}^{-1}$ , and 1577  $\text{cm}^{-1}$ ) in the emission spectra calculation. All remaining vibrational modes were treated by the IMDHO model.



For generating the vibronic emission spectra, a hybrid scheme between the independent-mode displaced harmonic oscillator (IMDHO) model<sup>11,12</sup> and anharmonic treatment via vibrational configuration interaction (VCI) was employed. For the IMDHO we applied an in-house Python code and with half-widths at half-maximum (HWHM) of 200 cm<sup>-1</sup>. For the anharmonic Franck–Condon factor calculations, we employed MidasCpp.<sup>13</sup> For the calculation of Franck–Condon factors in MidasCpp, see also ref. 14. In the construction of the potential energy surfaces (PESs), we made use of a multi-state<sup>15</sup> adaptive approach for grid point selection.<sup>16</sup> For expansion of the PES, we applied polynomials up to 8th order. No mode–mode couplings have been accounted for in the PES. The vibrational self consistent field (VSCF) calculations underlying the VCI calculations were performed using a b-spline basis.<sup>17</sup> 30 of the VSCF one-mode functions were considered in the following VCI calculations, which accounted for configurations with a maximum of three simultaneously excited modes (VCI[3]).

### 1.1.2 Molecular mechanics force field parameters of bTVBT4

Force field (FF) parametrization was carried out in two parts. In the first part, initial parameters were generated from the General Amber Force Field (GAFF).<sup>18,19</sup> In the second part, equilibrium bond distances, angles, and important dihedral potential were corrected based on optimized ground state geometry and dihedral potential calculated from DFT method.

#### Generation of initial parameters

To start with, rudimentary FF parameters of bTVBT4 were obtained from GAFF, following the steps outlined in an Amber tutorial,<sup>20</sup> summarized below.

- Step 1: Geometry Optimization

The initial geometry of bTVBT4 molecule is optimized in the Gaussian program by employing the B3LYP functional and 6-31+G(d,p) Pople’s basis set.

- Step 2: Restrained Electrostatic Potential (RESP)<sup>21</sup> charge calculation and defining

atom types

In Amber FFs, all residues (building blocks) on which macromolecules are built are restricted to integer charges. Having integer charges on building blocks help in transferring electrostatic parameters to macromolecules. RESP charges assure that rotationally degenerate atoms, such as hydrogens of methyl and methylene have equivalent charges. Determining the RESP charges and atom type involves; A) Calculating the electrostatic potential (ESP) at HF/6-31G\* on an already optimized geometry in gas phase. B) Fitting charges at each nucleus by minimizing the sum squared error between calculated ESP from charges and ESP obtain from step 2A. This is achieved with the help of Antechamber (part of Amber tools).<sup>18,19</sup>

- Step 3: Parameter check and building a parameter library

Once the atom types are defined, a parameter library for bTVBT4 was generated using LEaP (part of Amber tools).

- Step 4: Format conversion

The topology obtained from the above procedure is in Amber format, and since we will be using Gromacs<sup>22-25</sup> for molecular dynamics (MD), the topology is converted from Amber to Gromacs format using the ACPYPE<sup>26</sup> Python module.

### **Correcting bond distances, bond angles and dihedral angles**

During step 3 described above, the parmchk program (part of Amber tools) has created suggestions for some dihedrals which was not found in the GAFF database. We are interested in determining the dynamics of bTVBT4 in different environments and calculating an absorption spectrum on frames sampled from the dynamic trajectory. Hence, it is important to accurately describe bond distances, bond angles, and dihedral angles so that absorption spectra calculated on DFT and Molecular Mechanics (MM) optimized geometry give same result. In order to obtain more suitable GAFF parameters and validate the FF, the following steps were taken:

- Step 1: Dihedral scanning

A relaxed dihedral scan for each of the dihedral angles defined in Fig. S1(a) was performed using B3LYP/6-31+G(d,p) as well as Molecular Mechanics (MM) with the GAFF. Comparing the two (DFT and MM) potential energy curves in Fig. S4 (illustrating the SCCS dihedral angle), we see that the initial topology that was generated is not an adequate description of SCCS potential. Due to the artificially high barrier represented in GAFF, an interconversion between the *cis* and *trans* conformation of bTVBT4 gets hindered during the MD simulation. Similarly, it is important to represent other important dihedral angles SCCC, NCCC, and CCCC accurately, in order to sample conformation at the 300 K.

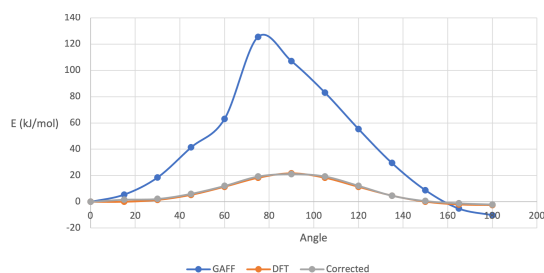


Figure S4: Dihedral potential of SCCS (marked in Fig. S1(a) of bTVBT4) from GAFF (blue), DFT (orange), and corrected FF (grey).

- Step 2: Modifying equilibrium bond distance and bond angle.

Bond distances and bond angles are crucial for absorption spectra, so all equilibrium bond distances and bond angles from GAFF were replaced with values obtained from DFT optimized geometries. All remaining parameters were kept the same as obtained from GAFF.

- Step 3: Excluding the energy contribution from dihedral of interest

For the required dihedral, first, a contribution of torsion energy is excluded from total

MM energy by setting appropriate torsion parameters to zero in a FF file. Using this modified topology, MM based relaxed dihedral scan (called MM\_zero) is obtained. Next, this MM\_zero energy profile is subtracted from DFT energy profile to get a missing energy contribution which will be added to torsion parameters of dihedral of interest. For the SCCS dihedral, the difference between DFT and MM\_zero curve is shown in Fig. S5 with red colour. Next task is to fit this curve with the Ryckaert-Bellemans function Eq. (1).

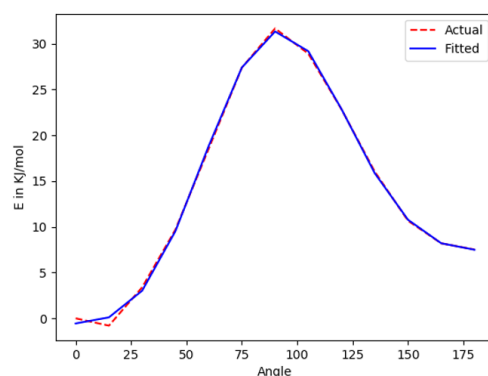


Figure S5: Curves shows the difference between potentials for DFT and MM\_zero (where all parameters that define SCCS are substituted to zero).

- Step 4: Fitting to DFT potential

The curve obtained by taking the difference between DFT and MM\_zero potential is fitted with Ryckaert-Bellemans function (Eq. (1)). This function has previously been used to fit dihedrals involved in bi-thiophene based ligands.<sup>27</sup> Curve fitting was done using the Scipy<sup>28</sup> Python module. The fitted curve is shown in blue in Fig. S5.

$$V(\phi) = \sum_{n=0}^5 C_n \cos^n(\phi - 180^\circ) \quad (1)$$

- Step 5: Adding corrected parameters and scanning the dihedral angle

Add corrected parameters for dihedral which was fitted and a relaxed dihedral scan

was performed using this modified topology. The result obtained in this step for the SCCS dihedral is shown in Fig. S4 (grey curve). The same procedure was employed for correcting all dihedrals (SCCC, NCCC, and CCCC, illustrated in Fig. S1a) iteratively. Note: So far in the process of fitting dihedral potential, except for the step 2 (correcting the bonds and angles based on DFT optimized geometry) we followed the procedure which has been previously described in Ref. 29.

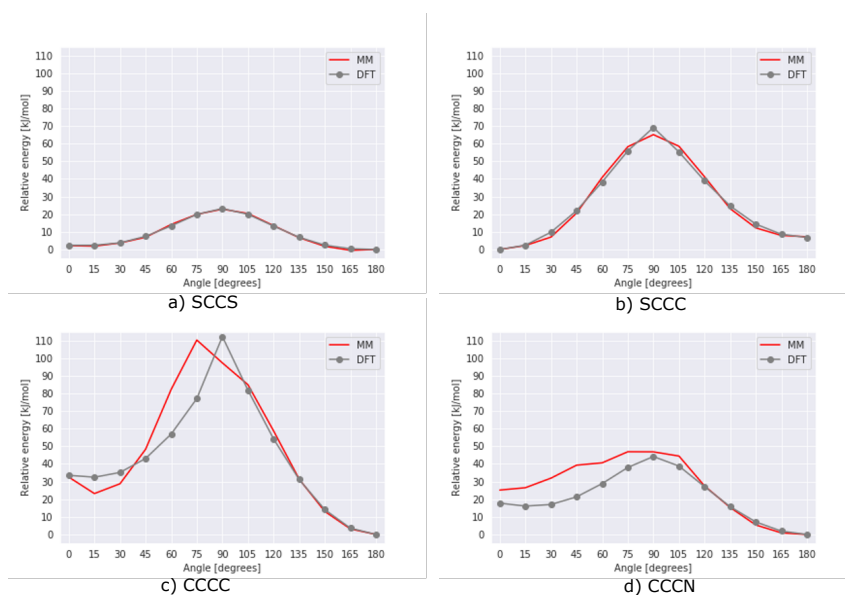


Figure S6: Fitting of important dihedrals (marked in Fig. S1(a) of bTVBT4 to the DFT potential. The dihedral scans are performed at B3LYP/6-31+G(d,p).

- Step 6: Validation

After correcting all important dihedrals, the final dihedral scan is shown in Fig. S6. In all the cases, the dihedral potential curve is in good agreement with DFT. For the energy barrier above 40 kJ/mol, the dihedral potential of the CCCN dihedral deviate from DFT potential, however this will not be the problem in current study as simulations are performed at room temperature (300 K). Next, if we compare the

relative energies of *cis* and *trans* obtained using different methods (DFT and the re-parametrized FF), it is in good agreement (Table S1). This can also be seen from Boltzmann distribution shown in Table S2.

**Table S1:** Comparison of QM vs. MM ground state energy difference for the most stable (*cis/trans*) bTVBT4 conformers.

Method	$\Delta E_{cis-trans}$ kJ/mol (kcal/mol)
QM (B3LYP/6-31+G(d,p))	2.47 (0.59)
QM (B3LYP/aug-cc-pVDZ)	1.97 (0.47)
MM (Modified GAFF)	2.22 (0.53).

**Table S2:** Comparison of QM vs. MM Boltzmann distribution for ground-state conformations.

Method	$\Delta E_{cis-trans}$ kcal/mol	$P_{cis}$	$P_{trans}$
QM (B3LYP/6-31+G(d,p))	0.59	0.27	0.73
MM (Modified GAFF)	0.53	0.29	0.71

- Step 7: Absorption spectrum test

Absorption spectra calculated on *cis/trans* geometries optimized using the re-parametrized FF are presented in Table S3, along with transition wavelengths computed at different levels of DFT. The transition wavelengths calculated using MM optimized geometries were in very good agreement with DFT values.

**Table S3:** Comparison of QM vs MM ground state absorption spectra of bTVBT4 conformations in vacuum, calculated at TDDFT/CAM-B3LYP(100%)/aug-cc-pVDZ using Dalton program.

Geom. opt. method	<i>cis</i> nm (eV)	<i>trans</i> nm (eV)
QM B3LYP/6-31+G(d,p)	471.7 (2.628)	472.3 (2.625)
QM B3LYP/aug-cc-pVDZ	473.4 (2.619)	472.9 (2.622)
MM (Modified GAFF)	471.3 (2.631)	472.4 (2.625)

### 1.1.3 Functional benchmark for absorption spectra

Generally, changing the DFT functional induce a global shift in transition energies and does not influence the absorption spectrum profile. In order to have computational transition en-

ergies close to experimental values, a choice of functional was based on benchmark calculation shown in Table S4.

**Table S4:** Functional benchmark for absorption spectra, calculated on B3LYP/6-31+G(d,p) optimized geometry of *trans* conformer.

Functional	<i>trans</i> nm (eV)
CAM-B3LYP(100%)/aug-cc-pVDZ	472 (2.625)
CAM-B3LYP/aug-cc-pVDZ	490 (2.532)
B3LYP/aug-cc-pVDZ	518 (2.395).

#### 1.1.4 Molecular dynamics simulation

All classical MD simulations were performed using GROMACS (version 19.3)<sup>22–25</sup> with the Amber ff14SB force field for the tau protein,<sup>30</sup> TIP3P for water solvents,<sup>31</sup> and re-parametrized GAFF for bTVBT4. In particular, all MD processes (energy minimization, equilibration, and production) were performed for different types of systems: bTVBT4 in a water solvent, periodic tau fibril in water solvent (with and without bTVBT4), bTVBT4 in small tau fibril model. The solvation box was neutralized with counter-ions ( $\text{Cl}^-$ ) for each of these models and described with the Amber forcefield (ff14SB). The protonation state of all amino acid residues of the tau protein are represented at the physiological pH of 7.4.

When the simulation is performed in the *NPT* ensemble (equilibration step), the reference temperature of 300 K was kept constant using the velocity rescaling weak coupling scheme with a coupling constant of  $\tau = 0.2$  ps. A Berendsen barostat with a coupling constant of 1 ps was employed to maintain the pressure at 1 atm. In the case of bTVBT4 in tau protein, Berendsen barostat was used under semi-isotropic conditions. Whereas for bTVBT4 in water, the isotropic condition was applied. Periodic boundary conditions were applied to all three directions of the simulated box. Electrostatic interactions were simulated with the Particle Mesh Ewald (PME)<sup>32</sup> approach using a long-range cutoff of 1.5 nm. The cutoff distance of Lennard-Jones (LJ) interactions was also equal to 1.5 nm. The MD simulation time step was 2 fs with a pair-list update period of 10 steps. All H-bond lengths were kept

constant using the LINCS routine.<sup>33</sup>

### 1.1.5 Molecular dynamics simulation of bTVTB4 in vacuum

The ligand bTVBT4 (with one Cl<sup>-</sup> counter ion) was first simulated in vacuum, and *cis* and *trans* conformations were sampled during a 4  $\mu$ s long MD simulation. Percentages of *cis* and *trans* conformations of bTVBT4 are presented in Table S5. Initially, the Cl<sup>-</sup> ion was positioned at 24 nm from bTVBT4. It moves closer ( $< 4$  Å) to bTVBT4 during the trajectory, and starts interacting with bTVBT4 from about 98 ns and stays close to the ligand during the remainder of the simulation. Due to this fact, the expected percentage of *cis* and *trans* conformations from Boltzmann distribution is not achieved.

**Table S5:** Percentage of *cis* and *trans* conformations from 4  $\mu$ s long MD simulation of bTVBT4 and Cl<sup>-</sup> ions in vacuum.

	<i>cis</i>	<i>trans</i>
Percentage from MD simulation	64	36
Percentage from Boltzmann distribution	71	29

### 1.1.6 Molecular dynamics simulation of bTVTB4 in water

One bTVBT4 molecule in *trans* conformer was solvated by a water solvent box of  $5.2 \times 5.2 \times 5.2$  nm<sup>3</sup>. One Cl<sup>-</sup> ion was added to neutralize the system. Next, the equilibration step was performed with the first energy minimization of the whole system, followed by a short 100 ps *NPT* simulation to stabilize the pressure of the system. Subsequently, we propagated the system in an *NVT* ensemble for the production run with accumulated time of 4  $\mu$ s from 4 independent simulations.

The converged dihedral distribution over snapshots extracted from 4  $\mu$ s data is shown in Fig. S7. The time step between each saved snapshot was 10 ps. The calculated percentage of *cis* and *trans* conformers of bTVBT4 in water are 50.5 and 49.5, respectively. Although, *trans* conformation that is the most populated in vacuum, the *cis* conformer is estimated



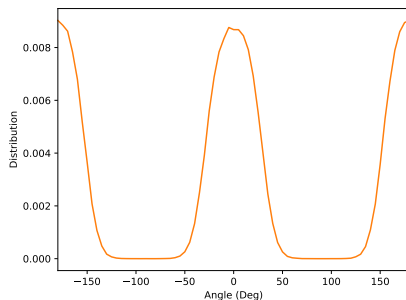


Figure S7: Dihedral SCCS (marked in Fig. S1(a) of bTVBT4) distribution in 4  $\mu$ s long MD of bTVBT4 in water.

**Table S6:** Percentage of *cis* and *trans* conformations from a 4  $\mu$ s MD simulation of bTVBT4 and Cl<sup>-</sup> in water. LJ and Coulomb interaction energies of bTVBT4 with water are given in kJ/mol.

	<i>cis</i>	<i>trans</i>
Percentage	50.5	49.5
LJ	$-135.9 \pm 10$	$-136.3 \pm 10$
Coulomb	$-68.3 \pm 14$	$-62.1 \pm 4$

to be equally populated in water due to an increase in interaction energy with water. The decomposition of interaction energy shows that the LJ interaction energy is approximately equal for both the conformers, as presented in Table S6. However, the Coulomb interaction of *cis* conformer in water is more stable than *trans* conformer by 6.2 kJ/mol. This increase in Coulomb interaction energy is attributed to interaction between thiophene moiety of *cis* and water.

### 1.1.7 Polarizable embedding and spectra in solution

In order to include the effect of the water environment while calculating the spectra of bTVBT4, we employed the method of Polarizable Embedding (PE)<sup>34,35</sup> on the snapshots extracted from the MD simulations. PE is based on a hybrid quantum-classical computational approach in which the total system is split into three regions: core, polarizable, and

non-polarizable. The core region having only a bTVBT4 molecule is calculated in the quantum chemical description. The polarizable region is described with Ahlström charges and isotropic polarizabilities,<sup>36</sup> while the non-polarizable region is described with just TIP3P charges. The shell thickness of 15 Å polarizable and 5 Å non-polarizable regions are considered to be reasonable to account for the effect of water. The Fig. S8 shows the test calculation of transition wavelength as a function of shell thickness of polarizable (red), non-polarizable (blue), and combination of both (green).

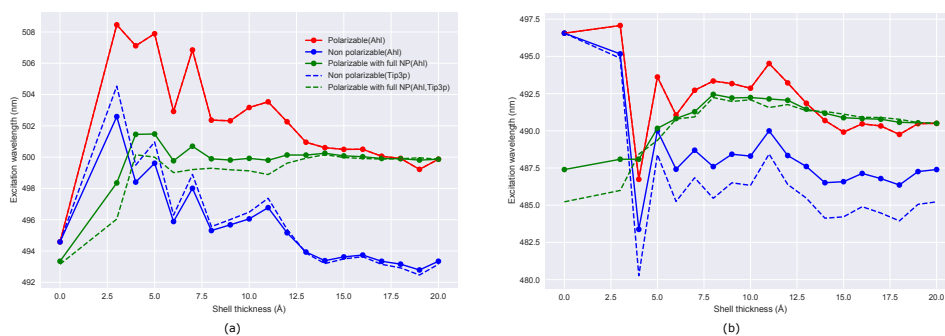


Figure S8: Transition wavelength of bTVBT4 as a function of shell thickness of polarizable (red), non-polarizable (blue), and polarizable with full NP (green) for (a) trans, and (b) cis snapshot extracted randomly from the simulation trajectory. The bold lines indicate when the non-polarizable region is represented by Ahlström charges and the dashed lines when the TIP3P charges are used. The absorption spectra calculation are performed in CAM-B3LYP/aug-cc-pVDZ.

From the MD simulation, a total of 200 number of snapshots were extracted of which in 100 frames bTVBT4 was in *cis* conformation and for another 100, in *trans* (respecting the *cis:trans* ratio found in the MD simulation, see Fig. S7), while ensuring converge spectral profile. The time step between two consecutive frames was at least 10 ps. For each snapshot, single point absorption spectra calculation was performed where the bTVBT4 was described in a QM core, all atoms within 15 Å of bTVBT4 with polarizable and following 5 Å shell in the non-polarizable region. The calculations were carried out for the ten lowest excited states at the CAM-B3LYP(100%) and aug-cc-pVDZ basis set using the Dalton program.

Subsequently, the calculated absorption spectrum on each snapshot was Gaussian convoluted with standard deviation of 0.15 eV and also with line broadening that was obtained from vibrationally-resolved electronic spectra of bTVBT4 in vacuum. The final spectral profile was obtained by taking an average over all the snapshots under consideration see in article Fig. 2(b). For the set of 100 *trans* conformer, average absorption maxima is at 480 nm (2.58 eV). Whereas for the set of 100 *cis* conformer, absorption maxima is at 485 nm (2.56 eV), see Table S7.

**Table S7:** Calculated transition wavelength averaged over 100 *cis/trans* snapshots for bTVBT4 in water.

	<i>trans</i>	<i>cis</i>	average
Transition wavelength nm(eV)	480 (2.58)	485 (2.56)	482.5 (2.57)

One snapshot from the set of 100 *trans* frames that had a peak at the average maxima at around 480 nm was selected for further studies on the absorption spectra. On this particular snapshot, the QM region was increased to also encompass 5 water molecules in addition to the bTVBT4 molecule. The effect of this was only a negligible shift in the absorption spectrum (Table S8). However, increasing QM region further to accommodate 14 water molecules, the absorption maxima shifted by 5 nm from 480 nm (average value) towards the experimental value (470 nm). Nevertheless, this improvement comes at a large computational cost and therefore, we decided against including water molecules in the QM region.

**Table S8:** Further improvement in absorption spectra of bTVBT4 in water, by increasing the QM region. One *trans* snapshot having absorption maxima at 480 nm was selected for test calculation

Method	Transition wavelength nm (eV)
QM(bTVBT4)	480 (2.58)
QM (bTVBT4) + PE	480 (2.58)
QM (bTVBT4 + 5 H2O) + PE	479 (2.59)
QM (bTVBT4 + 14 H2O) + PE	475 (2.61)
Experimental	470 (2.64)

## 1.2 Modeling bTVBT4 biomarker interactions with tau fibril

### 1.2.1 Tau fibril in cryo-EM structure

Formation of amyloid and tau protein fibrils are the main characteristics of Alzheimer's disease. Fitzpatrick *et al.*<sup>37</sup> deduced the cryo-EM structure of tau protein. It was found to exist in two ultrastructural polymorphs: paired helical filaments (PHFs) and straight filaments (SFs). They differ in an inter-protofilament arrangement. The core of the structure (protofilament) is made up of 73 amino acids. In the case of PHFs, the two protofilaments are bound by  $2_1$  screw symmetry to form a monomer. These monomers stack on each other to develop a long fibril structure. Consecutive monomers along the protofilament have an approximate twist of 1 degree and distance 4.7 Å.

### 1.2.2 Creating a tau fibril models

From the PDB ID 5O3l, first, coordinates of one protofilament out of ten were extracted. Next, by taking the  $2_1$  screw symmetry of the first protofilament, another protofilament was generated to form a monomer. The consecutive monomers were obtained by rotating a previous monomer by 0.97 degrees and translating it by 5 Å. This operation was continued until the last monomer had a twist of 180 degrees with respect to the first monomer. It is important to ensure a perfect fit between the last monomer and the first monomer of an adjacent periodic image of the simulation box. In this way a long fibril was modeled, which approximately resemble the experimentally found tau fibril while decreasing the computational cost by half. Employing the same strategy, tau models having adjacent monomer twist of 0.8, 0.9, 1.0, and 1.2 were also constructed to find a fibril model with least stress induced due to forced periodicity.

### 1.2.3 MD simulation of full tau fibril

For each of the tau model created, simulated annealing was performed to slowly warm the system under *NPT* conditions. The heating of system from 150 K to 300 K was achieved with 14 annealing points, at the temperature step size of 25 K.

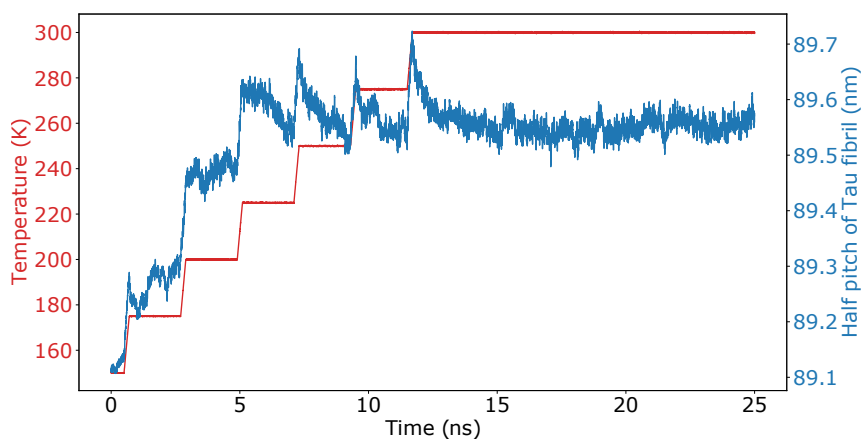


Figure S9: Length of half pitch tau fibril (blue) as a function of temperature (red) during 25 ns simulated annealing MD.

The temperature of the system was gradually increased every 200 ps, and at each step, it was simulated for 2 ns except for the last step of 300 K where it was simulated for 13 ns. It was observed that tau fibril generated with the adjacent monomer twist of 0.97 degrees had fewer kinks compared to the other tau models with different adjacent monomer twists (data not shown). The dynamics of the total fibril length along its axis is shown in the Fig. S9 for the tau model with monomer twist of 0.97 degrees. Simulation at the final step of 300 K shows that we have reached an equilibrium structure from 14 ns onward. In the equilibrium structure of the tau model, the average distance between two adjacent monomers was decreased from 5 Å to 4.8 Å.

#### 1.2.4 Molecular dynamics of full periodic model of tau with 60 bTVBT4 molecules

Starting from the thermally stable structure of the tau fibril model, two kinds of system were built to increase the sampling in MD simulation. In the first system, the simulation box contained the tau fibril with 60 bTVBT4 molecules inside the tau cavity (Fig. S10a), and in the second system, 60 bTVBT4 outside the tau cavity (Fig. S10b). Each ligand was placed at random position and in random orientation.

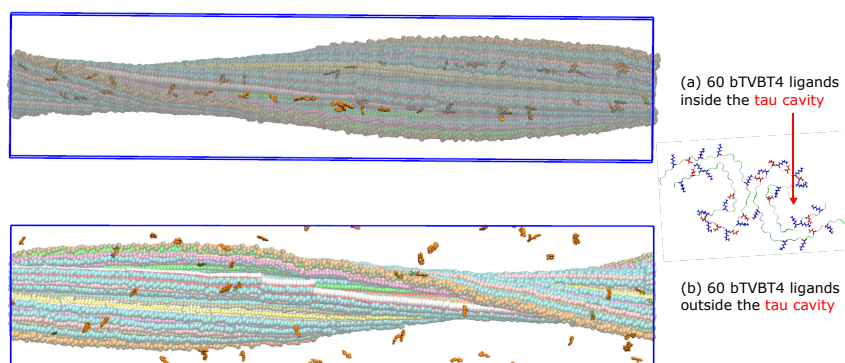


Figure S10: (a) 60 bTVBT4 ligands placed inside the cavity of tau fibril. (b) 60 bTVBT4 ligands placed outside the cavity of the tau fibril.

Initially, in the process of equilibration, both kinds of systems were energy minimized, followed by short 100 ps *NVT* simulation with position restrain on bTVBT4 and tau fibril to stabilize the temperature. Next 100 ps short *NPT* simulation to stabilize the pressure. Finally, the MD simulation production run was carried out in *NVT* ensemble, maintaining the temperature at 300 K. For the 60 bTVBT4 molecules inside tau cavity case, five independent simulations were simulated for 40 ns each, whereas for the 60 bTVBT4 molecules outside tau cavity case, four independent simulations were setup and simulated for approximately 35 ns each.

Binding mode code assignment for Site A :

Site A was distinguished as the strongest binding site at which bTVBT4 binds in various modes (see Fig. 3). These modes were characterized with respect to position of the ethyl

group, bi-thiophene group of bTVBT4 and amino acid residues Ile360, and His362 of the tau fibril that bTVBT4 interact with.

Modes are defined by four letters assigned based on the following criteria:

Fix the position of tau such that Ile360 is on top and His362 on the bottom side.

- Based on relative position of ethyl group with Ile360 and His362 assign "U" (ethyl pointing towards Ile360) or "D" (ethyl pointing towards His362)
- ethyl group is at left "L" or right "R" (keep position of Ile360 and His362 fixed as mention above)
- Bi-thiophene moiety in *cis* "C" or *trans* "T" conformation.
- CH<sub>3</sub> group on ethyl, inside "I" (pointing towards tau protein) or outside "O" (away from tau protein)

Considering all possible combination, in total there are 16 modes. The numeric and alphabetic codes for all 16 modes are defined in the Table S10. Frequency of interconversion between "I" , "O" and "C" , "T" is grater as compared to "U" , "D" and "L" , "R". Hence, when only the first two letters of alphabetic code are considered, we define it as major mode and when all four letters are considered we defined it as minor modes (see Fig. S11).

**Table S9:** Lennard-Jones (LJ) and Coulombic interaction energies (in kJ/mol) between bTVBT4 and residues Ile360, Thr361, and His632 that make up binding site A, and Arg349, Val350, and Gln351 that make up binding site B. Residue Arg349 has a charge of +*e*, whereas others are charge neutral.

Site	Type				Total
A		Ile360	Thr361	His362	
	LJ	$-50.46 \pm 7$	$-4.78 \pm 2$	$-36.67 \pm 8$	$-91.9 \pm 9$
	Coulomb	$4.82 \pm 1$	$-1.76 \pm 2$	$-24.91 \pm 14$	$-21.85 \pm 14$
B		Arg349	Val350	Gln351	
	LJ	$-73.23 \pm 13$	$-13.01 \pm 4.0$	$-31.96 \pm 11$	$-118.21 \pm 14$
	Coulomb	$5.60 \pm 12$	$-7.84 \pm 4$	$1.79 \pm 9$	$-0.45 \pm 16$

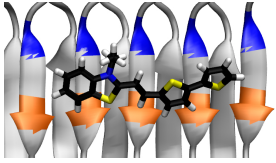
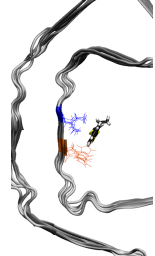

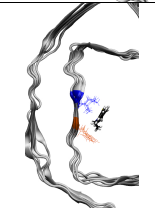
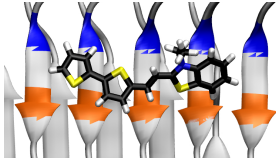
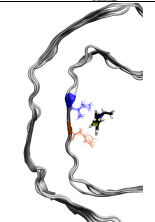
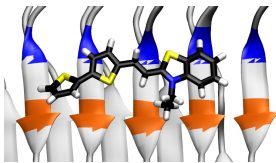
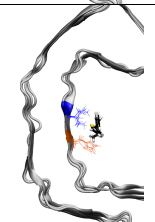
Code	Minor mode	Major mode	Front view	Top view
12	ULTO	UL		
4	DLTO	DL		
16	URTO	UR		
7	DRTI	DR		

Figure S11: Examples of naming convention. The alphabetic code up/down (U/D) and left/right (L/R) are based on relative position of ethyl group with respect to residues Ile360 (marked in blue) and His362 (marked in orange). The *cis/trans* (C/T) is with respect to the dihedral between bi-thiophene moiety and inside/outside (I/O) based on the dihedral rotations around the thiazole-ethyl bond.

Average total interaction energy of bTVBT4 with site A:

From an MD simulation of the full tau model with 60 bTVBT4 molecules, two binding sites



were identified A (Ile360, Thr361, His362) and B (Arg349, Val350, Gln351) as seen in article Fig. 3. The LJ and Coulombic interaction energies of bTVBT4 in the binding site A and site B are shown in the Table S9. These energies were calculated only for those bTVBT4, which bound to the site (A or B) continuously for 20 ns and had LJ < -70 kJ/mol. The result shows that the ligand binding to site A is driven by Coulombic interaction while to site B by LJ interaction.

**Table S10:** Average of interaction energy (LJ + Coulomb) and number of bTVBT4 molecules at each mode. Energies are given in kJ/mol.

code	Mode	Number of molecules	Average energy	Standard deviation
1	DLCI	8	-160.5	19
2	DLCO	7	-157.3	17
3	DLTI	8	-157.7	18
4	DLTO	16	-151.5	17
5	DRCI	14	-160.5	21
6	DRCO	9	-156.0	19
7	DRTI	23	-154.4	18
8	DRTO	12	-150.4	17
9	ULCI	6	-147.5	18
10	ULCO	5	-151.3	16
11	ULTI	9	-144.8	15
12	ULTO	10	-150.8	20
13	URCI	2	-142.9	17
14	URCO	3	-153.0	16
15	URTI	10	-156.4	18
16	URTO	19	-147.0	17

The majority of the bTVBT4 population is found at site A at which it binds in 16 different minor modes (or 4 major modes) as defined in Table S10. The average total interaction energy (LJ+Coulomb) between bTVBT4 molecules and tau fibril calculated for each of the minor modes from the five independent MD simulations are shown in Table S10. Often interconversion from one minor mode to another is observed in the binding pocket during the trajectory, mainly between minor modes of type  $XXTX \leftrightarrow XXCX$  and  $XXXI \leftrightarrow XXXO$ , where X represent any other letter codes for minor modes as defined in Table S10. The Fig. S12 shows all the interconversion between the minor modes. The average of the

total interaction energy over all minor modes is calculated to be  $-153 \pm 19$  kJ/mol. The bTVBT4 interaction with the site A residues account for 75% of the average total interaction energy when it interacts with the whole tau, as shown in Table S11.

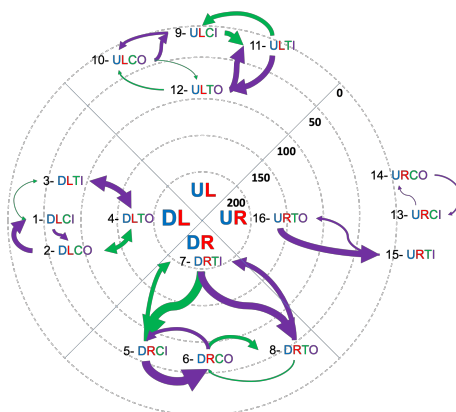


Figure S12: Exchange between minor modes at site A in the system of 60 bTVBT4 in the cavity of full tau model. Radius represent the time spent (ns) in each mode (accumulated time for all molecules in the mode), the thickness indicates the number of exchanges (largest 14).

**Table S11:** Comparison of bTVBT4 average total interaction energy (in kJ/mol) with protein and site A residues.

	with site A residues	with protein
bTVBT4 interaction	$-114 \pm 17$	$-153 \pm 19$

### 1.2.5 Smaller binding site model

The smaller binding site models were generated from the system of 60 bTVBT4 molecules in the cavity of the tau fibril by cropping out the four major binding modes. Each smaller binding site model consist of one bTVBT4 molecule and ten protein chains at the side where the ligand was bound (see Fig. S13). Next, for each model, a simulation box of size 15 nm x 9 nm x 9 nm was created and filled with water molecules and counter anions  $\text{Cl}^-$ . Further, after the energy minimization and equilibration process of a system, the MD production

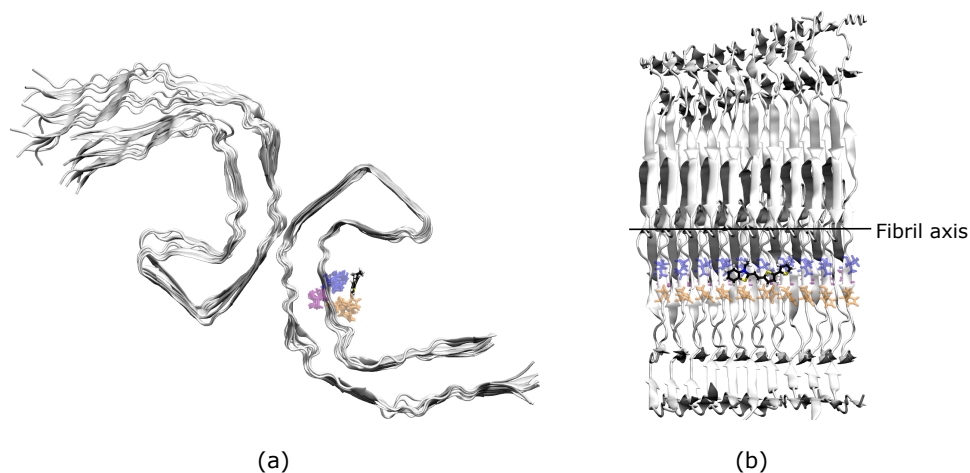


Figure S13: Side (a) and top (b) view of the small binding site tau model cropped from the full periodic tau model.

run of 250 ns in *NVT* was performed on two parallel trajectories for each major mode (i.e., 500 ns time for each major mode). During the whole procedure, position restraints on the outer-most tau protein chains were applied to maintain adjacent chains distance. The results obtained from this short model of binding modes are summarized in Fig. S14 and Fig. S15.

	DL				DR				UL				UR							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16				
Trj 1	1	21	5	1	0	5	1	0	1	0	9	0	2	0	0	13	0	0	0	0
	2	7	64	0	0	6	1	25	0	38	10	1	40	0	3	14	0	0	0	0
	3	0	0	2	5	7	0	0	1	2	11	1	0	48	5	15	0	0	3	5
	4	0	2	4	40	8	0	39	1	171	12	0	2	6	63	16	0	0	6	106
Trj 2	1	60	1	1	0	5	1	0	1	0	9	3	0	0	0	13	0	0	0	0
	2	1	6	0	0	6	0	31	0	32	10	0	0	0	1	14	0	0	0	0
	3	2	0	24	1	7	2	0	37	2	11	1	0	37	4	15	0	0	6	10
	4	0	0	2	22	8	0	32	2	103	12	0	1	5	115	16	0	0	11	133

Figure S14: Time spent (ns) in each minor mode (diagonal) and number of exchanges between the minor modes (non-diagonal) from row to column. Characterized based on four major modes. Data obtained from two trajectories simulated for 250 ns for each major mode.

Initial Mode	Trajectories 1st time											Avg. of 4 modes
	Modes explored	Time (ns)	Coul avg	Coul std	Coul var	LJ avg	LJ std	LJ var	Tot (LJ+Coul)	Tot std	Tot var	
DLTO (4)	1	21	-14.3	6.1	37.6	-97.4	7.6	58.3	-111.7	9.8	95.9	-112.5 ± 10
	2	64	-13.9	6.2	38.0	-96.3	8.0	64.4	-110.2	10.1	102.4	
	3	2	-12.5	6.0	36.3	-98.5	8.3	68.8	-111.0	10.3	105.1	
	4	40	-17.4	7.2	52.3	-99.8	8.0	64.0	-117.1	10.8	116.3	
DRTI (7)	5	1	-10.7	3.9	14.9	-103.3	8.8	78.2	-114.1	9.7	93.1	-117.8 ± 11
	6	25	-14.8	6.2	38.9	-104.2	8.4	71.0	-119.0	10.5	109.8	
	7	1	-8.4	2.7	7.3	-96.1	9.1	83.1	-104.5	9.5	90.4	
	8	171	-23.5	9.8	95.8	-110.1	8.8	77.9	-133.6	13.2	173.7	
ULTO (12)	9	~0	-28.1	8.7	76.5	-81.6	8.6	73.6	-109.7	12.3	150.2	-121.3 ± 13
	10	40	-31.5	11.5	131.3	-97.0	8.0	63.7	-128.5	14.0	195.0	
	11	48	-25.5	9.6	91.9	-96.5	8.9	79.7	-121.9	13.1	171.6	
	12	63	-30.1	8.7	76.5	-95.3	10.2	104.3	-125.4	13.4	180.8	
URTO (16)	13	0										-132.0 ± 9
	14	0										
	15	3	-28.7	9.3	86.3	-102.9	8.0	64.5	-131.7	12.3	150.8	
	16	106	-28.6	9.1	83.3	-103.8	7.5	56.6	-132.4	11.8	139.9	

(a)

Initial Mode	Trajectories 2nd time											Avg. of 4 modes
	Modes explored	Time (ns)	Coul avg	Coul std	Coul var	LJ avg	LJ std	LJ var	Tot (LJ+Coul)	Tot std	Tot var	
DLTO (4)	1	60	-12.3	5.0	25.1	-97.6	7.4	55.1	-109.9	9.0	80.2	-111.9 ± 10
	2	6	-11.6	5.2	27.1	-99.3	7.6	58.4	-110.9	9.2	85.5	
	3	24	-11.6	5.0	25.1	-96.0	7.5	56.3	-107.5	9.0	81.4	
	4	22	-20.5	7.9	61.8	-98.7	7.6	57.5	-119.2	10.9	119.2	
DRTI (7)	5	1	-11.3	3.9	15.4	-96.6	8.2	67.1	-107.9	9.1	82.5	-116.8 ± 11
	6	31	-15.0	6.0	36.2	-99.3	7.6	57.7	-114.3	9.7	93.9	
	7	37	-15.3	7.5	55.6	-101.4	8.5	72.2	-116.7	11.3	127.8	
	8	103	-19.3	9.6	93.1	-109.3	9.4	89.2	-128.6	13.5	182.3	
ULTO (12)	9	3	-28.1	8.7	76.5	-81.6	8.6	73.6	-109.7	12.3	150.2	-119.2 ± 11
	10	~0	-31.5	11.5	131.3	-97.0	8.0	63.7	-128.5	14.0	195.0	
	11	37	-25.5	9.6	91.9	-96.5	8.9	79.7	-121.9	13.1	171.6	
	12	115	-30.1	8.7	76.5	-95.3	10.2	104.3	-125.4	13.4	180.8	
URTO (16)	13	0										-130.9 ± 9
	14	0										
	15	6	-27.9	8.9	76.9	-103.1	8.2	66.7	-131.0	12.1	145.6	
	16	133	-29.3	9.6	91.9	-101.5	7.7	58.7	-130.8	12.3	150.5	

(b)

Figure S15: Statistics drawn from the two parallel trajectories (a and b) for each major mode (smaller binding site model) simulated for 250 ns.

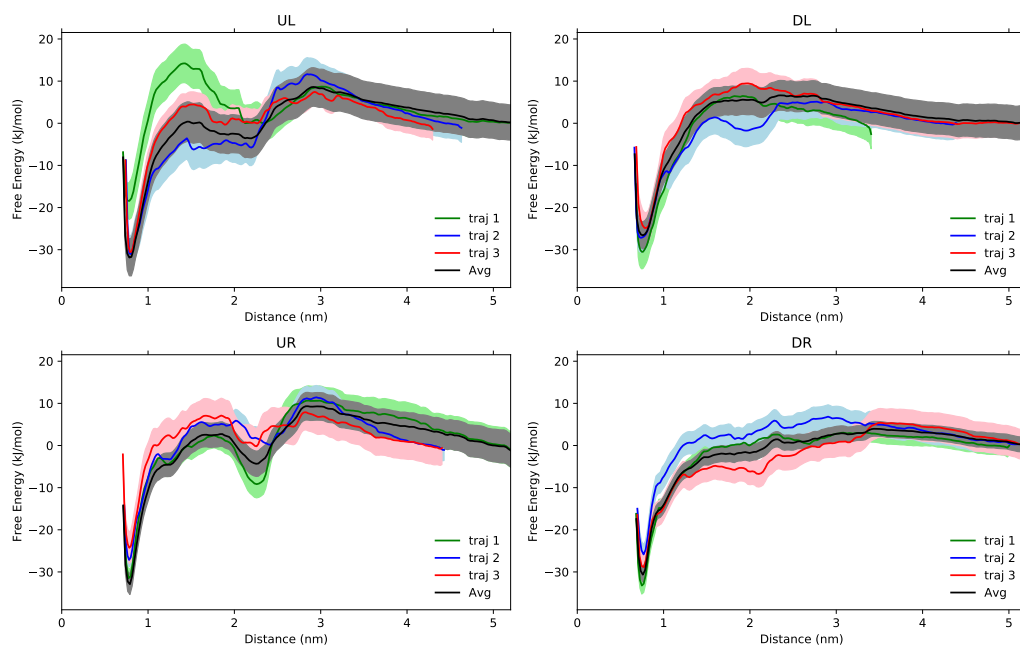


Figure S16: The PMF calculation for three pulling trajectory (green, red, blue) and its average (black) for each major modes.

### 1.2.6 PMF calculation details

We performed potential of mean force (PMF) calculations to obtain the height of the atomistic free-energy barrier for the bTVBT4 to bind at site A. It was achieved by using an umbrella sampling technique and a weighted histogram analysis method. Starting from the last snapshot of each major mode (short models) equilibrium MD simulation, bTVBT4 has pulled away from the binding site A to outwards. The center of mass (COM) pulling was applied between two groups Thr360 (middle residue at site A) and bTVBT4. The pulling rate of 0.001 nm/ps combined with a spring constant of 3000 kJ/mol was used. The ligand was pulled out from major modes of binding site A for about 5 nm. From these pulling trajectories, snapshots were taken to generate the starting configurations for the umbrella sampling windows. In each window, short 100 ps *NVT* followed by 2 ns *NVT* simulation was performed. Finally, analysis of results obtained from umbrella sampling was performed with the weighted histogram analysis method (WHAM).<sup>38,39</sup> Note that, for each major mode, three pulling simulations followed by umbrella sampling were conducted to get the average PMF curve finally (Fig. S16).

### 1.2.7 Absorption spectra calculation of bTVBT4 at the binding site A

To calculate absorption spectra, 300 uncorrelated snapshots were chosen based on the percentage of time spent in each minor mode and the major mode's Boltzmann percentage. Boltzmann percentage for major modes was calculated based on their respective binding energies obtained from PMF. Table S12 shows the number of snapshots chosen for each minor mode.

For each snapshot, absorption spectra were calculated by employing the polarizable embedding method. The total system was split into two regions: core and polarizable region. The core region having only a bTVBT4 molecule is treated in the quantum chemical description. The polar region consists of water molecules, counter anions ( $\text{Cl}^-$ ), and amino acid residues within 20 Å of bTVBT4. For water, Ahlström charges and isotropic polarizabilities

**Table S12:** Number of snapshots adopted for each minor modes for the calculation of the absorption spectrum of bTVBT4 in tau environment.

Major mode	Minor mode	Number of snapshots
DL	1	3
	2	3
	3	1
	4	3
DR	5	0
	6	9
	7	6
	8	42
UL	9	1
	10	12
	11	25
UR	12	53
	13	0
	14	0
	15	5
	16	137
total		300

were used. Standard charge and isotropic polarizability for  $\text{Cl}^-$  atoms were taken from the library in the PyFrame module.<sup>40</sup> For the amino acid residues of tau protein, the average of charges and isotropic polarizabilities over 10 tau protofilament is used. The charges and isotropic polarizabilities were determined using the LoProp program.<sup>41,42</sup> The absorption calculations were carried out for the ten lowest excited states at the CAM-B3LYP(100%) and aug-cc-pVDZ basis set using the Dalton program. Subsequently, for each snapshot, the calculated absorption spectrum was Gaussian convoluted with standard deviation of 0.15 eV. The final spectral profile was obtained by taking an average over all the snapshots under consideration.

### 1.2.8 Planarity calculation

The planarity parameter used for the assessment of the molecular structure of bTVBT4 is adopted from Ref. 43 and reads

$$P = \sum_{i=1}^N \frac{||\theta_i| - 90|}{90}, \quad (2)$$

where  $\theta_i$  are the dihedral angles in the molecule, and  $N$  is the number of dihedral angles for which the planarity is calculated. For the bTVBT4 case, the planarity was calculated over four dihedrals (marked in Fig. S1 (a)). Fig. S17 shows the correlation between planarity and transition energy of the first excited state, calculated in water (using 200 snapshots) and tau (using 300 snapshots) environment. The average value of planarity for bTVBT4 in water is  $3.43 \pm 0.20$  and in tau is  $3.49 \pm 0.18$ .

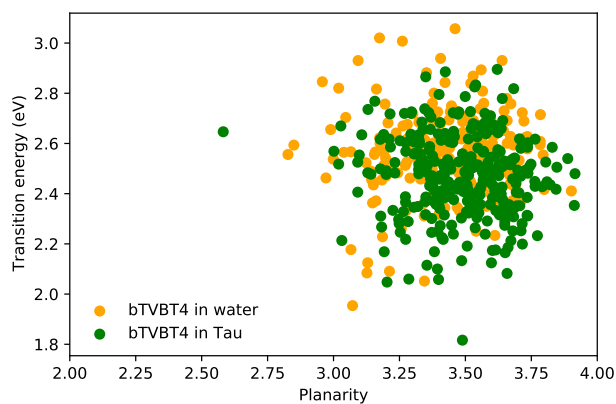


Figure S17: Correlation between planarity (defined in Eq. 2) vs. transition energy for the first excited state of bTVBT4. In total, 200 points are shown for bTVBT4 in water (orange), and 300 points shown for bTVBT4 in the tau fibril environment of binding site A (green), corresponding to the snapshots used in spectrum calculation. The average value of planarity for bTVBT4 in water is  $3.43 \pm 0.20$  and in tau is  $3.49 \pm 0.18$ .



### 1.2.9 Qualitative picture of interactions between bTVBT4 and siteA

The intramolecular and intermolecular interactions are accounted within the Columbic term (electrostatic interactions) and Lennard-Jones function (van der Waals interactions) in the forcefield. Fig. S18 shows RESP charges mapped on site A and bTVBT4 atoms. The strongest Columbic interaction is observed between the imidazole ring of His362 (having negative charge cloud) and the bTVBT4.

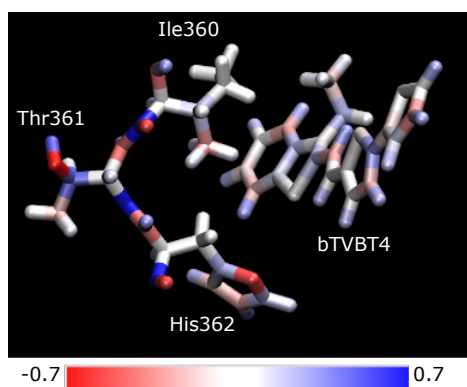


Figure S18: RESP charges used in MD simulation for site A and bTVBT4 are mapped on atoms with color code.

## 2 Experimental materials, methods, and supplementary results

### 2.1 Antibody and ligand double labelling

Frozen human brain tissue sections (10  $\mu\text{m}$ , frontal cortex) with Alzheimers disease (AD) or Picks disease (PiD) pathology were fixed in 70% EtOH for 3 min at 4°C and then incubated in dH<sub>2</sub>O for 2×2 min and phosphate buffered saline (PBS, 10 mM phosphate, 140 mM NaCl, 2.7 mM KCl, pH 7.4) for 10 min at RT. Next, a permeabilization and blocking step in PBS with 0.1% triton x-100 (PBS-T) and 5% normal goat serum was performed for 1 h at RT.

Monoclonal anti-phospho-tau-antibody AT8 (Thermo Scientific) was diluted 1:500 in PBS-T with 5% normal goat serum and added to the sections. After incubation over night at 4°C, the sections were washed in PBS-T for 3×10 min and then incubated for 1 h at RT with goat anti-mouse secondary antibody conjugated to Alexa 488 (Thermo Scientific) diluted 1:400 in PBS-T with 5% normal goat serum. The sections were washed in PBS for 3×10 min and ligand bTVBT4, diluted to 100 nM in PBS, was added. After 30 min at RT, excess ligand was removed by repeated washings in PBS and the sections were mounted using Dako mounting medium for fluorescence (Agilent). The result was analyzed using an inverted Zeiss LSM 780 laser scanning confocal microscope (Zeiss) exciting the samples at 405 nm (lipofuscin), 490 nm (Alexa 488) and 580 nm (bTVBT4).

## 2.2 Hyperspectral and fluorescence lifetime imaging

Frozen brain tissue section (10  $\mu\text{m}$ , frontal cortex) with AD or PiD pathology was fixed in 99.7% EtOH for 10 min at RT, rehydrated in 50% EtOH for 2 min, dH<sub>2</sub>O for 2×2 min and then PBS for 10 min. The section was incubated in 100 nM b-TVBT4 ligand for 30 min at RT and then washed repeatedly with PBS. After mounting with Dako mounting medium for fluorescence (Agilent), fluorescence lifetime imaging was performed using an inverted Zeiss LSM 780 laser scanning confocal microscope (Zeiss) equipped with a 32 channel QUASAR GaAsP spectral array detector. Emitted photons were routed through the direct coupling confocal port of the Zeiss LSM 780 scanning unit and detected by a Becker & Hickl HPM-100-40 hybrid photomultiplier tube (Becker & Hickl GmbH). Data were recorded by a Simple-Tau 152 system (SPC-150 TCSPC FLIM module) with the instrument recording software SPCM version 9.42 in the FIFO image mode using 256 time-channels. A Plan-Apochromat 20×/1.3 Oil DIC objective lens was used, and the pinhole set to 20.2  $\mu\text{m}$ . For excitation at 535 nm, a pulsed tunable In Tune laser with a repetition rate of 40 MHz was used. Data were analyzed using SPCImage version 3.9.4. Emission spectra from bTVBT4 bound to tau pathology in AD were recorded between 561 nm to 687 nm using an inverted LSM

780 confocal microscope (Carl Zeiss, Oberkochen, Germany) and excitation spectra were collected with the same microscope system using a tunable In Tune laser and the excitation wavelength was scanned between 490 to 600 nm having the emission fixed at 612 nm.

When bound to tau aggregates in AD, bTVBT4 displayed strikingly longer decay times, 1.7 to 2.4 ns, than for the ligand in different solvents, verifying that bTVBT4 adopts a distinct conformation when bound to the aggregates.

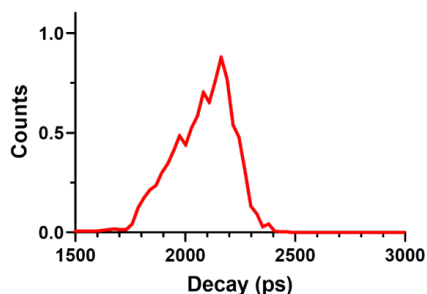


Figure S19: Intensity-weighted mean lifetime distributions of bTVBT4-stained tau deposits in AD brain tissue sections. The fluorescence lifetimes were collected with excitation at 535 nm.

### 2.3 Spectroscopy

Steady-state absorption spectra were recorded using a Shimadzu UV-1601PC spectrophotometer. Measurements were performed with 10 mm quartz cuvettes (Hellma Precision). Steady-state photoluminescence measurements were carried out employing a PTI Quanta-master 8075-22 (Horiba Scientific) equipped with Double Monochromator 300 spectrometer chambers for both excitation and emission. A Hamamatsu R928 PMT was used for detection in the range 185–950 nm. A OB-75X (75 W Xenon arc lamp) was used as the light source. Data acquisition and basic data-handling of steady state luminescence data were carried out with the Felix Data Analysis software and further processed and presented using Origin Pro.

Time-correlated single photon counting (TC-SPC) was used to register and analyze decay

traces of bTVBT4 in a selection of solvents using an IBH system as described recently in Ref. 44. In the experiments a picosecond laser diode at 469 nm was used for excitation and the emission was recorded at 600 nm with a 16 nm slit.

**Table S13:** Experimental absorption ( $\lambda_{\text{abs}}$  and  $\Delta E$ ) and emission ( $\lambda_{\text{em}}$ ) maxima and quantum efficiencies for bTVBT4 in different solvents with dielectric constants  $\epsilon_r$ .<sup>a</sup>

Solvent	$\lambda_{\text{abs}}$ (nm)	$\Delta E$ (eV)	$\lambda_{\text{em}}$ (nm)	QE (%)	$\epsilon_r$
PBS (H <sub>2</sub> O)	463	2.678	604	0.24±0.03	81
methanol	476	2.605	605	0.22±0.02	33
ethanol	479	2.588	604	0.29±0.03	24
n-butanol	490	2.530	605	0.39±0.04	18
n-pentanol	492	2.520	604	0.76±0.09	15
n-octanol	495	2.505	603	0.67±0.08	10
acetonitrile	471	2.632	607	0.23±0.03	37
DMF	477	2.599	614	0.74±0.09	37
CHCl <sub>3</sub>	509	2.436	605	0.05±0.007	4.8

<sup>a</sup>Dielectric constants are taken from <http://www.stenutz.eu/chem/solv6.php?name=octanol> (n-octanol) and [https://www.engineeringtoolbox.com/liquid-dielectric-constants-d\\_1263.html](https://www.engineeringtoolbox.com/liquid-dielectric-constants-d_1263.html) (others).

The blue shift of the absorption spectra of bTVBT4 in more polar solvent might be due to a negative solvatochromic effect associated with an excited state that is less polar than the ground state. In such a situation the ground state becomes more effectively relaxed in the reaction field of the solvent and the ground-to-excited state energy difference thus becomes larger upon introducing more polar solvent. Interestingly, there is almost no, or at least very small, differences in the emission spectra with respect to solvent polarity, corroborating the notion that the excited state is less polar.

Attempted lifetime measurements of bTVBT4 at a concentration of 2.5  $\mu\text{M}$  in selected solvents showed very rapid decays, at the limit of the resolution of our TC-SPC system. Typically, for polar protic solvents such as PBS, methanol, and ethanol, the lifetimes were in the order of 10–20 ps ( $\pm 10$  ps) and for non-protic solvents such as acetonitrile, CHCl<sub>3</sub>, and DMF, around 30 ps ( $\pm 5$  ps). Two representative decay traces are shown in Fig. S21.

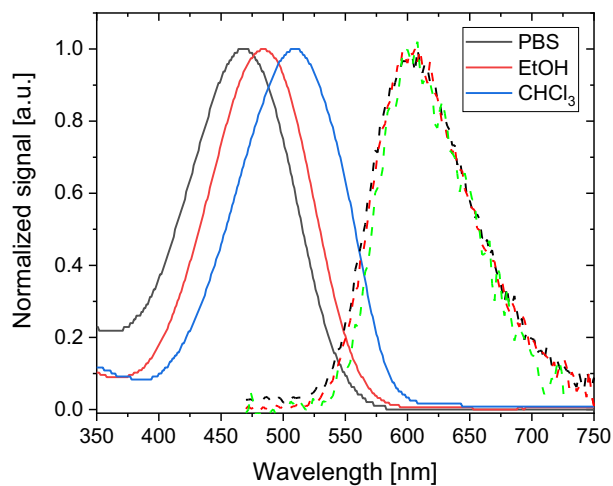


Figure S20: Representative absorption (solid) and emission (dashed) spectra of bTVBT4 in various solvents: PBS, ethanol and chloroform. All spectra are normalized. Photophysical parameters are summarized in Table S13.

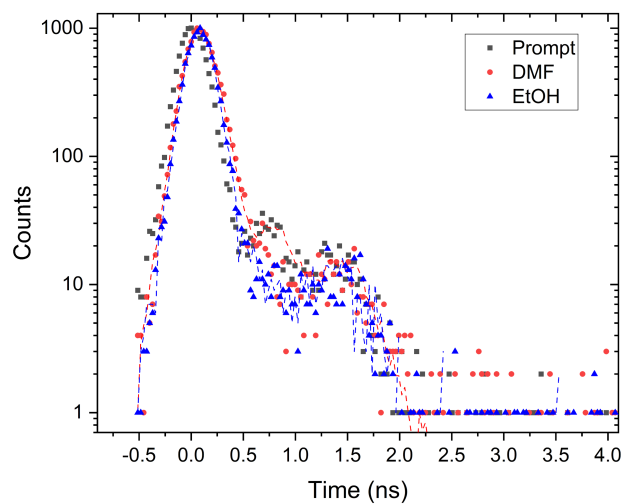


Figure S21: Representative TC-SPC decays of bTVBT4 in EtOH and DMF (2.5 mM) along with attempted re-convolution fits (dashed lines). "Prompt" denotes the system response acquired from the scatter of the pure solvent at the excitation wavelength.

## References

- (1) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (2) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P. et al. The Dalton Quantum Chemistry Program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.
- (3) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
- (4) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (5) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.
- (6) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; AllLaham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closedshell atoms and hydrides of the firstrow elements. *The Journal of Chemical Physics* **1988**, *89*, 2193–2218.
- (7) Petersson, G. A.; AllLaham, M. A. A complete basis set model chemistry. II. Openshell systems and the total energies of the firstrow atoms. *The Journal of Chemical Physics* **1991**, *94*, 6081–6090.
- (8) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett* **2004**, *393*, 51–57.

- (9) Dunning, T. H. Gaussian Basis Sets for use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (10) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **1992**, *96*, 6796–6806.
- (11) Champion, P.; Albrecht, A. On band shapes of electronic transitions in the multimode weak coupling limit. *J. Chem. Phys.* **1980**, *72*, 6498–6506.
- (12) Myers, A. B.; Mathies, R. A. Resonance Raman intensities: a probe of excited-state structure and dynamics. *Biological applications of Raman spectroscopy* **1987**, *2*, 1–58.
- (13) Christiansen, O.; Artiukhin, D.; Godtlielsen, I.; Gras, E.; Györfy, W.; Hansen, M.; Hansen, M.; Kongsted, J.; Klitting, E.; König, C. et al. MidasCpp (molecular interactions, dynamics and simulation chemistry program package in C++). 2021; <https://midascpp.gitlab.io/>.
- (14) Moitra, T.; Madsen, D.; Christiansen, O.; Coriani, S. Vibrationally resolved coupled-cluster x-ray absorption spectra from vibrational configuration interaction anharmonic calculations. *J. Chem. Phys.* **2020**, *153*, 234111.
- (15) Madsen, D.; Christiansen, O.; Norman, P.; König, C. Vibrationally resolved emission spectra of luminescent conjugated oligothiophenes from anharmonic calculations. *Phys. Chem. Chem. Phys.* **2019**, *21*, 17410–17422.
- (16) Sparta, M.; Toffoli, D.; Christiansen, O. An adaptive density-guided approach for the generation of potential energy surfaces of polyatomic molecules. *Theor. Chem. Acc.* **2009**, *123*, 413–429.
- (17) Toffoli, D.; Sparta, M.; Christiansen, O. Accurate multimode vibrational calculations

using a B-spline basis: theory, tests and application to dioxirane and diazirinone. *Mol. Phys.* **2011**, *109*, 673–685.

- (18) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.
- (19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- (20) Lelanda, B.; Paul, D.; Krueger, B.; Walker, R. <http://ambermd.org/tutorials/advanced/tutorial1/section1.htm>.
- (21) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.
- (22) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.
- (23) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual* **2001**, *7*, 306–317.
- (24) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.
- (25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–447.



- (26) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Research Notes* **2012**, *5*, 367.
- (27) Wildman, J.; Repiščák, P.; Paterson, M. J.; Galbraith, I. General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution. *Journal of Chemical Theory and Computation* **2016**, *12*, 3813–3824.
- (28) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.
- (29) Diaz-Quijada, G. A.; Weinberg, N.; Holdcroft, S.; Pinto, B. M. Investigation of Barriers To Conformational Interchange in Oligothiophenes and Oligo(Thienyl)furans. *The Journal of Physical Chemistry A* **2002**, *106*, 1266–1276.
- (30) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (32) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- (33) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

- (34) Olsen, J. M. H.; Kongsted, J. *Advances in Quantum Chemistry, Chapter 3 – Molecular Properties Through Polarizable Embedding*; Academic Press: San Diego, USA, 2011; Vol. 61.
- (35) List, N. H.; Norman, P.; Kongsted, J.; Jensen, H. J. Aa. A Quantum-Mechanical Perspective on Linear Response Theory Within Polarizable Embedding. *J. Chem. Phys.* **2017**, *146*, 234101.
- (36) Ahlström, P.; Wallqvist, A.; Engström, S.; Jönsson, B. A Molecular Dynamics Study of Polarizable Water. *Mol. Phys.* **1989**, *68*, 563–581.
- (37) Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garinger, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W. Cryo-EM structures of tau filaments from Alzheimers disease. *Nature* **2017**, *547*, 185–190.
- (38) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
- (39) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g-whamA Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010**, *6*, 3713–3720.
- (40) Olsen, J. M. H. {PyFraME}: Python Framework for Fragment-based Multiscale Embedding (version 0.2.0). 2018; <https://doi.org/10.5281/zenodo.1443314>.
- (41) Gagliardi, L.; Lindh, R.; Karlström, G. Local Properties of Quantum Chemical Systems: The LoProp Approach. *J. Chem. Phys.* **2004**, *121*, 4494–4500.
- (42) Vahtras, O. LoProp for Dalton. 2014; <http://dx.doi.org/10.5281/zenodo.13276>.
- (43) Sjöqvist, J.; Maria, J.; Simon, R. A.; Linares, M.; Norman, P.; Nilsson, K. P. R.;

Lindgren, M. Toward a molecular understanding of the detection of amyloid proteins with flexible conjugated oligothiophenes. *J. Phys. Chem. A* **2014**, *118*, 9820–9827.

- (44) Gustafsson, C.; Shirani, H.; Leira, P.; Rehn, D. R.; Linares, M.; Nilsson, K. P. R.; Norman, P.; Lindgren, M. Deciphering the Electronic Transitions of Thiophene-Based Donor-Acceptor-Donor Pentameric Ligands Utilized for Multimodal Fluorescence Microscopy of Protein Aggregates. *ChemPhysChem* **2021**, *22*, 323–335.

### A.3 THE ROLE OF MICRO-SOLVATION ON COMPUTED EMISSION SPECTRA: THE CASE OF OXAZINES

The manuscript:

**The role of micro-solvation on computed emission spectra: The case of oxazines,**

T.M.N. Nguyen and C. König,

is in preparation.

# **The role of micro-solvation on computed emission spectra: The case of oxazines**

Nghia Nguyen Thi Minh and Carolin König\*

*Institut für Physikalische Chemie und Elektrochemie, Leibniz Universität Hannover,  
Callinstr. 3A, 30167 Hannover, Germany.*

E-mail: [carolin.koenig@pci.uni-hannover.de](mailto:carolin.koenig@pci.uni-hannover.de)

## Abstract

Oxazine dyes act as reporters of their near environment by the response of its fluorescence spectra. At the same time, their fluorescence spectra exhibit a pronounced vibrational progression. In this work, we computationally investigate the impact of explicit and implicit solvation as well as aggregated betaine on the vibrational profile of fluorescence spectra of different oxazine derivatives. For aggregated betaine as well as a water molecule located above the plane of the dyes, we observe a distinct modification of the vibrational profile, which is more pronounced than the effect of implicit embedding. Our analysis shows that this effect cannot be explained by a pure change in the electrostatic environment, but that also vibrational degrees of freedom of the environment can be decisive for the vibrational profile and can, hence, should not be neglected.

## Introduction

The fluorescent properties of oxazine dyes, such as Nile blue, are highly sensitive to the immediate chemical environment of the dye: Oxazine dyes exhibit a pronounced solvchromism<sup>1-3</sup> and have been suggested as fluorescent probes in biological applications.<sup>4-10</sup> The latter includes the detection of amyloid- $\beta$  deposits,<sup>6</sup> which are hallmarks of Alzheimer's disease.<sup>11,12</sup>

To understand the sensitivity of the fluorescence pattern of oxazine dyes on immediate environment and therefore the oxazine's function as fluorescent probes, computational assistance is required. This assistance, is, however, challenged by many aspects to be considered simultaneously, these are (i) direct interaction of the environmental molecules with the dye (ii) dynamic effects on the spectra, and (iii) vibrational couplings. The challenge is, hence, to treat the (anyways computationally demanding) vibronic coupling in the environment. For this, one may apply, implicit embedding models, such as the polarizable continuum model (PCM),<sup>13-15</sup> which allows for the description of mean-field solvent effects on the position of the spectra and the associated vibronic structure.<sup>15,16</sup> This method can cover the solvent shift

( $\lambda_{max}$ ) of oxazine dyes<sup>17</sup> and can also alter the vibrational profile of fluorescence spectra.<sup>18–20</sup> It is, however, not able to account for inhomogeneous spectral broadening effects, which arise from explicit solute-solvent interactions.<sup>16,21,22</sup> A more explicit description of these interactions is obtained by using cluster models. This method involves treating solute molecules that interact with the solvent at the quantum mechanical level. Although this approach may provide higher accuracy, it poses several challenges when performing excited state calculations with explicit solvent molecules. The main challenge is the increased computational cost resulting from the larger system size and the need to sample solute-solvent configurations. This limits the feasibility of the method to a small number of solvent molecules. It is worth noting that the solute-solvent interactions may be significant at short-range, that is, within the first solvation shell, but their influence decreases rapidly with distance. Hence, for long-range interactions, a more approximate solvent model can often be employed.<sup>16</sup>

In this study, we consider water solution, being the most abundant molecule in the biological context, and individual betaine molecules in close proximity to the oxazine dyes. Betaine (N,N,N-trimethylglycine) is a zwitterionic compound that has a quaternary ammonium group and a carboxylate group, making it a highly polar but neutral molecule.<sup>23,24</sup> It possesses a very large dipole moment of 11.9 D, which is much higher than the dipole moment of a single water molecule, which is only 1.85 D.<sup>24–26</sup> This large dipole moment of betaine is attributed to the presence of two opposite charges separated by a considerable distance, resulting in a significant separation of positive and negative charges within the molecule. This unique property of betaine makes it an excellent candidate for stabilizing biological macromolecules, especially in harsh environments such as extreme pH.<sup>27,28</sup> Due to the large dipole moment, betaine has been used to probe the environmental sensitivity of fluorescent dyes.<sup>29</sup>

In the following, we study the solvation effects on the vibrationally resolved fluorescence spectra of oxazine dyes (*c.f.* Figure 1), namely nile blue (NB+), oxazine 1 (Ox1+), darrow red (DR+), crestyl violet (CV+), oxazine 4 (Ox4+), oxazine 170 (Ox170+). First, we

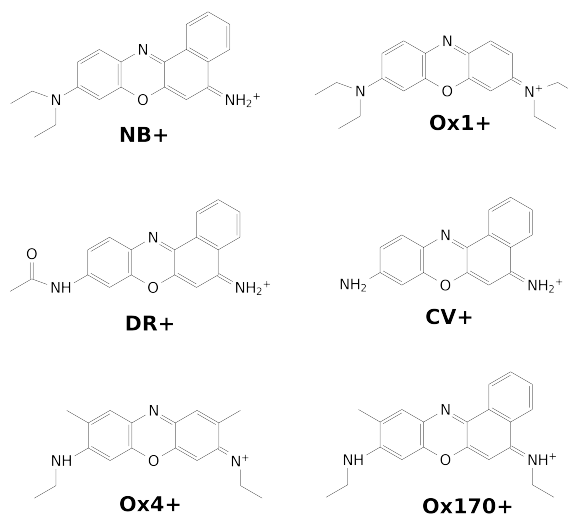


Figure 1: Chemical structures of nile blue (NB+), oxazine 1 (Ox1+), darrow red (DR+), crestyl violet (CV+), oxazine 4 (Ox4+), oxazine 170 (Ox170+)

validate our computational setup for isolated oxazines against experimental data in the gas phase.<sup>30</sup> Thereafter, we investigate the effect of individual water molecules on the vibrational profile on the emission spectra of Ox4+ and Ox170+ with and without the additional consideration of a continuum environment description. We, hence, also study the impact of a nearby betaine molecule on the vibrationally resolved fluorescence spectra of Ox4+ and Ox170+. For both water and betaine microsolvation, we find that the impact of explicit solvation on the vibrationally resolved fluorescence spectra can go beyond a pure electronic effect but also the environmental vibrational degrees of freedom need to be considered.



## Methodology and computational details

All dye geometries were initially prepared in Gaussian16<sup>31</sup> with a preliminary optimization using B3LYP<sup>32-34</sup> density functional theory in conjunction with the 6-31+G(d,p)<sup>35,36</sup> basis set. They were then used as starting points for a conformational search with TINKER 7.1<sup>37</sup> software generating a library of conformers for each dye by perturbing the geometries with a hopping algorithm<sup>38,39</sup> in the *scan* tool. Structures were saved if their energies satisfied a 10 kcal/mol cutoff from the lowest energy structure found, which resulted in a pool of several conformers for each fluorescent molecule. The conformers were ranked based on their MM3<sup>40</sup> energies and re-optimized using the CAM-B3LYP<sup>41</sup> density functional and a def2-TZVP basis set<sup>42,43</sup> in Gaussian16. Vibrational analyses were performed on the resulting structures to confirm that they were local minima before calculating their first excited state. The time-dependent density-functional theory (TD-DFT) method with CAM-B3LYP/def2-TZVP was used to determine the first excited state of the most populated conformers of each dye.

For the calculation of the vibrational profiles, we considered our recently introduced VCI-in-IMDHO, scheme combining accurate vibration configuration interaction (VCI)<sup>44-47</sup> profiles for selected degrees of freedom with the independent mode displaced harmonic oscillator model (IMDHO)<sup>48,49</sup> for the remaining vibrational space.<sup>18</sup> This scheme includes an *a priori* estimate of the importance of an anharmonic treatment for all vibrational degrees of freedom separately based on three criteria. The first criterion is the importance of the mode for vibrational progression, namely harmonic displacement ( $\Delta_i$ ) based on the IMDHO model. This first measure estimates the impact of the mode under consideration on the vibrational progression. Here, we used 0.2 as a cut-off of the  $\Delta_i$ . Besides the ratio of the Franck-Condon factors, it is also important that their effect is also resolved when assuming a line broadening, i.e., that the additional peaks are not hidden below the broadened lines. Thus, in this study, we only include modes for which the crossing point between the Lorentzian broadening the 0-0 and 1-0 transition (denoted  $\sigma$  in ref. 18) is greater than the chosen half width

at half maximum (HWHM). Finally, by comparing the one-mode vibrational profiles within the IMDHO model with the one-mode anharmonic results, we assessed the impact of one-mode anharmonicity for each mode. For the potential energy surface (PES) generation and anharmonic vibrational wave function calculations of the one-mode vibrational profiles, we employed a locally modified version of the Molecular Interactions Dynamics And Simulation C++ package (MidasCpp)<sup>50</sup> version 2019.04.0. For calculating Franck–Condon factors, we employed Vibrational Configuration Interaction (VCI) wave function, and a Vibrational Self-Consistent Field (VSCF) wave function was used as a reference state. Further details about the criteria and anharmonicity calculation can be found in ref. 18. The outlined analysis revealed for most investigated systems no important anharmonic contributions (see ESI) so that we resorted to the IMDHO model for most calculations. The only exception is one Ox170+–betaine complex (with the betaine at position no.5, see below). For which we identified one important and anharmonic mode, which was treated by VCI, with all other modes treated by the IMDHO model via the VCI-in-IMDHO scheme. This particular mode of the Ox170+–betaine complex no.5 corresponds to scissoring vibration of the C–H groups in the benzene ring that is fused to the oxazine ring.

We applied an HWHM of 200  $\text{cm}^{-1}$  in the time-dependent formulation of IMDHO<sup>51–54</sup> using the Python implementation described in ref. 18. To account for the solvation effect, different solvation models were employed. In particular, we have considered the polarizable continuum model (PCM)<sup>55,56</sup> using the integral equation formalism variant and microsolvation (both in gas phase and with PCM) models. For the microsolvation, we investigated several positions of the solvent (water/betaine) around the solute. These structures were optimized in the ground state and then excited state by QM calculations at the DFT and TD-DFT level of theory (using the same functional and basis set as for the calculation of the emission energies in vacuum).

The experimental spectra were transformed into line shapes by applying an intensity correction proportional to  $\omega^2$ ,<sup>57,58</sup> and the area under the peak was normalized to 1. The

maxima of the isolated harmonic spectrum in gas phase were used as a reference to shift the maxima of the experiment to the same position, if not specified otherwise.

## Results and discussions

### Isolated oxazine dyes

Before analyzing the solvent effect on the vibronic spectra of oxazines, we validated our method for the bare molecule in vacuum against experimental references.<sup>30</sup> The backbone of the investigated oxazines consists of a benzophenoxazine ring and the amino-based side chains which can adopt several distinct orientations with respect to the ring giving rise to different conformers. To find the local minima, the conformational landscape of these dyes has been explored. The results from the scan were collected and the conformers were re-optimized with CAM-B3LYP/def2-TZVP. The calculated free energies are presented in Table S-1 in the electronic supporting information (ESI). Several conformers exist for each molecule, however, only a limited number of them (highlighted in red in Table S-I of the ESI) are responsible for over 90% of the total Boltzmann-weighted population. The main difference between the conformations of each molecule lies in the orientation of the side chains. The respective emission spectra of other highly populated conformers are depicted in Figure S7 in the ESI, showing that the gas-phase emission spectra calculated for different conformers are very similar. This suggests that the electronic structure of the different conformers is conserved, despite their structural differences. Therefore, in the following, we focus on the spectra of the most stable conformer of each oxazine molecule.

In Figure 2, the calculated vibrationally resolved emission spectra of the most populated conformer for each oxazine molecule in their isolated form are presented along with an experimental reference for the emission spectra in a vacuum, as reported by Kjær and Nielsen.<sup>30</sup> The comparison demonstrates that the calculated vibrational shape of the fluorescence spectrum aligns well with the experimental results. We note, however, that the

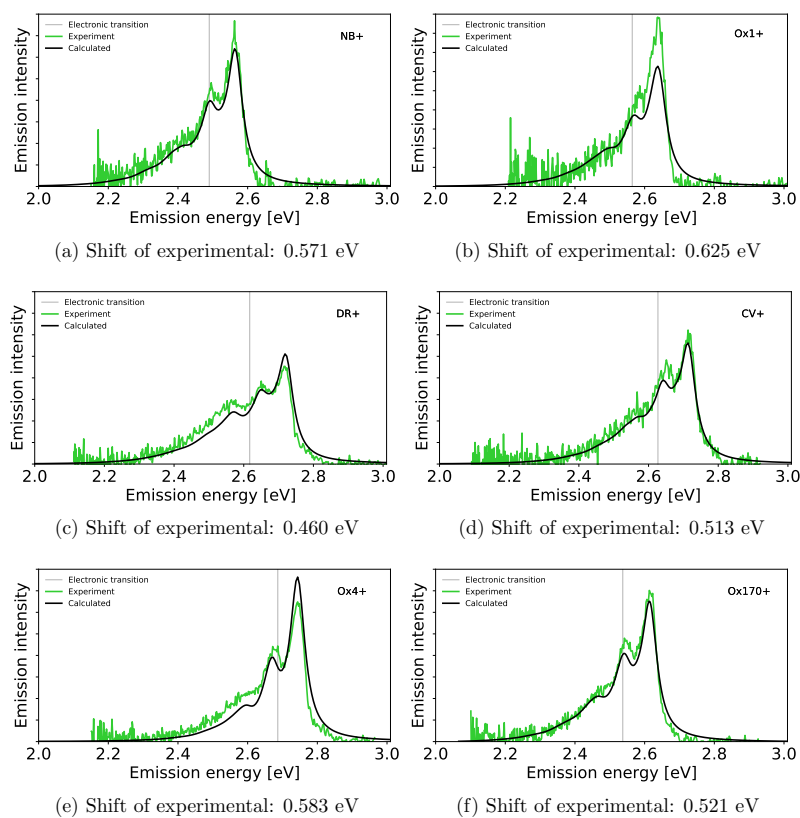


Figure 2: Calculated computational (of the most stable conformer, black) and experimental<sup>30</sup> (green) emission spectra in vacuum of a: Nile blue (NB+), b: Oxazine 1 (Ox-1+), c: Darlow red (DR+), d: Crestyl violet (CV+), e: Ox4+, f: Ox170+ using CAM-B3LYP/def2-TZVP. The experimental spectra are shifted as shown in the respective caption to get the same maximum position as the calculated spectrum of the respective most stable conformer.

calculated spectra have an offset of 0.460 to 0.626 eV compared to the experiment, which can most likely mainly be attributed to shortcomings in the description of the electronic structure by (TD)DFT. The agreement of the vibrational profiles, however, provides further confidence for the application of our recently developed computational protocols<sup>18</sup> to study the vibrational line shape of the oxazine molecules.

### **Effect of individual water molecules aggregated to Ox4+ and Ox170+**

Most experimental measurements are conducted in complex environments, such as water or proteins. Therefore, the study of the interaction between fluorescent dyes and their environment is essential for accurately predicting the photophysical properties of these molecules in real-world applications. By focusing on Ox4+ and Ox170+ in this section, we aim to gain a deeper understanding of how the microenvironment affects vibronic emission. To firmly understand how a microenvironment (such as a nearby solvent dipole) influences a fluorescent dye's optical spectra, it is necessary to know the intrinsic emission by the cation in vacuo.

The effect of the presence of one water molecule on the emission spectra of Ox4+ and Ox170+ was investigated at different positions, as shown in figure 3 and figure 4. In position no.1, the water molecule is located above the 1,4-oxazine scaffold. The other positions refer to the location of water around the substituent on the aromatic ring as depicted in Figure 3-a and Figure 4-a. The calculated emission spectra for the bare dye (black) and the water-dye complexes (color) are also depicted in these figures. The emission spectra for the water-dye complexes (no.2 and no.3) show a small blue shift compared to the spectrum of the respective bare molecule, while the vibrational profile is hardly altered. In contrast to that the side-on coordination over the oxazine ring (position no.1) leads to a larger change in the calculated emission profile: The peak gets broadened, so that the shoulder pattern observed for isolated oxazines and the other water positions is blurred out in this case. These results indicate that the specific position of the water molecule does have a significant impact.

It is worth mentioning that the presence of water molecules results in the emergence

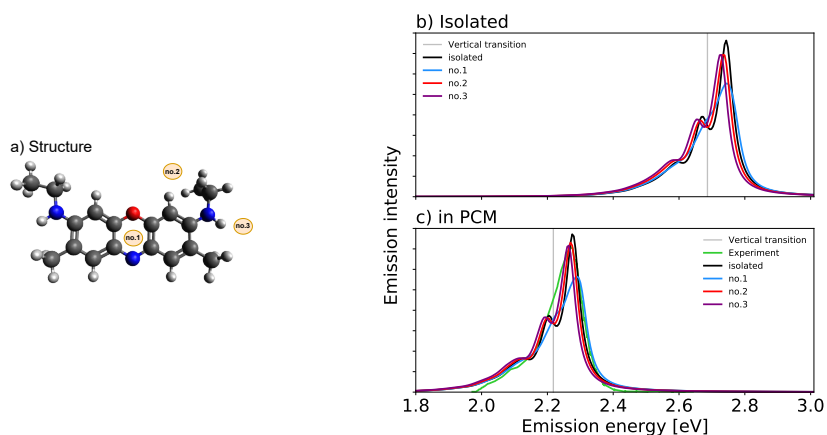


Figure 3: a) Optimized structure of Ox4+ with the inclusion of a single water molecule at various positions (no.1, no.2, no.3) and emission spectra of isolated (black), with a water molecule (color) in vacuum (b) and in PCM (c) using CAM-B3LYP/def2-TZVP with HWHM of 0.025 eV. The experimental spectrum in water<sup>59</sup> (green) is shifted by 0.317 eV to get the same maximum position as the calculated spectrum of the isolated molecule in PCM.

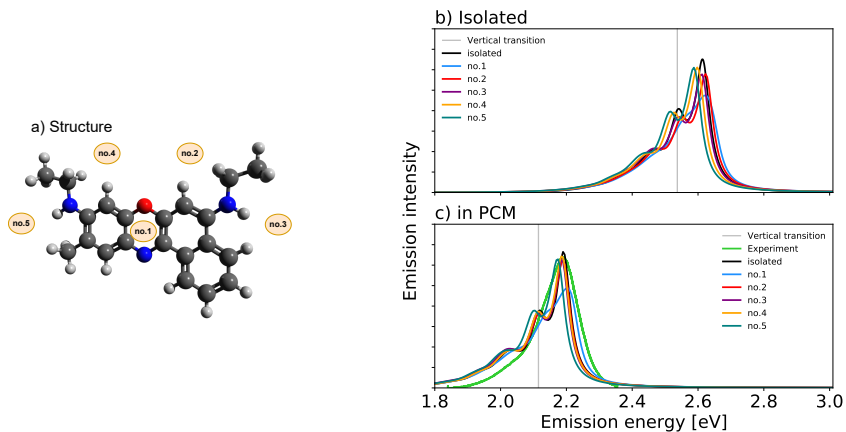


Figure 4: a) Optimized structure of Ox170+ with the inclusion of a single water molecule at various positions (no.1, no.2, no.3, no.4, no.5) and emission spectra of isolated (black), with a water molecule (color) in vacuum (b) and in PCM (c) using CAM-B3LYP/def2-TZVP with HWHM of 0.025 eV. The experimental spectrum in water<sup>60</sup> (green) is shifted by 0.290 eV to get the same maximum position as the calculated spectrum of the isolated molecule in PCM.

of additional vibrational modes with  $\Delta_i$  and  $\sigma$  values surpassing the cut-off. These modes are compiled in Table S-IX and S-X of the ESI. Particularly large values are obtained for a translational motion of the water molecules on top of the oxazine ring of Ox4+ (mode 3). Despite its low harmonic frequency of  $49 \text{ cm}^{-1}$ , the  $\sigma$  value larger than the current HWHM of  $200 \text{ cm}^{-1}$  is observed as well as negligible anharmonic effects on its one-mode vibrational progression (c.f., Table S-IX and Figure S-8 of the ESI). For Ox170+, we do not observe a similarly ranked low frequency mode (c.f., Table S-X of the ESI).

This observation let us examine the impact of additional vibrational modes arising from microsolvation on the vibrational profile. The respective vibrationally resolved spectra of Ox4+ and Ox170+ with one water molecule and different setups, i.e., the isolated dye, including all solvent modes, excluding all solvent modes (and specifically excluding the most significant water mode, mode 3, in Ox4+) are depicted in Figure 5 and 6. In all cases, but the location no.1, the impact of solvent modes on the vibronic spectra are negligible. For Ox4+, we can further assign this impact to the above mentioned mode 3, which also had large  $|\Delta_i|$  and  $\sigma$  values. For the smaller effect of Ox170+, such an assignment is not as clear. The assignment for Ox4+, however, indicates that the important measures obtained from the IMDHO model may also be used as indication for the impact of environmental modes.

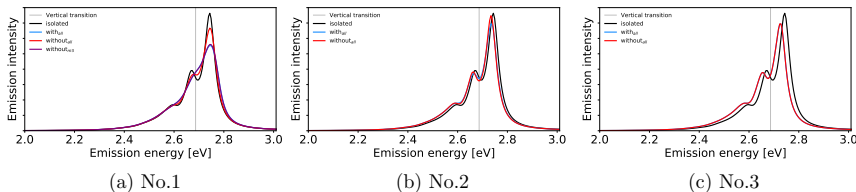


Figure 5: Comparison of emission spectra for Ox4+ with one water molecule, illustrating the following scenarios: isolated dye (black), including all solvent modes (blue), excluding all solvent modes (red), and excluding the most significant water mode, namely mode number 3 ( $\omega_i = 49 \text{ cm}^{-1}$ , purple).

As a next step, we investigate the effect of two water molecules on this particular position. The vibronic spectra of Ox4+ and Ox170+ with one and two water molecules attached

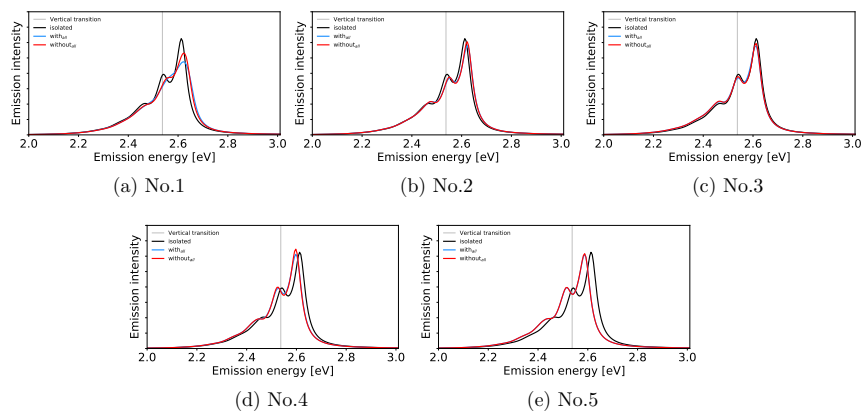


Figure 6: Comparison of emission spectra for Ox170+ with one water molecule, illustrating the following scenarios: isolated dye (black), including all solvent modes (blue), and excluding all solvent modes (red).

at position 1 are shown in Figure 7. The broad overall band observed for the one-water (represented in blue) and two-water clusters (represented in red) are similar, indicating that the number of water molecules at position no.1 has a minimal effect on the shape of the oxazine spectra.

### Continuum water solvation with and without explicit water molecules

As outlined in the introduction, for the description of condense solvents, often continuum models such as PCM are employed. The respective calculated spectra with and without explicit consideration of individual water molecules are shown in Figure 3.c and Figure 4.c: The spectra obtained through the PCM model exhibit a blue shift compared to the spectra in vacuo leading to a closer match to the position of experimental data in aqueous solutions. Still, in both cases, an overall shift between the experimental and theoretical fluorescence spectra of about 0.3 eV are observed, which can likely largely be attributed to the use of TD-DFT in the present study. The overall vibrational shapes of the emission spectra in the vacuum and with the PCM model are very similar for both isolated and microsolvated cases.



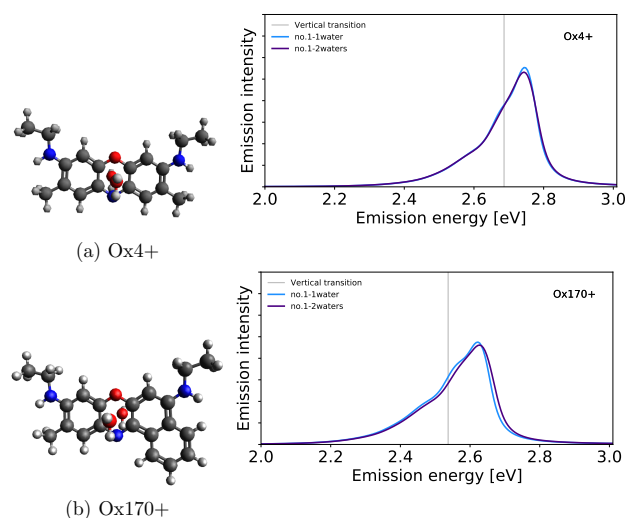


Figure 7: Comparison of the harmonic spectra of a) Ox4+ and b) Ox170+ with one water (blue), and two water molecules (violet) at position no.1 in vacuum.

Hence, the noticeable effect of the on-ring position (position no.1) of the water molecule observed in the vacuum calculation is also obtained in PCM. In this case, the effect of the individual water molecule on the vibrational line shape (not the peak position) is more pronounced than that of the continuum model.

### Microsolvation of Ox-4+ and Ox-170+ by betaine

Due to the larger dipole moment of betaine compared to water, we expect a larger effect of individual betaine molecules attached to the oxazine dyes. We have, hence, placed betaine molecules around the Ox-4+ and Ox-170+ at similar positions as above for the individual water molecules (for the structures, see the SI). The resulting calculated spectra are shown in Figure 8 and Figures S-8–9 of the ESI. These results confirm the overall larger effect for betaine microsolvation compared to the water case: The betaine molecule leads to a larger blue shift in the emission spectra compared to water, and the band shape is also noticeably altered for all locations of betaine. The main difference, next to the peak position is an

overall broadening, which for Ox-4+ is most pronounced for the location no.2, followed by the no.3 and no.1. Similarly also for the Ox170+-betaine complexes, we observe a significant effect for all betaine positions. Here, however, the broadening effect of the on-top position (no.1) is larger than for the respective Ox4+-betaine complex.

To disentangle the electronic and vibrational effects on the changed vibrational profiles, we again analyzed the impact of the additional vibrational degrees of freedom on the vibrational profile, as shown in Figure 9 and 10. For all Ox4+-betaine complexes, the impact of the additional vibrational degrees of freedom is small (c.f., Figure 9). Hence, the change in the vibrational profile due to the aggregation of the betaine molecule can mainly be attributed to electronic effects. The same holds for the Ox170+-betaine complexes with the betaine at positions no.4 and no.5 (see Figure 10). For positions no.1 and no.2, a more pronounced but still small effect of the additional vibrational degrees of freedom is observed, while for position no.3, we observe the largest effect. This is why we investigated this case further: The position no.3 is the only of the Ox170+-betaine complexes for which we find

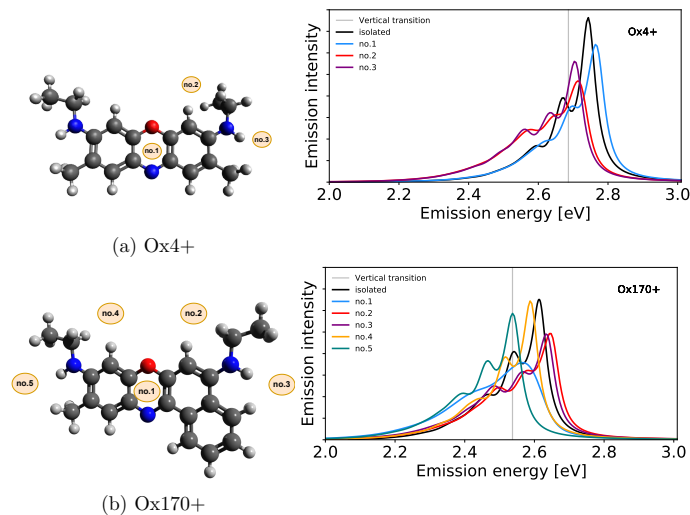


Figure 8: Harmonic spectra of a) Ox4+ and b) Ox170+ for isolated (black) and one betaine cluster (color) at different positions (no.1, no.2, no.3, no.4, no.5) in vacuum.

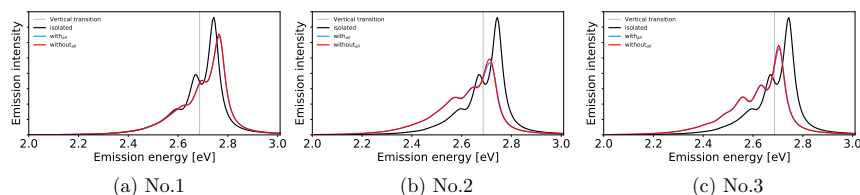


Figure 9: Comparison between the emission spectra of the isolated (black), including all modes (blue), and excluding all betaine modes (red) for Ox4+ with one betaine molecule with different position of the betaine molecule.

vibrational modes with  $|\Delta_i| > 0.6$  and  $\sigma > 200 \text{ cm}^{-1}$  that can be assigned to betaine modes. These are modes 54 and 98 in Table SXIII in the ESI. These specific modes alone cover roughly half of the effect on the betaine modes on the vibronic spectrum as shown in Figure S19 in the ESI.

The overall findings suggest that the location of betaine has a significant impact on the emission spectra of oxazine dyes. This is in line with the proposal of Nielsen to probe the electronic character of the chromophore's transitions and their susceptibility to electric field perturbations through comparing gas-phase action spectra for bare ions with spectra for ion-betaine complexes.<sup>26</sup> Similar to the oxazine-water complexes, we find also here, that the effect of the surrounding may be not only of electronic nature, but that also environmental vibrational degrees of freedom may be important.

## Summary, conclusions and outlook

In summary, the study investigates the impact of explicitly different solvents, including water and betaine, on the fluorescence properties of oxazine dyes. The gas-phase emission spectra of different conformers of each oxazine were found to be similar, indicating that their electronic structure remains conserved despite structural differences in the substituents. An analysis of the importance and anharmonicity of the modes showed that the highly approximate independent-mode harmonic oscillator (IMDHO) model is likely a good approximation in

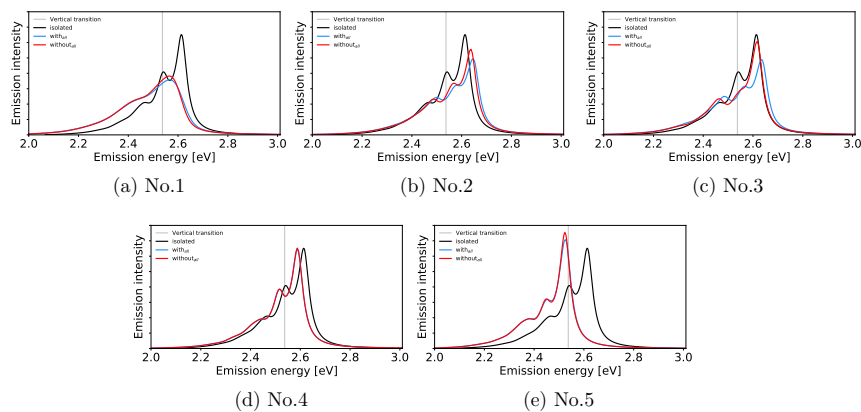


Figure 10: Comparison between the emission spectra of the isolated (black), including all modes (blue), excluding all betaine modes (red) for Ox170+ with one betaine molecule with different position of the betaine molecule. For position no.3, mode 54 ( $\omega_i = 611 \text{ cm}^{-1}$ ) and mode 98 ( $\omega_i = 1153 \text{ cm}^{-1}$ ) are the vibration mode of betaine are also excluded (purple for mode 54, orange for mode 98 and green for both) in the calculated spectra. For position no.5, mode 52 is treated by VCI in the hybrid VCI-in-IMDHO model.

most present cases. This was supported by the finding that the IMDHO spectra of the most stable conformer in a vacuum align well with experimental results for all six investigated oxazine derivatives.

In the following, we found that the position of the water molecule had a significant impact on the spectra of oxazines, particularly when at position no.1 (on top of the 1,4-oxazine ring), which broadened the main peaks of the emission spectra of Ox4+ by the presence of one particular mode from the solvent. Effects of two water molecules on this particular position were also studied indicating an insignificant effect on the shape compared to the one water case. The presence of betaine, which has a very large dipole moment, led to a larger blue shift in the emission spectra compared to water, with the band shape noticeably altered. These findings indicate that the direct molecular environment has a significant impact on the emission spectra of oxazine dyes. This does not only concern the peak position, but also the vibrational profile. We further find that this effect is in some cases not purely

electronic, but also the additional vibrational degrees of freedom can affect the vibrational profile of the fluorescence spectra. This finding is interesting in the sense that the vibrational contribution of the environment is not included in typical multi-level approaches with the explicit environment.

Overall we observe the dependence of the dye's vibrational profile (not only peak position) on the environment. This dependence is more pronounced for betaine aggregation than for a single water molecule. These findings suggest, that detailed environmental information can be extracted from the fluorescent spectra of these dyes when also the change in vibrational profile is considered.

## **Author contributions**

NNTM: Data curation, formal analysis, validation, visualization, writing

CK: Conceptualization, methodology, writing

## **Conflicts of interest**

There are no conflicts to declare.

## **Supporting Information**

Conformational and vibrational mode analysis of isolated oxazine dyes (Section S1 and S2 in the ESI), IMDHO emission spectra of the most populated conformers of oxazine dyes (Section S3 in the ESI), detailed mode analysis of Ox4+ and Ox170+ with the inclusion of a single water (Section S4 in the ESI) and betaine (Section S5 in the ESI) molecule in vacuum.

## Acknowledgements

We thank Steen Brøndsted Nielsen for original data for the emission spectra of the bare oxazine dyes. This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) through the Emmy Noether Young Group Leader Programme (project KO 5423/1-1). The computations were carried out on the cluster system at the Leibniz University Hannover, Germany, which is funded by the Leibniz University Hannover, the Lower Saxony Ministry of Science and Culture (MWK) and the DFG.

## References

- (1) Deye, J. F.; Berger, T. A.; Anderson, A. G. Nile Red as a solvatochromic dye for measuring solvent strength in normal liquids and mixtures of normal liquids with supercritical and near critical fluids. *Anal. Chem.* **1990**, *62*, 615–622.
- (2) Ghanadzadeh, A.; Zeini, A.; Kashef, A.; Moghadam, M. Solvent polarizability and anisotropy effects on the photophysical behavior of oxazine 1: An appropriate polarizability indicator dye. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2009**, *73*, 324–329.
- (3) Tajalli, H.; Gilani, A. G.; Zakerhamidi, M.; Tajalli, P. The photophysical properties of Nile red and Nile blue in ordered anisotropic media. *Dyes Pigm.* **2008**, *78*, 15–24.
- (4) Isak, S. J.; Eyring, E. M. Fluorescence quantum yield of cresyl violet in methanol and water as a function of concentration. *J. Phys. Chem.* **1992**, *96*, 1738–1742.
- (5) Knemeyer, J.-P.; Marmé, N.; Sauer, M. Probes for detection of specific DNA sequences at the single-molecule level. *Anal. Chem.* **2000**, *72*, 3717–3724.
- (6) Hintersteiner, M.; Enz, A.; Frey, P.; Jatón, A.-L.; Kinzy, W.; Kneuer, R.; Neumann, U.; Rudin, M.; Staufenbiel, M.; Stoeckli, M. et al. In vivo detection of amyloid- $\beta$  deposits

- by near-infrared imaging using an oxazine-derivative probe. Nat. Biotechnol. **2005**, 23, 577–583.
- (7) Jose, J.; Burgess, K. Benzophenoxazine-Based fluorescent dyes for labeling biomolecules. Tetrahedron **2006**, 62, 11021–11037.
- (8) Vogelsang, J.; Cordes, T.; Forthmann, C.; Steinhauer, C.; Tinnefeld, P. Controlling the fluorescence of ordinary oxazine dyes for single-molecule switching and superresolution microscopy. Proc. Natl. Acad. Sci. **2009**, 106, 8107–8112.
- (9) Wan, Q.; Song, Y.; Li, Z.; Gao, X.; Ma, H. In vivo monitoring of hydrogen sulfide using a cresyl violet-based ratiometric fluorescence probe. ChemComm. **2013**, 49, 502–504.
- (10) Liu, X.-D.; Fan, C.; Sun, R.; Xu, Y.-J.; Ge, J.-F. Nile-red and Nile-blue-based near-infrared fluorescent probes for in-cellulo imaging of hydrogen sulfide. Anal. Bioanal. Chem. **2014**, 406, 7059–7070.
- (11) Price, J. L.; Morris, J. C. Tangles and plaques in nondemented aging and “preclinical” Alzheimer’s disease. Ann. Neurol. **1999**, 45, 358–368.
- (12) Chiti, F.; Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. Annu. Rev. Biochem. **2017**, 86, 27–68.
- (13) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem. Phys. **1981**, 55, 117–129.
- (14) Mennucci, B. Polarizable continuum model. Wiley Interdiscip. Rev. Comput. Mol. Sci. **2012**, 2, 386–404.
- (15) Zuehlsdorff, T. J.; Shedge, S. V.; Lu, S. Y.; Hong, H.; Aguirre, V. P.; Shi, L.; Isborn, C. M. Vibronic and environmental effects in simulations of optical spectroscopy. Annu. Rev. Phys. Chem. **2021**, 72, 165–188.

- (16) Zuehlsdorff, T. J.; Isborn, C. M. Modeling absorption spectra of molecules in solution. Int. J. Quantum Chem. **2019**, 119, e25719.
- (17) Fleming, S.; Mills, A.; Tuttle, T. Predicting the UV–vis spectra of oxazine dyes. Beilstein J. Org. Chem. **2011**, 7, 432–441.
- (18) Minh, N. N. T.; König, C. Tailored anharmonic–harmonic vibrational profiles for fluorescent biomarkers. Phys. Chem. Chem. Phys. **2022**, 24, 14825–14835.
- (19) De Mitri, N.; Monti, S.; Prampolini, G.; Barone, V. Absorption and emission spectra of a flexible dye in solution: A computational time-dependent approach. J. Chem. Theory Comput. **2013**, 9, 4507–4516.
- (20) Borrelli, R.; Ellena, S.; Barolo, C. Theoretical and experimental determination of the absorption and emission spectra of a prototypical indolenine-based squaraine dye. Phys. Chem. Chem. Phys. **2014**, 16, 2390–2398.
- (21) Loco, D.; Cupellini, L. Modeling the absorption lineshape of embedded systems from molecular dynamics: A tutorial review. Int. J. Quantum Chem. **2019**, 119, e25726.
- (22) DeFusco, A.; Minezawa, N.; Slipchenko, L. V.; Zahariev, F.; Gordon, M. S. Modeling solvent effects on electronic excited states. J. Phys. Chem. Lett **2011**, 2, 2184–2192.
- (23) Lever, M.; Slow, S. The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism. Clin. Biochem. **2010**, 43, 732–744.
- (24) Shikata, T. Dielectric relaxation behavior of glycine betaine in aqueous solution. J. Phys. Chem. **2002**, 106, 7664–7670.
- (25) Rak, J.; Skurski, P.; Gutowski, M. An ab initio study of the betaine anion–dipole-bound anionic state of a model zwitterion system. J. Chem. Phys. **2001**, 114, 10673–10681.



- (26) Stockett, M. H.; Boesen, M.; Houmøller, J.; Brøndsted Nielsen, S. Accessing the intrinsic nature of electronic transitions from gas-phase spectroscopy of molecular ion/zwitterion complexes. *Angew. Chem.* **2017**, *129*, 3544–3549.
- (27) Scheibler, C. Ueber das Betain, eine im Saft der Zuckerrüben (*Beta vulgaris*) vorkommende Pflanzenbase. *Ber. Dtsch. Chem. Ges.* **1869**, *2*, 292–295.
- (28) Scheibler, C. Ueber das Betain und seine Constitution. *Ber. Dtsch. Chem. Ges.* **1870**, *3*, 155–161.
- (29) Kjær, C.; Lisy, J. M.; Nielsen, S. B. Gas-phase ion spectroscopy of congo red dianions and their complexes with betaine. *J. Phys. Chem. A* **2018**, *122*, 3211–3217.
- (30) Kjær, C.; Nielsen, S. B. Luminescence spectroscopy of oxazine dye cations isolated in vacuo. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4600–4605.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (32) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
- (33) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (34) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (35) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J. Chem. Phys.* **1988**, *89*, 2193–2218.

- (36) Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. J. Chem. Phys. **1991**, 94, 6081–6090.
- (37) Ponder, JW, Tinker–software tools for molecular design. Washington University: St. Louis, 2015; <https://dasher.wustl.edu/tinker/>.
- (38) Wales, D. J.; Doye, J. P. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. J. Phys. Chem. A **1997**, 101, 5111–5116.
- (39) Supady, A.; Blum, V.; Baldauf, C. First-principles molecular structure search with a genetic algorithm. J. Chem. Inf. Model. **2015**, 55, 2338–2348.
- (40) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. J. Am. Chem. Soc. **1989**, 111, 8551–8566.
- (41) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). Chem. Phys. Lett. **2004**, 393, 51–57.
- (42) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. Phys. Chem. Chem. Phys. **2005**, 7, 3297–3305.
- (43) Weigend, F. Accurate Coulomb-fitting basis sets for H to Rn. Phys. Chem. Chem. Phys. **2006**, 8, 1057–1065.
- (44) Bowman, J. M.; Christoffel, K.; Tobin, F. Application of SCF-SI theory to vibrational motion in polyatomic molecules. J. Phys. Chem. **1979**, 83, 905–912.
- (45) Christiansen, O. Vibrational coupled cluster theory. J. Chem. Phys. **2004**, 120, 2149–2159.

- (46) Christiansen, O. Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties. Phys. Chem. Chem. Phys. **2007**, 9, 2942–2953.
- (47) Christiansen, O. Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations. Phys. Chem. Chem. Phys. **2012**, 14, 6672–6687.
- (48) Champion, P. M.; Albrecht, A. C. On band shapes of electronic transitions in the multimode weak coupling limit. J. Chem. Phys. **1980**, 72, 6498–6506.
- (49) Myers, A. B.; Mathies, R. A. Resonance Raman intensities: a probe of excited-state structure and dynamics. Biological applications of Raman spectroscopy **1987**, 2, 1–58.
- (50) Christiansen, O.; Godtlielsen, I. H.; Gras, E. M.; Györfy, W.; Hansen, M. B.; Hansen, M. B.; Klinting, E. L.; Kongsted, J.; König, C.; Losilla, S. A. et al. Molecular Interactions Dynamics And Simulation C++ (MidasCPP) package. 2019.
- (51) Myers, A. B.; Mathies, R. A.; Tannor, D. J.; Heller, E. J. Excited state geometry changes from preresonance Raman intensities: Isoprene and hexatriene. J. Chem. Phys. **1982**, 3857–3866.
- (52) Tannor, D. J.; Heller, E. J. Polyatomic Raman scattering for general harmonic potentials. J. Chem. Phys. **1982**, 77, 202–218.
- (53) Neese, F.; Petrenko, T.; Ganyushin, D.; Olbrich, G. Advanced aspects of ab initio theoretical optical spectroscopy of transition metal complexes: Multiplets, spin-orbit coupling and resonance Raman intensities. Coord. Chem. Rev. **2007**, 251, 288–327.
- (54) Petrenko, T.; Krylova, O.; Neese, F.; Sokolowski, M. Optical absorption and emission properties of rubrene: insight from a combined experimental and theoretical study. New J. Phys. **2009**, 11, 015001.

- (55) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. J. Chem. Phys. **1997**, 107, 3032–3041.
- (56) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. Chem. Rev. **2005**, 105, 2999–3094, PMID: 16092826.
- (57) Avila Ferrer, F. J.; Cerezo, J.; Stendardo, E.; Improta, R.; Santoro, F. Insights for an accurate comparison of computational data to experimental absorption and emission spectra: beyond the vertical transition approximation. J. Chem. Theory Comput. **2013**, 9, 2072–2082, PMID: 26583553.
- (58) Charaf-Eddin, A.; Planchat, A.; Mennucci, B.; Adamo, C.; Jacquemin, D. Choosing a functional for computing absorption and fluorescence band shapes with TD-DFT. J. Chem. Theory Comput. **2013**, 9, 2749–2760.
- (59) Czímerová, A.; Iyi, N.; Bujdák, J. Energy transfer between rhodamine 3B and oxazine 4 in synthetic-saponite dispersions and films. J. Colloid Interface Sci. **2007**, 306, 316–322.
- (60) Saha, S.; Majhi, D.; Bhattacharyya, K.; Preeyanka, N.; Datta, A.; Sarkar, M. Evidence of homo-FRET in quantum dot–dye heterostructured assembly. Phys. Chem. Chem. Phys. **2018**, 20, 9523–9535.

Electronic Supplementary Information to  
The role of micro-solvation on the computed emission spectra: The case  
of oxazines

Nghia Nguyen Thi Minh and Carolin König

### S1 Conformational analysis

Table SI: Calculated relative Gibbs's free energies with respect to the most stable conformer  $\Delta G$  (kcal/mol), Maxwell-Boltzmann distribution (%) of the ground state of different conformers of oxazine dyes at CAM-B3LYP/def-TZVP level, T = 298.15 K, R = 8.314 J/molK. The conformers highlighted in red are responsible for over 90% of the total Maxwell-Boltzmann distribution.

Compound	Conformer	$\Delta G$ (kcal/mol)	Distribution (%)
NB+	<b>no.01</b>	0.00	77.2
	<b>no.02</b>	0.75	21.8
	no.03	2.83	0.6
	no.04	3.09	0.4
Ox1+	<b>no.01</b>	0.00	23.8
	<b>no.02</b>	0.01	23.3
	<b>no.03</b>	0.75	6.7
	<b>no.04</b>	0.74	6.8
	no.05	2.87	0.2
	no.06	2.91	0.2
	no.07	1.79	1.2
	no.08	3.14	0.1
	no.09	3.16	0.1
	no.10	1.81	1.1
	no.11	4.67	0.0
	no.12	4.65	0.0
	no.13	4.96	0.0
	no.14	4.89	0.0
	<b>no.15</b>	0.74	6.8
	<b>no.16</b>	0.75	6.7
	<b>no.17</b>	0.90	5.2
	no.18	2.70	1.4
	<b>no.19</b>	0.90	5.2
	no.20	1.70	1.4
	no.21	1.70	1.4
	no.22	1.70	1.4
	<b>no.23</b>	0.75	6.7
	no.24	2.53	0.3
	no.25	2.54	0.3
	no.26	3.77	0.0
Ox-170+	<b>no.01</b>	0.00	83.5
	no.02	8.24	0.0

	no.03	1.40	7.9
	no.04	1.45	7.2
	no.05	6.68	0.0
	no.06	9.61	0.0
	no.07	9.64	0.0
	no.08	7.25	0.0
	no.09	2.84	0.7
	no.10	2.83	0.7
	no.11	8.06	0.0
	no.12	8.04	0.0
	no.13	15.46	0.0
	no.14	15.50	0.0
	no.15	8.60	0.0
	no.16	8.69	0.0
	no.17	8.69	0.0
	no.18	8.24	0.0
	no.19	16.86	0.0
	no.20	16.82	0.0
	no.21	13.93	0.0
	no.22	13.89	0.0
	no.23	10.07	0.0
	no.24	10.04	0.0
	no.25	9.61	0.0
	no.26	9.64	0.0
	no.27	15.22	0.0
	no.28	15.23	0.0
	no.29	16.29	0.0
	no.30	15.46	0.0
	no.31	24.40	0.0
	no.32	24.38	0.0
	no.33	17.71	0.0
	no.34	17.67	0.0
	no.35	16.86	0.0
	no.36	16.82	0.0
	no.37	22.70	0.0
	no.38	22.78	0.0
<hr/>			
Ox-4+	no.01	0.00	89.3
	no.02	1.34	9.3
	no.03	9.85	0.0
	no.04	3.08	0.5
	no.05	2.68	1.0
	no.06	8.54	0.0
	no.07	8.25	0.0
	no.08	8.28	0.0
	no.09	9.38	0.0
	no.10	9.85	0.0
	no.11	13.78	0.0
	no.12	13.78	0.0
	no.13	17.63	0.0
	no.14	15.40	0.0
	no.15	15.4	0.0
	no.16	16.99	0.0
	no.17	18.95	0.0
	no.18	18.92	0.0
	no.19	24.39	0.0
	no.20	26.13	0.0

	no.21	36.94	0.0
	no.22	36.55	0.0
	no.23	26.15	0.0
CV+	no.01	0.00	100.0
DR+	no.01	1.88	4.0
	no.02	0.00	96.0

## S2 Mode analysis

### S2.1 NB+

Table SII: Mode ranking for NB+ using CAM-B3LYP/def2-TZVP. The most important mode with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub>i</sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
34	606	0.776	398	weak

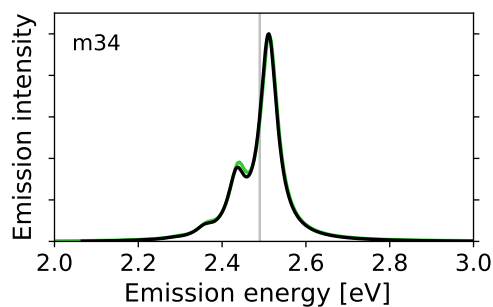


Figure S1: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of NB+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S2.2 Ox170+

Table SIII: Mode ranking for Ox170+ using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
37	612	0.706	353	weak
89	1420	0.254	259	weak
108	1594	0.223	254	weak

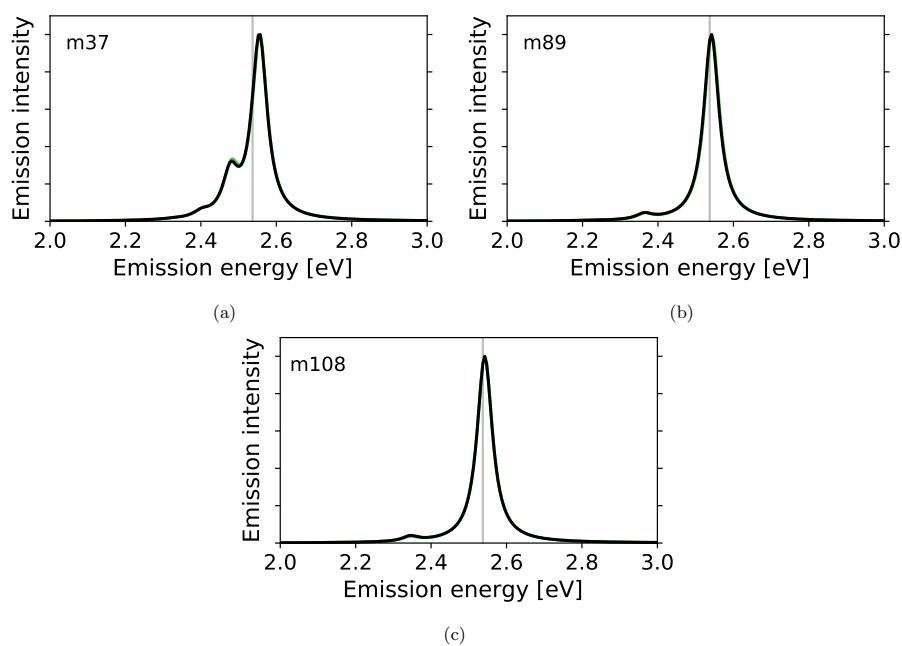


Figure S2: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox170+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.



### S2.3 Ox1+

Table SIV: Mode ranking for Ox1+ using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
36	573	0.670	308	weak
82	1341	0.212	203	weak

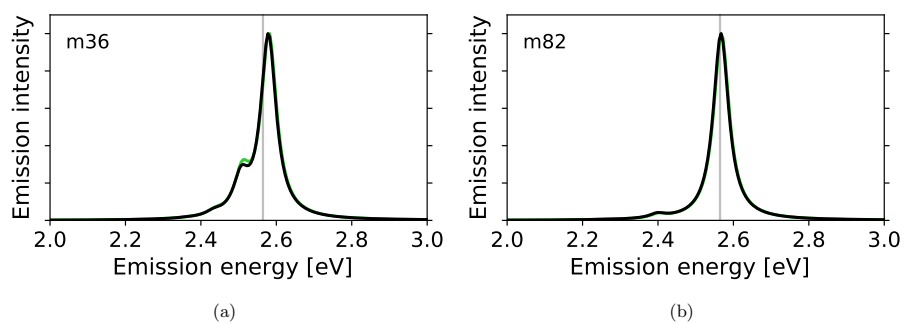


Figure S3: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox1+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S2.4 DR+

Table SV: Mode ranking for DR+ using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
30	597	0.811	418	weak
66	1263	0.292	266	weak
74	1406	0.290	295	weak
67	1273	0.264	242	weak
62	1191	0.262	224	weak
84	1573	0.245	277	weak

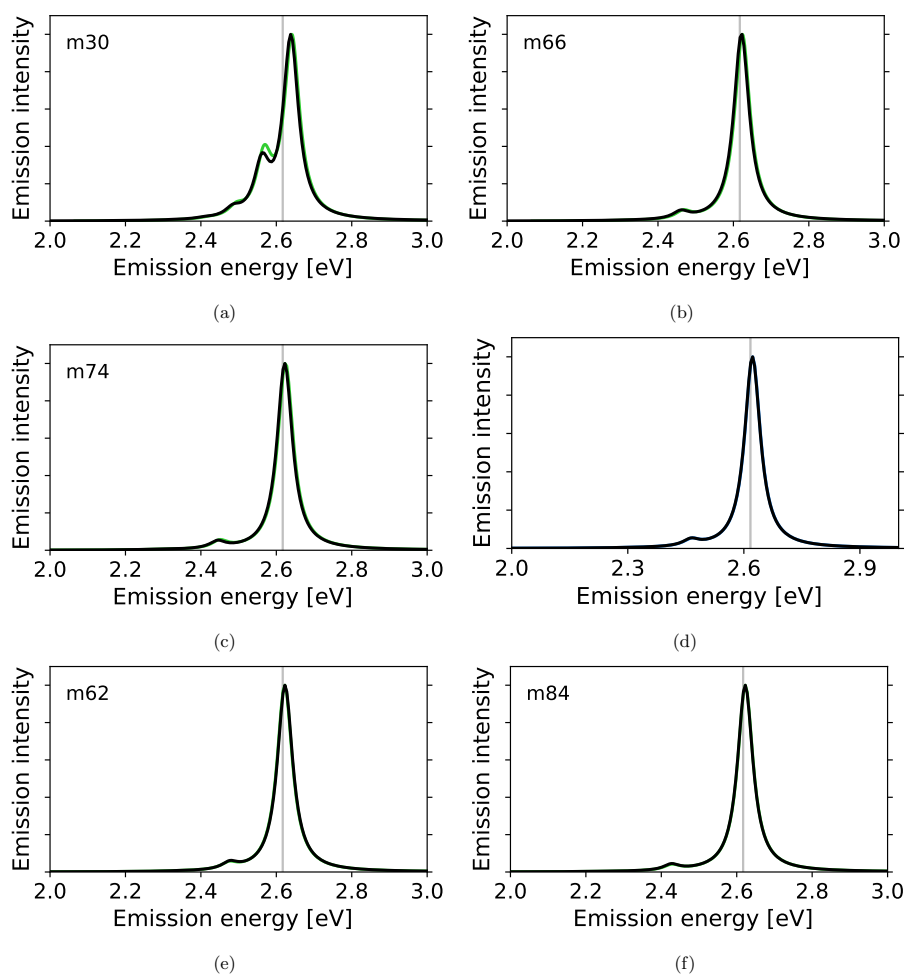


Figure S4: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of DR+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S2.5 CV+

Table SVI: Mode ranking for CV+ using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
26	602	0.848	451	weak
65	1414	0.301	308	weak
72	1579	0.262	298	weak

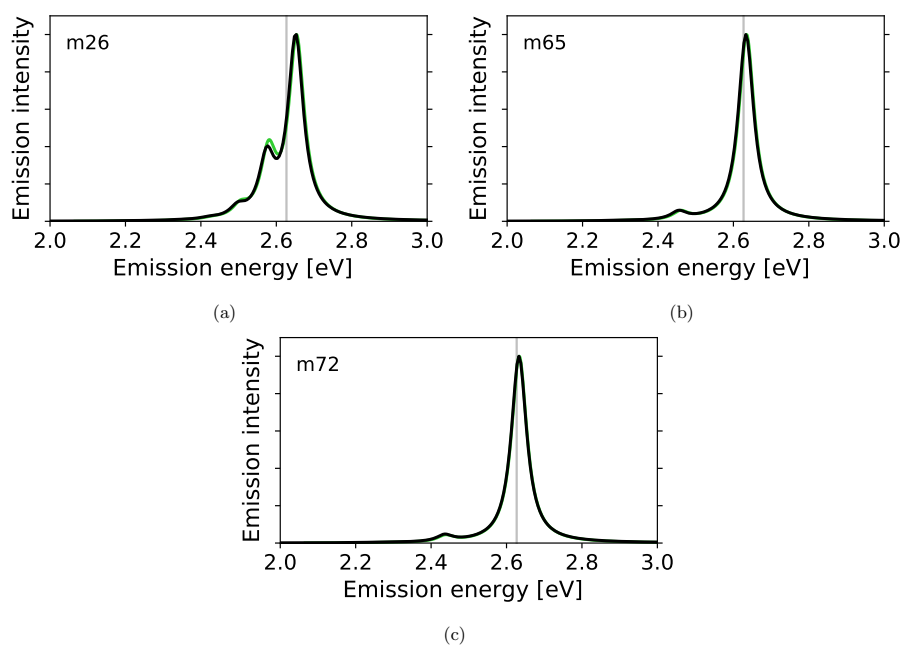


Figure S5: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of CV+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S2.6 O<sub>x</sub>4+

Table SVII: Mode ranking for O<sub>x</sub>4+ using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ , and the effect of one-mode anharmonicity on the vibronic profile (1-mode AH).

Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	1-mode AH
34	594	0.647	306	weak
35	601	0.566	262	weak
100	1583	0.246	280	weak
80	1422	0.205	208	weak

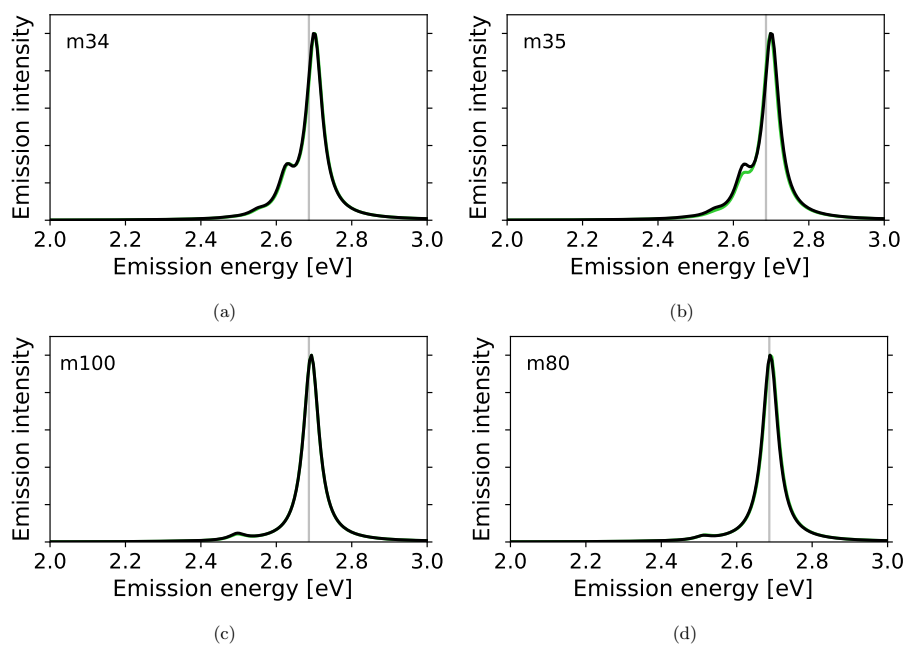


Figure S6: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of O<sub>x</sub>4+ using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

### S3 Isolated dye in gas phase

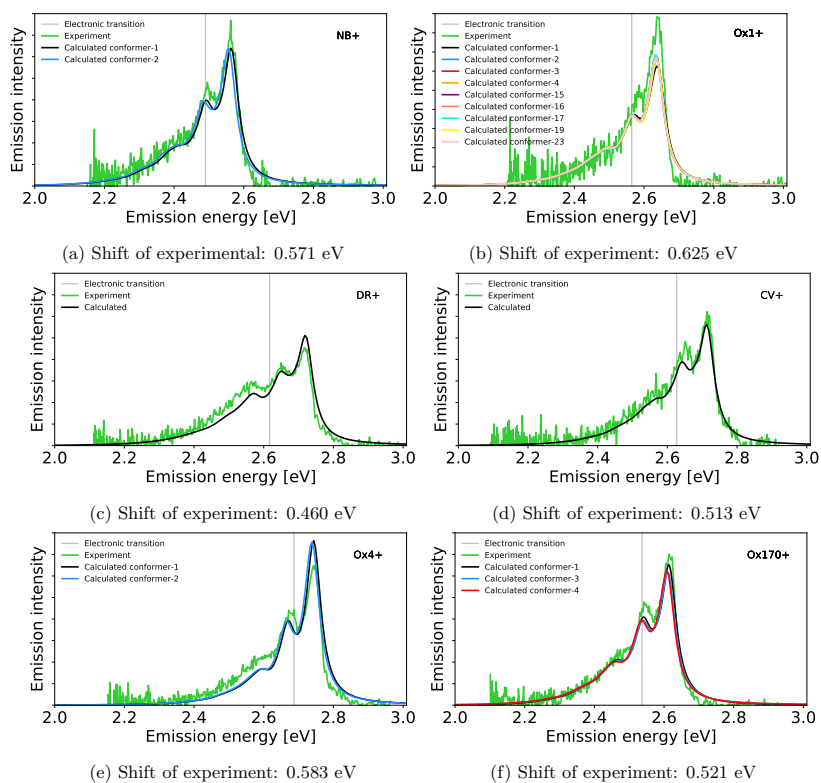


Figure S7: Calculated IMDHO emission spectra of the most stable conformer (black), other conformers (colored) using CAM-B3LYP/def2-TZVP, and experiment [1] (green) in vacuum of NB+ (a), Ox1+ (b), DR+ (c), CV+ (d), Ox4+ (e), Ox170+ (f). The experimental spectra are shifted, as indicated in the subcaptions, to align the maximum position with the calculated spectrum of the corresponding most stable conformer.

## S4 Microsolvation effect of water in gas phase

Table SVIII: Calculated relative Gibbs free energy  $\Delta G$  (kcal/mol) of the ground ( $S_0$ ) and first excited ( $S_1$ ) state, absorption  $E_{abs}$ , emission  $E_{ems}$  energies (eV), oscillator strength (Osc.), and the corresponding Stokes shift ( $\Delta$ ) for Ox4+ and Ox170+ with the inclusion of one single water molecule at various positions (no.1, no.2, no.3, no.4, no.5) using CAM-B3LYP/def2-TZVP.


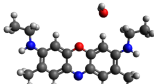
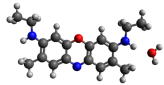
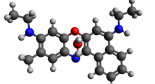
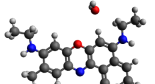
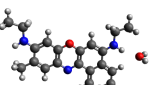
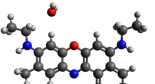
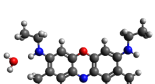
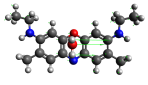
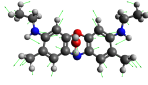
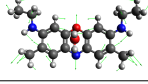
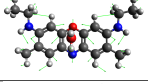
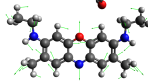
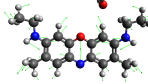
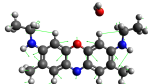
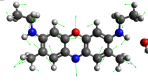
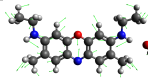
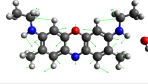
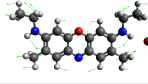
Position no.	Structure	$\Delta G (S_0)$	$\Delta G (S_1)$	$E_{abs}(\text{Osc.})$	$E_{ems}(\text{Osc.})$	$\Delta$
Ox4+ no.1		1.90	2.96	2.82 (1.0035)	2.69 (0.9542)	0.13
Ox4+ no.2		1.28	1.75	2.79 (1.0040)	2.67 (0.9546)	0.12
Ox4+ no.3		0.00	0.00	2.78 (1.0354)	2.66 (0.9896)	0.12
Ox170+ no.1		0.00	3.42	2.71 (0.9476)	2.55 (0.8961)	0.16
Ox170+ no.2		19.27	2.54	2.70 (0.9485)	2.54 (0.9018)	0.16
Ox170+ no.3		17.34	0.00	2.69 (0.9674)	2.53 (0.9195)	0.16
Ox170+ no.4		19.69	2.37	2.69 (0.9487)	2.52 (0.8931)	0.17
Ox170+ no.5		18.48	0.61	2.65 (0.9803)	2.51 (0.9245)	0.14

Table SIX: Mode ranking for Ox4+ with the inclusion of a single water molecule at various positions (no.1, no.2, no.3) using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , and the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ .

Position	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure
no.1	3*	49	1.339	342		41	603	0.632	301	
	40	596	0.535	243		106	1588	0.267	305	
no.2	40	594	0.617	289		41	601	0.576	268	
	106	1584	0.245	279						
no.3	39	595	0.699	338		40	602	0.505	230	
	106	1584	0.236	268		86	1424	0.204	208	

\*The mode is associated with the vibrations of water

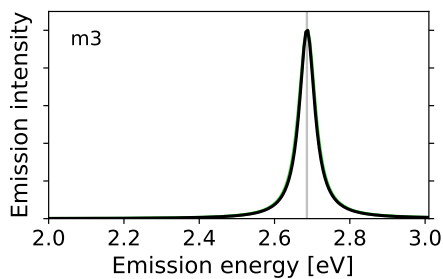
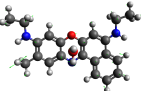
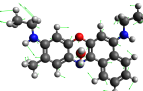
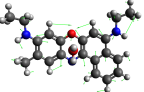
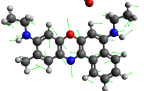

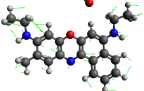
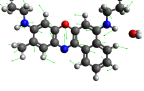
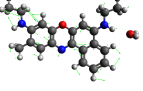
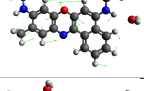
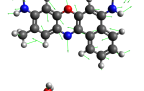
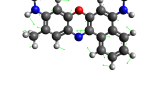
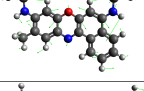
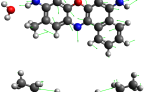
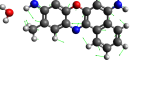
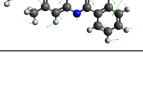


Figure S8: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the mode 3 of Ox4+ with the inclusion of a single water molecule at position no.1 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.



Table SX: Mode ranking for Ox170+ with the inclusion of a single water molecule at various positions (no.1, no.2, no.3, no.4, no.5) using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , and the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ .

Position	Mode <sub>i</sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure	Mode <sub>i</sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure
no.1	43	613	0.681	337		95	1420	0.255	260	
	114	1597	0.248	284						
no.2	43	611	0.672	330		114	1594	0.233	266	
	95	1418	0.2226	230						
no.3	42	612	0.687	340		95	1419	0.250	255	
	82	1232	0.250	221						
no.4	43	612	0.706	352		95	1423	0.241	246	
	114	1595	0.220	251						
no.5	42	612	0.718	361		95	1423	0.261	267	
	114	1594	0.215	245						

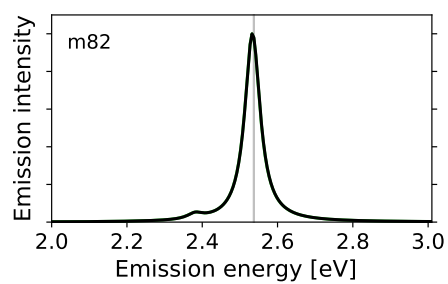
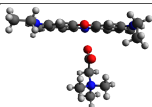

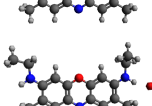
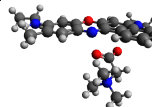
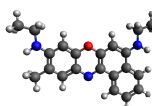
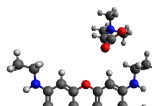
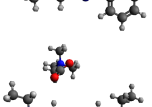
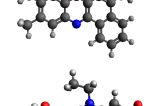


Figure S9: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the mode 82 of Ox170+ with the inclusion of a single water molecule at position no.3 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

## S5 Microsolvation effect of betaine in gas phase

Table SXI: Calculated relative Gibbs free energy  $\Delta G$  (kcal/mol) of the ground ( $S_0$ ) and first excited ( $S_1$ ) state, absorption  $E_{abs}$ , emission  $E_{ems}$  energies (eV), oscillator strength (Osc.), and the corresponding Stokes shift ( $\Delta$ ) for Ox4+ and Ox170+ with the inclusion of one single betaine molecule at various positions (no.1, no.2, no.3, no.4, no.5) using CAM-B3LYP/def2-TZVP.

	Structure	$\Delta G (S_0)$	$\Delta G (S_1)$	$E_{abs}(\text{Osc.})$	$E_{ems}(\text{Osc.})$	$\Delta$
Ox4+ no.1		37.18	38.38	2.83 (0.9142)	2.70 (0.8476)	0.13
Ox4+ no.2		35.39	8.55	2.78 (0.9424)	2.61 (0.8290)	0.17
Ox4+ no.3		0.00	0.00	2.77 (1.1009)	2.62 (1.044)	0.15
Ox170+ no.1		6.02	6.97	2.77 (0.8821)	2.47 (0.7546)	0.30
Ox170+ no.2		4.36	6.67	2.73 (0.9142)	2.54 (0.8701)	0.19
Ox170+ no.3		0.00	1.14	2.73 (1.0383)	2.53 (0.9891)	0.20
Ox170+ no.4		5.09	33.10	2.66 (0.8824)	2.51 (0.8252)	0.15
Ox170+ no.5		0.95	0.00	2.61 (1.0542)	2.45 (0.9822)	0.16

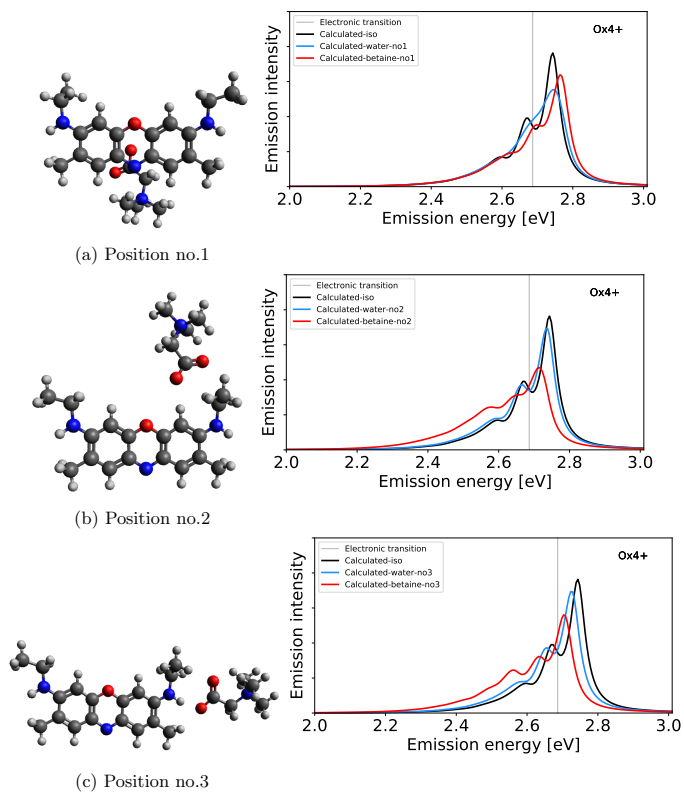


Figure S10: Calculated IMDHO emission spectra of isolated (black), with the inclusion of a single water (blue) and betaine (red) molecule in various positions in vacuum for Ox4+ using CAM-B3LYP/def2-TZVP.

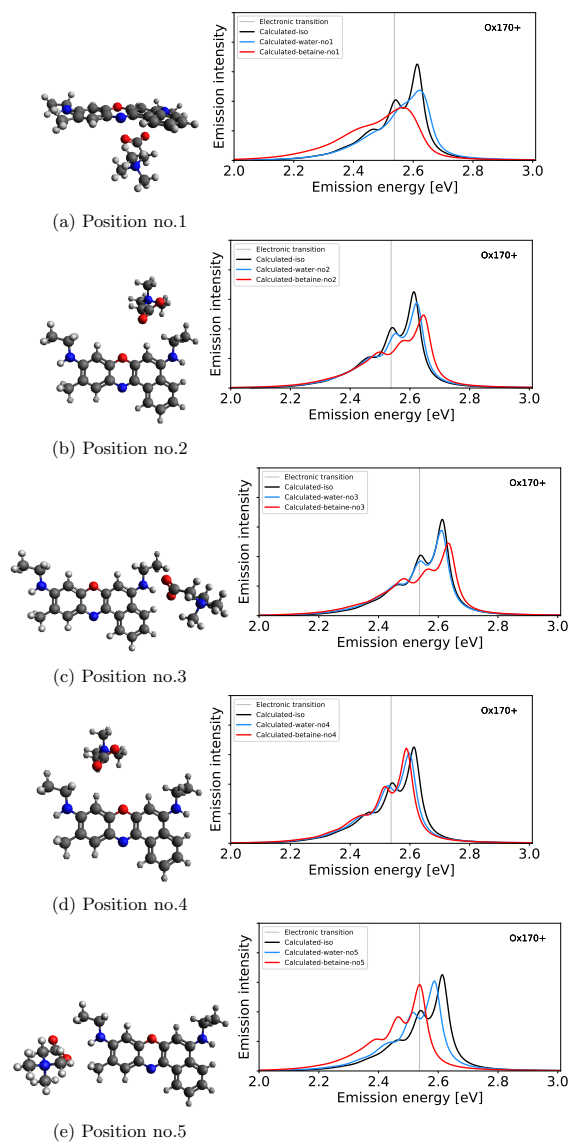
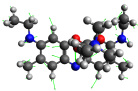
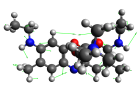
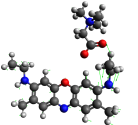
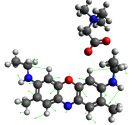
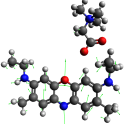
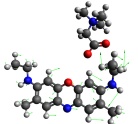
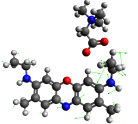
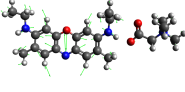
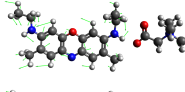
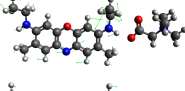
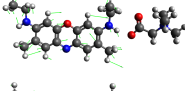
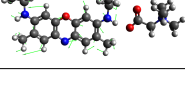
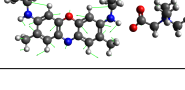


Figure S11: Calculated IMDHO emission spectra of isolated (black), with the inclusion of a single water (blue) and betaine (red) molecule in various positions in vacuum for Ox170+ using CAM-B3LYP/def2-TZVP.

Table SXII: Mode ranking for Ox4+ with the inclusion of a single betaine molecule at various positions (no.1, no.2, no.3) using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , and the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ .

Position	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure
no.1	53	600	0.774	392		145	1588	0.280	321	
	88	1115	0.279	224		145	1587	0.254	290	
no.2	53	599	0.712	349		89	1131	0.470	399	
	93	1172	0.243	204						
no.3	52	597	0.732	361		89	1139	0.376	314	
	93	1172	0.289	245		96	1211	0.238	207	
	99	1262	0.228	206		145	1584	0.219	248	

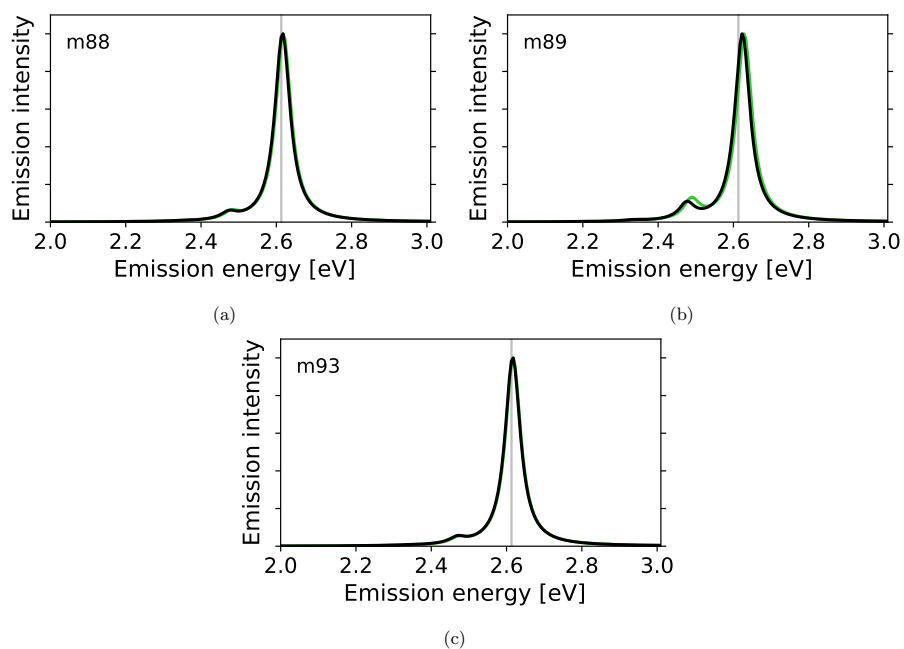


Figure S12: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox4+ with the inclusion of a single betaine molecule at position no.2 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

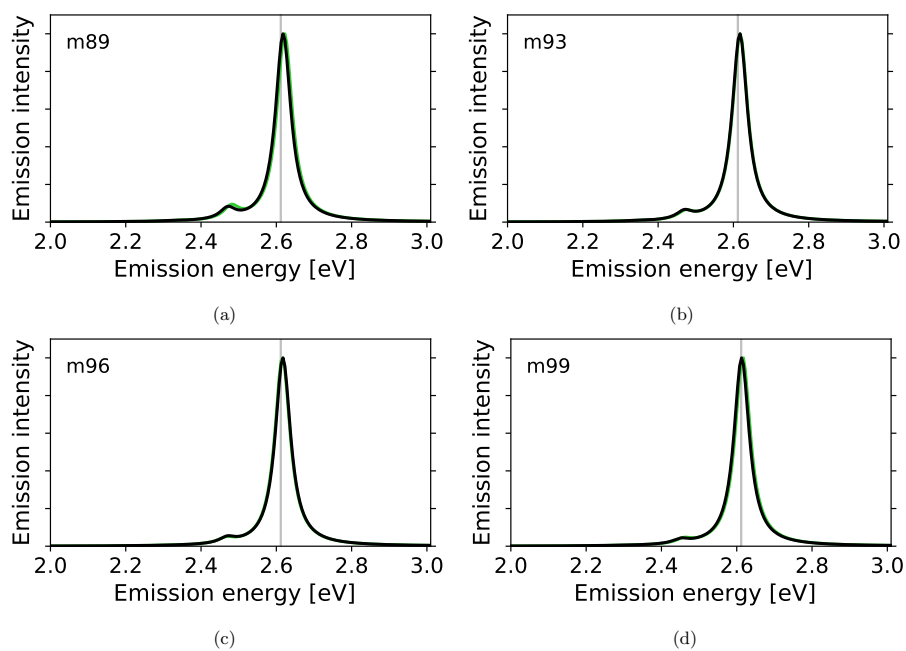
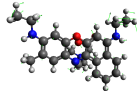
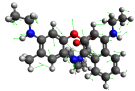
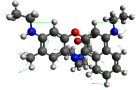
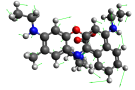
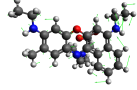
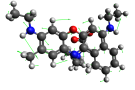
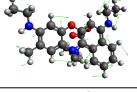
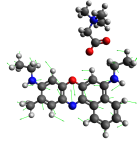
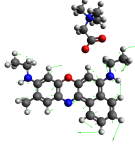
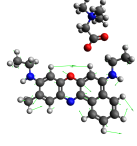
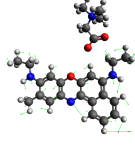
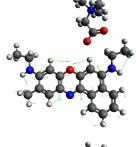
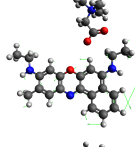
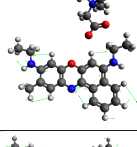
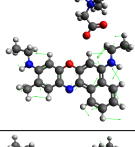
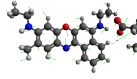
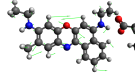
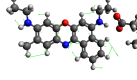
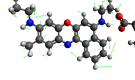




Figure S13: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox4+ with the inclusion of a single betaine molecule at position no.3 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

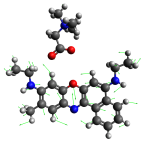
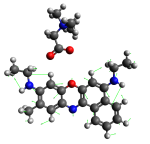
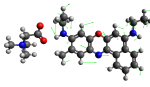
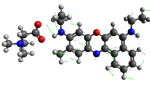
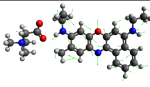
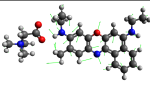
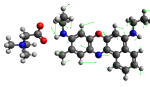
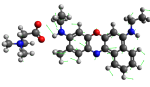


Table SXIII: Mode ranking for Ox170+ with the inclusion of a single betaine molecule at various positions (no.1, no.2, no.3) using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , and the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ .

Position	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure
no.1	14	123	1.340	364		56	615	0.498	231	
	106	1243	0.364	331		98	1157	0.294	246	
	103	1204	0.292	254		153	1598	0.257	295	
	120	1378	0.235	232						
no.2	56	611	0.585	278		96	1149	0.321	268	
	106	1241	0.301	270		103	1195	0.275	237	
	153	1594	0.254	291		104	1220	0.247	216	
	107	1265	0.236	214		120	1378	0.210	207	
no.3	54*	611	0.452	206		106	1240	0.415	381	
	107	1265	0.257	234		98*	1153	0.256	212	
	152	1578	0.210	237		124	1417	0.204	206	

\*The mode is associated with the vibrations of betaine

Table SXIV: Mode ranking for Ox170+ with the inclusion of a single betaine molecule at various positions (no.4, no.5) using CAM-B3LYP/def2-TZVP. The most important modes with the harmonic frequency  $\nu_i$  in  $\text{cm}^{-1}$ , the absolute value of the harmonic dimensionless displacement  $|\Delta_i|$ , and the resolution measure  $\sigma_i$  in  $\text{cm}^{-1}$ .

Position	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure	Mode <sub><i>i</i></sub>	$\nu_i$	$ \Delta_i $	$\sigma_i$	Structure
no.4	56	612	0.688	341		154	1599	0.223	255	
	96	1141	0.352	293		125	1427	0.247	253	
no.5	55	613	0.650	317		52	578	0.624	284	
	96	1141	0.352	293		125	1427	0.247	253	

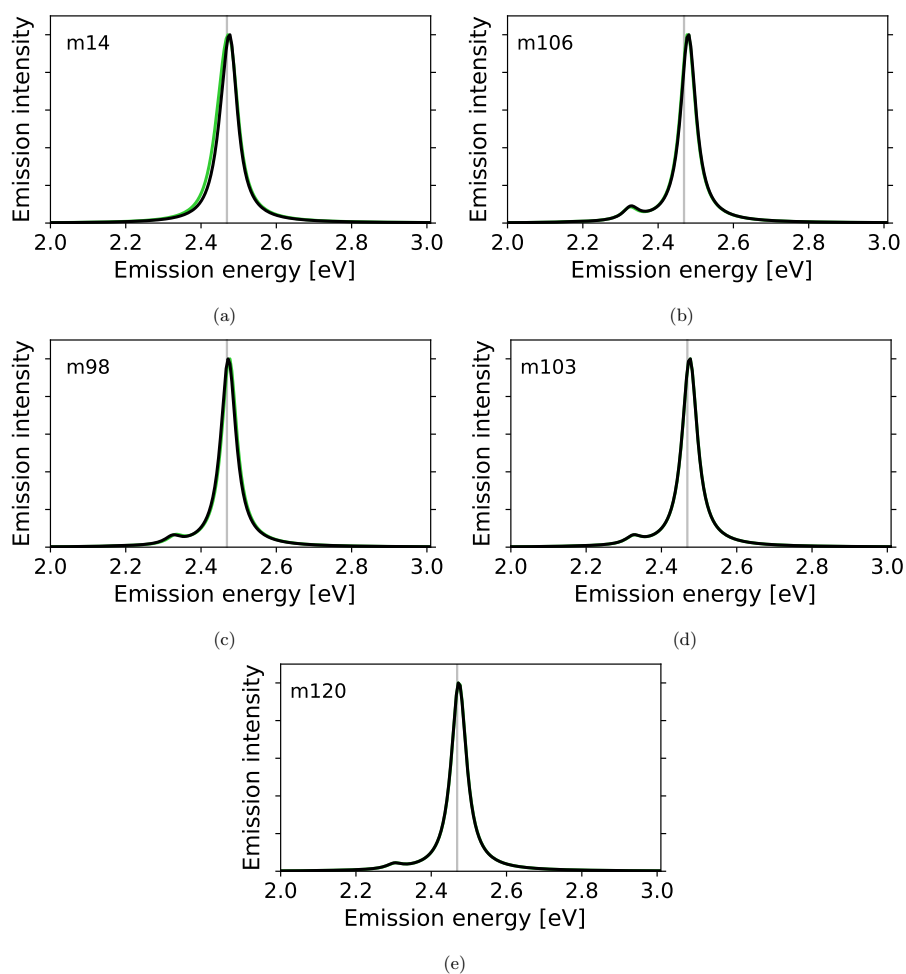


Figure S14: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox170+ with the inclusion of a single betaine molecule at position no.1 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

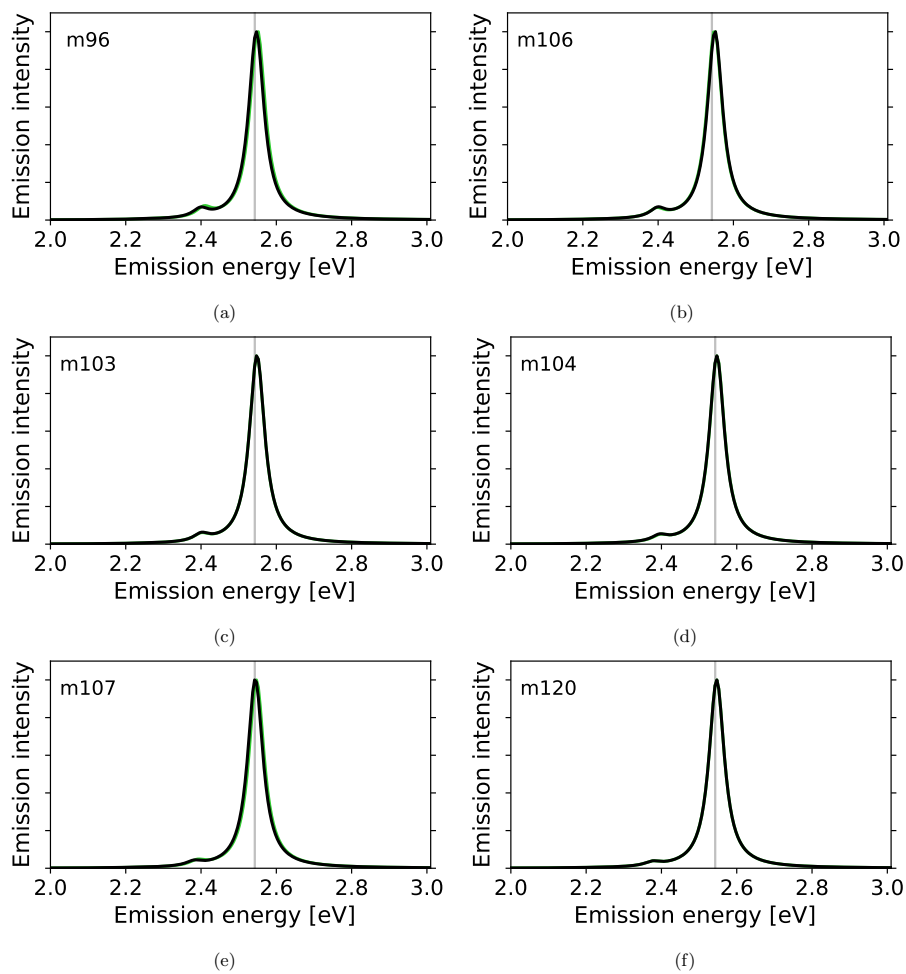


Figure S15: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox170+ with the inclusion of a single betaine molecule at position no.2 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

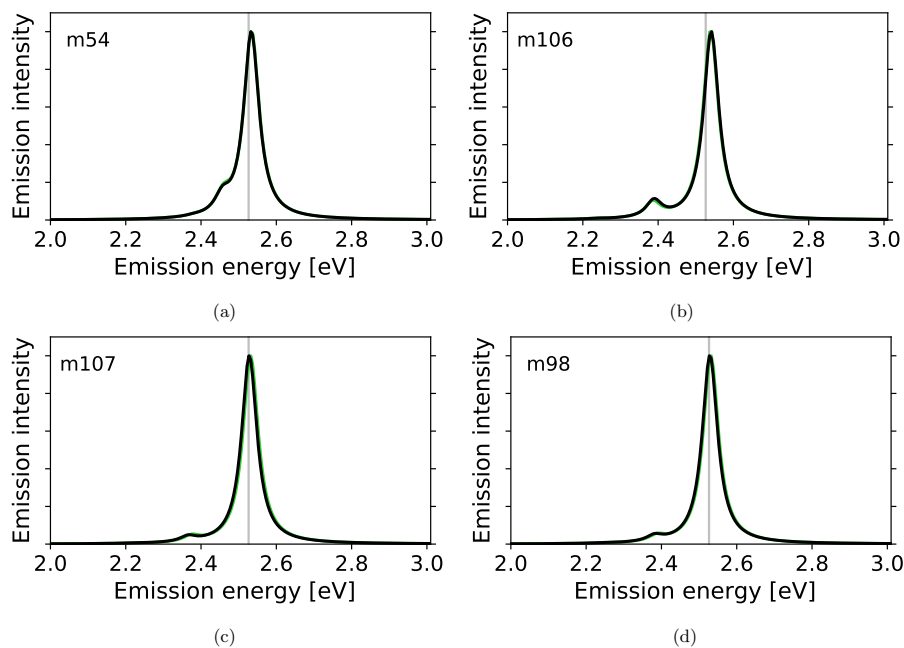


Figure S16: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox170+ with the inclusion of a single betaine molecule at position no.3 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

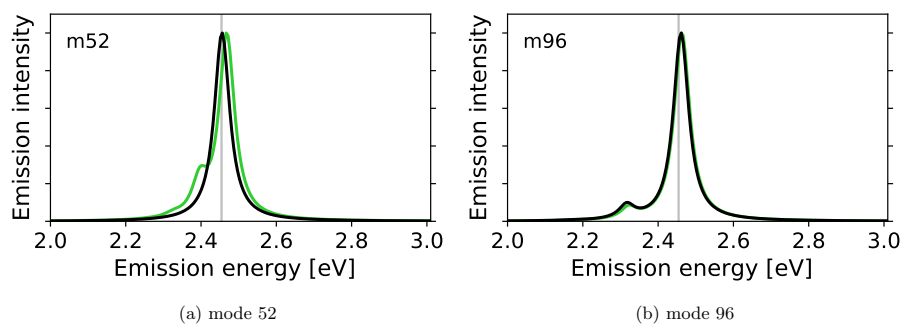


Figure S17: IMDHO treated (green) and VCI treated (black) vibrational profile with a HWHM of  $\gamma = 0.025$  eV for the  $S_0 \leftarrow S_1$  emission for the different modes of Ox170+ with the inclusion of a single betaine molecule at position no.5 using CAM-B3LYP/def2-TZVP. The vertical transition energy is given as a vertical grey line.

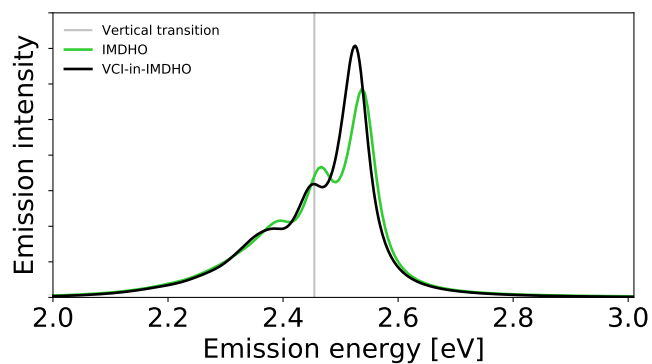


Figure S18: Calculated IMDHO (green) and hybrid VCI-in-IMDHO (black) emission spectra of Ox170+ with the inclusion of a single betaine molecule at position no.5 using CAM-B3LYP/def2-TZVP. The mode treated by VCI in the hybrid approach is mode 52. The vertical electronic emission energy is represented with grey lines.

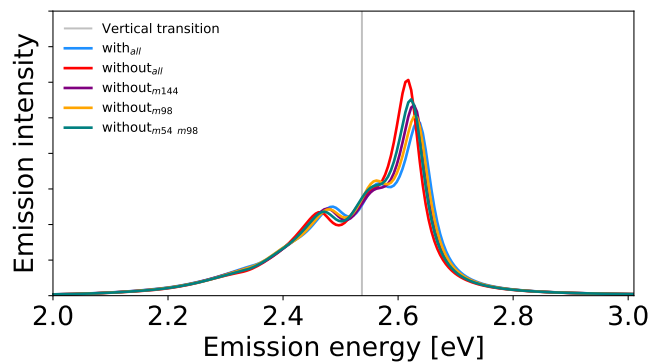


Figure S19: Calculated IMDHO emission spectra of Ox170+ with the inclusion of a single betaine molecule at various positions using CAM-B3LYP/def2-TZVP. The spectra include all betaine modes (blue) and exclude betaine modes (red), mode 54 (purple), mode 98 (orange), and both modes (green).

## References

- [1] C. Kjær and S. B. Nielsen, Phys. Chem. Chem. Phys., 2019, **21**, 4600–4605.

A.4 BINDING OF A PYRENE-BASED FLUORESCENT AMYLOID LIGAND TO  
TRANSTHYRETIN: A COMBINED CRYSTALLOGRAPHIC AND MOLEC-  
ULAR DYNAMICS STUDY

The manuscript:

**Binding of a pyrene-based fluorescent amyloid ligand to transthyretin:  
A combined crystallographic and molecular dynamics study,**

**T.M.N. Nguyen**, A. Begum, J. Zhang, P. Leira, Y. Todarwal, M. Linares, P.  
Norman, D. Derbyshire, E. von Castelmur, M. Lindgren, P. Hammarström and  
C. König,

*J. Phys. Chem. B*, 2023, 127, 30, 6628–6635,

DOI: 10.1021/acs.jpcc.3c02147,

is included in this thesis under the terms of the Attribution 4.0 International (CC  
BY 4.0) license.



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

Ngha Nguyen Thi Minh, Afshan Begum, Jun Zhang, Petter Leira, Yogesh Todarwal, Mathieu Linares, Patrick Norman, Dean Derbyshire, Eleonore von Castelmuur, Mikael Lindgren, Per Hammarström,\* and Carolin König\*

Cite This: *J. Phys. Chem. B* 2023, 127, 6628–6635

Read Online

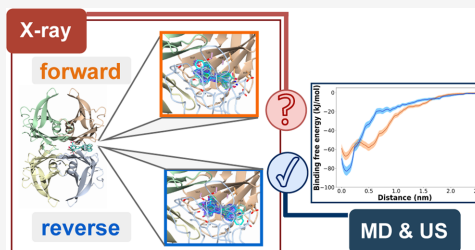
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Misfolding and aggregation of transthyretin (TTR) cause several amyloid diseases. Besides being an amyloidogenic protein, TTR has an affinity for bicyclic small-molecule ligands in its thyroxine (T4) binding site. One class of TTR ligands are trans-stilbenes. The trans-stilbene scaffold is also widely applied for amyloid fibril-specific ligands used as fluorescence probes and as positron emission tomography tracers for amyloid detection and diagnosis of amyloidosis. We have shown that native tetrameric TTR binds to amyloid ligands based on the trans-stilbene scaffold providing a platform for the determination of high-resolution structures of these important molecules bound to protein. In this study, we provide spectroscopic evidence of binding and X-ray crystallographic structure data on tetrameric TTR complex with the fluorescent salicylic acid-based pyrene amyloid ligand (Py1SA), an analogue of the Congo red analogue X-34. The ambiguous electron density from the X-ray diffraction, however, did not permit Py1SA placement with enough confidence likely due to partial ligand occupancy. Instead, the preferred orientation of the Py1SA ligand in the binding pocket was determined by molecular dynamics and umbrella sampling approaches. We find a distinct preference for the binding modes with the salicylic acid group pointing into the pocket and the pyrene moiety outward to the opening of the T4 binding site. Our work provides insight into TTR binding mode preference for trans-stilbene salicylic acid derivatives as well as a framework for determining structures of TTR–ligand complexes.



### INTRODUCTION

Transthyretin (TTR) is a 55 kDa homotetrameric secreted protein with 127 amino acids in each subunit. TTR is synthesized mainly by the liver and the choroid plexus, and to some extent in the pancreas and the eye.<sup>1</sup> It is an important protein for metabolic homeostasis and has also been suggested to function as a molecular chaperone to prevent aggregation of proteins associated with neurodegenerative diseases.<sup>2</sup> TTR circulates in human plasma and cerebrospinal fluid and functions as a transport protein of the metabolic hormone thyroxine (T4) and retinol (Vitamin A) through complex formation with retinol-binding protein (RBP).<sup>1</sup> T4 binds directly to TTR. The T4 binding site is characterized by three subsites, each composed of pairs of symmetric and hydrophobic halogen binding pockets (HBPs). HBP 1 and 1' are in the outer cavity; HBP 3 and 3' are the inner binding subsite; and HBP 2 and 2' are the intervening interfaces between them.<sup>3,4</sup> Hydrogen bonding to the bound ligand is enabled through Ser-117 and Thr-119 buried at the bottom of the binding site. The entrance to the binding site is flanked by

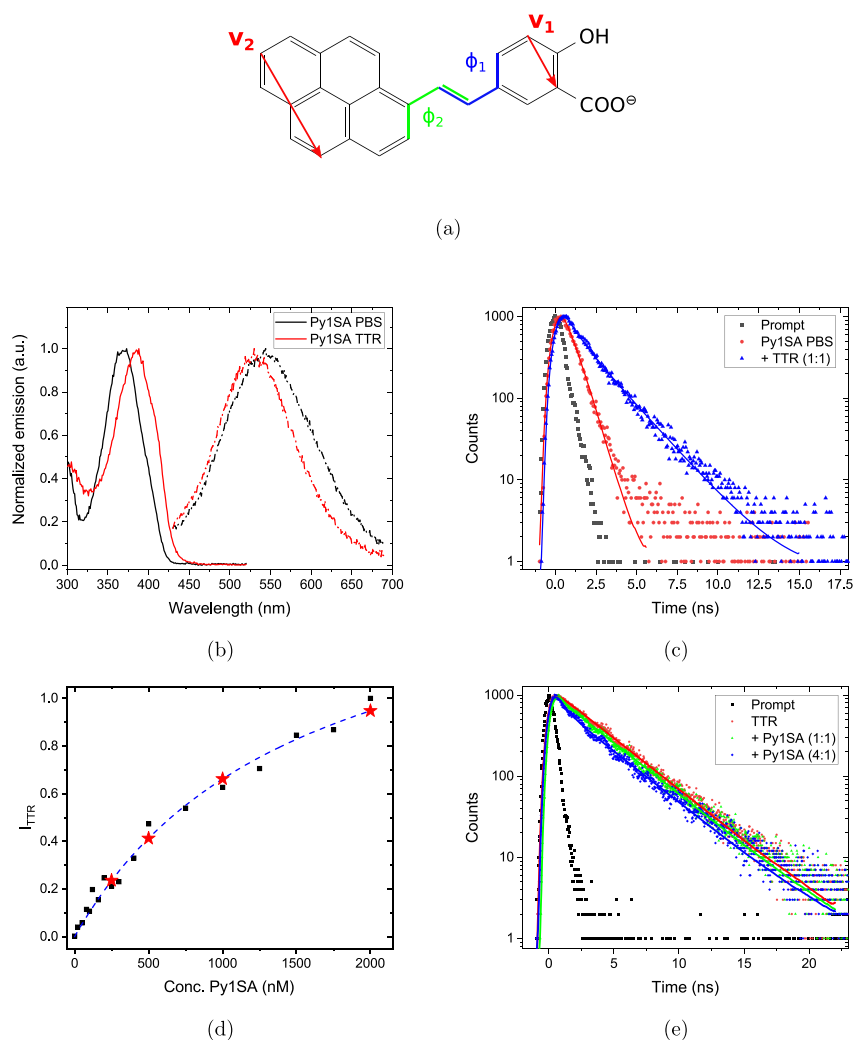
opposing Lys15 residues. In elderly individuals, wild-type TTR misfolds and aggregates into amyloid fibrils mainly manifesting as cardiac amyloidosis.<sup>5</sup> There are also numerous familial forms of TTR amyloid disease caused by over 140 point mutations in the TTR gene often causing familial amyloid polyneuropathy (FAP).<sup>5</sup> Point mutations are inherited dominantly and cause production of a destabilized tetrameric TTR protein, elevating the risk for amyloidosis. There are several treatment options for TTR amyloid diseases including liver transplantation, which surgically removes the liver-produced familial TTR protein.<sup>6</sup> Transcription downregulation can be achieved using small interfering RNA (siRNA) and antisense oligonucleotides (ASOs).<sup>7</sup> Small-molecule stabilizers

Received: March 31, 2023

Revised: July 3, 2023

Published: July 21, 2023



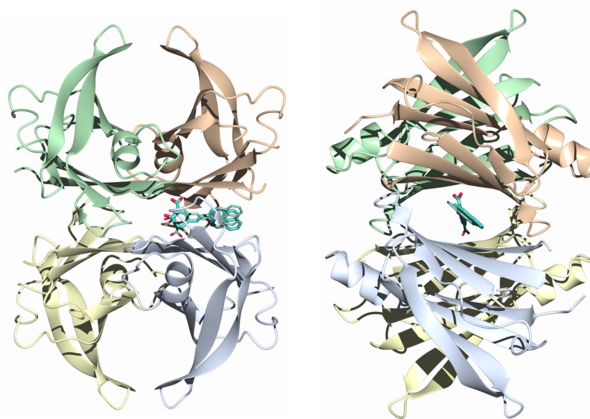


**Figure 1.** Photophysical properties of the Py1SA fluorescent ligand in PBS with and without the presence of TTR. (a) Schematic representation of Py1SA ligand. (b) Excitation (@525 nm emission) and emission (@350 nm excitation) spectra. (c) TC SPC traces of 1  $\mu\text{M}$  Py1SA with and without TTR. (d) Plot of the difference in the fluorescence intensity as a function of the concentration of Py1SA in the presence of a 1  $\mu\text{M}$  solution of TTR in PBS buffer. The Py1SA/TTR ratios 0.5:1, 1:1, 2:1, and 4:1 are indicated by stars. (e) TC SPC—Changes in the decay traces of TTR (1  $\mu\text{M}$ ) tryptophans for different concentrations of Py1SA.

can kinetically stabilize the TTR tetramer, thereby reducing protein misfolding.<sup>8</sup> A first clinical trial of CRISPR/Cas9 has recently been performed with successful results in the reduction of TTR production in humans.<sup>9</sup> Currently, work is ongoing to better diagnose patients early, to follow up and monitor the various available treatments, and to find improved small-molecule ligands as kinetic stabilizers.<sup>10–14</sup>

We are using TTR as a research platform to understand small-molecule recognition and binding specificity in our work to facilitate these efforts.<sup>15,16</sup> While the structure of TTR was

originally solved in 1978<sup>17</sup> and hundreds of structures of TTR variants and TTR complexes are present in the protein database (PDB), the binding modes of TTR ligands are still a matter of intense research: Recently, a quinoline-derived D–A–D-type fluorescent probe was utilized by Sun et al. to study its binding to wild-type TTR,<sup>18</sup> while previous work by some of us<sup>19</sup> has identified a pyrene-based trans-stilbene ligand with a salicylic acid moiety (Py1SA) as an amyloid fibril probe for several different amyloid proteins. Py1SA is an amyloid fluorophore containing a combination of a salicylic acid



**Figure 2.** Orthogonal views of the crystal structures of the TTR–Py1SA complex. The individual molecules of the TTR tetramer (shown as ribbons) are highlighted in colors (yellow: A, green: A', lilac: B, peach: B'). Py1SA (in sticks, here reverse binding mode) binds in the T4 binding site, formed at the B/B' interface. The equivalent A/A' site is devoid of ligand.

trans-stilbene conjugated with a pyrene.<sup>19</sup> Its design was based on the pan-amyloid Congo red analogue X-34 and the pyrene moiety from its extensive fluorescence lifetime. The spectroscopic features of Py1SA with a variety of solvents were reported previously in the context of amyloid binding.<sup>19</sup> Excited-state intramolecular proton transfer along with intramolecular charge transfer was observed for the anionic form in polar solvents.<sup>19</sup>

In the present work, we reveal that Py1SA can in addition to A $\beta$ 1-42 fibrils<sup>19</sup> also bind to the native state of TTR through its T4 binding site. We have previously observed that trans-stilbene-based small molecules bind to the TTR T4 binding site,<sup>20,21</sup> but a compound comprising a pyrene was rather surprising due to its bulkiness. The crystallographic data of the complex, however, do not allow unambiguously demonstrating the binding mode, mainly due to partial occupancy. Consequently, the most likely orientation of the ligand in the binding pocket could not be resolved experimentally. We, therefore, complement the experimental study with molecular dynamics and umbrella sampling calculations that give clear evidence for a dominant orientation of the ligand and thereby describe the TTR–Py1SA complex at atomic resolution.

## METHODS

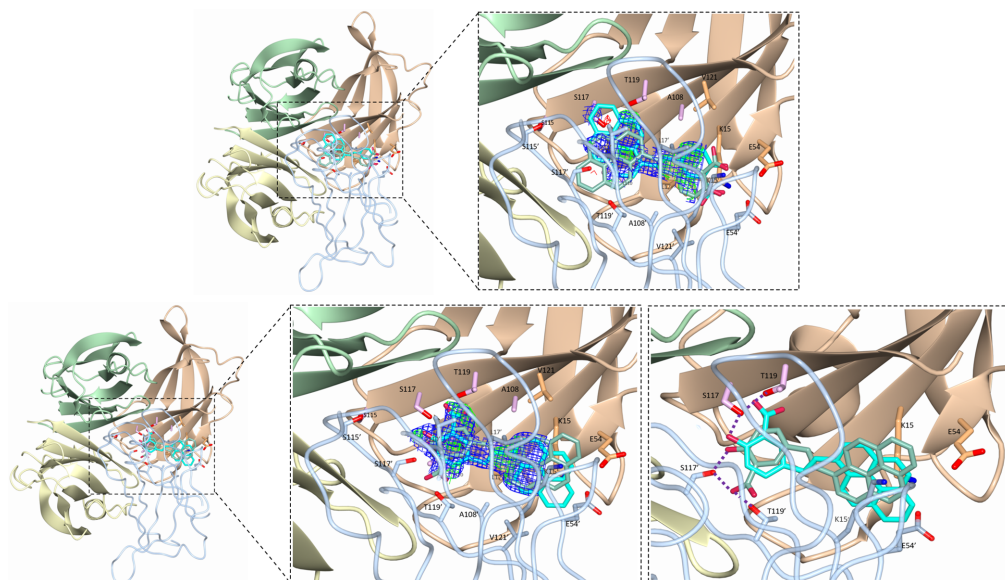
The photophysical properties of typically 1–4  $\mu$ M TTR and Py1SA solutions in phosphate-buffered saline (PBS) were measured using steady-state and time-resolved fluorescence spectroscopy at room temperature (20 °C) and similar procedures as in earlier work,<sup>22,23</sup> see also Section S-1.1 in the Supporting Information (SI). The expression and purification of human TTR as well as the crystallization were carried out as described previously<sup>24</sup> and are briefly summarized in Sections S-1.2 and S-1.3 in the SI. The X-ray diffraction (XRD) data were collected under cryogenic conditions at the MAX VI facility, Sweden, and processed to a resolution of 1.4 Å.<sup>25,26</sup> Phasing was done by molecular replacement.<sup>27</sup> The search model was derived from the published coordinates 1F41, omitting terminal residues and a known flexible region. Hence, the protein structure was

determined for residues 11–98 and 104–122 within each monomer. It was refined against the diffraction data<sup>28</sup> including manual map inspection.<sup>29</sup> For more details on the X-ray analysis, we refer to Section S-1.4 in the SI.

For the molecular dynamics (MD) simulations, we applied Gromacs 2019.3<sup>30</sup> with the Amber ff14SB force field<sup>31</sup> for the TTR protein, TIP3P<sup>32</sup> for water, and a General Amber Force Field (GAFF)<sup>33,34</sup> re-parametrized for Py1SA against B3LYP data (see Section S-1.5 in the SI). The re-parameterization procedure has been described in earlier work.<sup>35</sup> As initial structures, we applied refined (but ambiguous) structures from the X-ray analysis after the required preprocessing.<sup>36,37</sup> In these structures, the ligand is located in the binding pocket. After the optimization step, all protein atoms, except the ones that are within 4 Å of the binding pocket, as well as all bonds have been constrained in the further MD simulations. For each of the four equilibrated structures, we conducted a 1  $\mu$ s MD simulation in the NVT ensemble at 300 K, i.e., under ambient conditions (see also Section S-1.6 in the SI). The umbrella sampling (US)<sup>38–40</sup> simulations were carried out with Gromacs version 2021.3. As coordinate, we choose the center-of-mass (COM) distance between the pyrene or benzene group of the Py1SA ligand and the binding pocket starting from the TTR–Py1SA complex for the reverse and forward binding modes, respectively. Here, the binding pocket is defined by all atoms within 4 Å from the ligand in the equilibrated structure. In the US simulations, we constrained all backbone movement. Further information on the US details can be found in Section S-1.7 in the SI.

## RESULTS AND DISCUSSION

**Spectroscopic Evidence for Py1SA Ligand Binding to TTR.** Differences in the photophysical properties of the fluorescent Py1SA ligand<sup>19</sup> [see Figure 1a] in buffer and with TTR can provide valuable evidence for ligand binding. The excitation and emission spectra in PBS and with a 1:1 stoichiometric amount of TTR are shown in Figure 1b: We observe a distinct red shift for the excitation and a concomitant blue shift of the emission, which indicates that the Py1SA



**Figure 3.** Close-up of TTR–Py1SA binding site, comparing forward (top) and reverse (bottom) binding modes. Forward is defined as when the pyrene moiety is encaged predominantly by HBP 1 (highlighted in pink), while in the reverse case, this moiety is directed along HBP 3 (orange) toward the solvent. In the reverse binding mode, the salicylic group formed hydrogen bonds with Ser-117 and Thr-119. Molecule B' is depicted in “worm” representation with color-matched side chain “sticks” for clarity. Ligand is represented as sticks with the associated sigma-A weighted maps drawn as a mesh (2mFo-DFc in blue contoured at 1.0  $\sigma$  and mFo-DFc in green/red contoured at 3.0  $\sigma$ ).

ligand associates with the hydrophobic binding site. The corresponding time-correlated single photon counting (TC SPC) decay traces at excitation 337 nm were collected at emission in the 520–540 nm region, as shown in Figure 1c. There is a fast single exponential decay in PBS with a decay time of 0.62 ( $\pm 0.008$ ) ns that becomes considerably elongated upon adding equimolar amounts of TTR. In the latter case, a two-component model can be used to fit the decay. Keeping the first component fixed at 0.62 ns, a slower component of 1.97 ( $\pm 0.023$ ) ns contributes with 80% amplitude weight. This indicates that not all of the Py1SA are associated with the binding site; however, the elongated decay time upon binding is also consistent with a considerable increase of the fluorescence quantum efficiency from 6.9 ( $\pm 0.4$ ) to 12.7 ( $\pm 1.7$ )% (see Figure S-3 in the SI). Furthermore, a fluorescence binding assay was set up giving a binding curve shown in Figure 1d that could be fitted to a  $K_D$  of approx. 1.5  $\mu\text{M}$ , being in rough agreement with the time-decay analysis. Another indicator of binding within TTR is to monitor the effect on the fluorescence of tryptophan moieties in terms of fluorescence resonance energy transfer. By stepwise addition of Py1SA to a 1  $\mu\text{M}$  solution of TTR, there is a clear shortening of the tryptophan decay time as shown in Figure 1e. The decay traces go from being single exponential in PBS to a more complicated decay with Py1SA present and was modeled with a double decay. It consists of a slow component associated with only TTR in solution (fixed at: 3.42  $\pm$  0.011 ns) and a much faster component (fitted: 0.561  $\pm$  0.034 ns) gradually growing with increased Py1SA/TTR ratio. At a 4:1 molar ratio of Py1SA/TTR, the latter contributed with 14% amplitude

weight. Taken together, the spectroscopic data suggest that Py1SA binds specifically and efficiently to the native protein also at these low concentrations.

**X-ray Diffraction.** To gain insight into the ligand binding on the atomistic level, the crystallized TTR–Py1SA complex was investigated by X-ray diffraction. The crystals belong to the space group  $P2_122_1$  with two molecules (the AB dimer) in the asymmetric unit; the second biological dimer (A'B') forming the tetramer can be obtained by rotation along the crystallographic 2-fold  $c$ -axis.<sup>17</sup> The inner  $\beta$ -sheets of the dimer-dimer (AB-A'B') interface form two ligand-binding site cavities referred to as sites AA' and BB', respectively (see Figure 2). These two binding sites are symmetry-equivalent.

Clear electron density, deviating from solvent (apo TTR), confirmed the presence of the ligand Py1SA as depicted in Figure 3. For model building, ligand “atoms” were “placed” with an occupancy of 0.3 if visible in the electron density, all remaining atoms were “placed” at 0.1 occupancy. Occupancy was increased in line with (i) developing density and (ii) consistency with surrounding B-factors. However, ambiguity in the initial electron density, and lack of any subsequent improvement during the refinement process, did not permit ligand placement with any degree of confidence.

**Molecular Dynamics Simulations.** Due to the ambiguity in the interpretation of the X-ray data, we additionally performed molecular dynamics (MD) simulations. The starting points for the MD simulations are the four different possible structures of the TTR–Py1SA complex obtained from the refinement of the X-ray data labeled forward-B, forward-B', reverse-B, and reverse-B'. Note that forward-B and forward-B'

as well as reverse-B and reverse-B' represent slightly different structures of symmetry-equivalent binding modes of the TTR–Py1SA complex. These different structures are a result of the electron density being averaged over the AA' BB' symmetry axis. In the MD simulations, all protein atoms except for the ones that are within 4 Å from the ligands were under harmonic constraints.<sup>41</sup>

The subsequent 1  $\mu$ s MD simulations suggest that the reverse mode is more stable than the forward mode. This is supported by several observations: In one of the forward trajectories, the Py1SA ligand tends to move somewhat out of the pocket. In all other trajectories, the ligand is more stationary in the pocket. The unbinding tendency in this trajectory is, e.g., illustrated by the longer distance of the centers of mass of the dye and the pocket (see Figure S-8 in the SI). This movement is also accompanied by larger fluctuations of the Coulombic interaction of the protein with the ligand during the simulation compared to the other trajectories (Figure S-7 in the SI) as well structural changes of the ligand upon leaving the pocket (Figure S-4 in the SI). Further, the reverse mode can establish hydrogen bonds with the pocket (c.f. Figure 3). It shows an average of 3–5 hydrogen bonds (Figure S-6 in the SI). In particular, hydrogen bonds are formed to the residues Ser-117 and Thr-119. This is in line with previous studies, which show that both residues are capable of forming hydrogen bonds with the natural ligand, T4, to stabilize the TTR–T4 complex.<sup>3,42</sup>

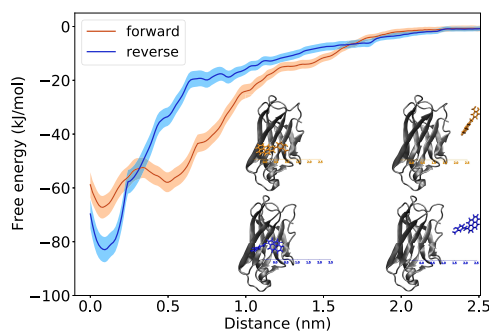
We have further investigated the motion of the ligand in the binding pocket during the simulation time. In particular, we find no rotation of the entire ligand, while located in the binding pocket (Figure S-4 in the SI). For the two dihedral angles [ $\phi_1$  and  $\phi_2$  indicated in Figure 1a], we find simultaneous rather than individual switches within the binding pocket. This can be rationalized by a similar molecular shape and rather small movements necessary for this simultaneous switch compared to rotating about only  $\phi_1$  or  $\phi_2$ . However, the simultaneous switch seems to be a rather rare event (Figure S-5 in the SI).

The Lennard-Jones short-range (LJ-SR), Coulombic short-range (Coul-SR) potential, and their sum are listed in Figure S-7 in the SI. Due to the binding and unbinding process, large fluctuations of the forward-B trajectory are observed in particular in the Coulomb contribution. The overall short-range binding energy for the other forward trajectory is higher than that for the reverse trajectories by about 160–170 kJ/mol. The shifts observed in the optical spectra suggest a hydrophobic character of the binding pocket. This observation is in line with the identified pocket, as it contains apart from Ser-117 and Thr-119 at the inner binding pocket (which establish hydrogen bonds to the ligand) mainly amino acids of hydrophobic nature.

**Umbrella Sampling Simulations.** This difference in the binding energy, however, is not conclusive of the actual free binding energies. For this purpose, we additionally performed umbrella sampling (US) simulations to construct the potential mean force (PMF) surface starting from all four initial structures of the TTR–Py1SA complex. In the pulling trajectories, we increased the distance of the center of the TTR binding pocket to the Py1SA molecule.

The corresponding data for all calculated pathways can be found in Figures S-9 to S-12 in the SI. We have selected the trajectories for the potential of mean force analysis and binding free energy calculation as outlined in Section S-4 in the SI. The

averaged PMF profiles for the forward and reverse modes obtained by evaluating the distance histograms of the simulations by employing the WHAM algorithm are shown in Figure 4.



**Figure 4.** Free energy profiles for Py1SA in the forward (orange) and reverse (blue) modes obtained by the potential of mean force approach by pulling the ligand from the binding site to become free in solution.

For both binding modes, we observe the deepest minimum within the first 0.5 nm from the starting point. The binding free energy was calculated by taking the difference between the last and the lowest values of the PMF graphs. We observe binding free energies of  $67 \pm 4$  and  $83 \pm 5$  kJ/mol for the forward and reverse modes, respectively. This corresponds to a clear domination of the reverse mode. This finding is in line with the MD results discussed above and a previous study where the structure–activity relationships<sup>4</sup> revealed the importance of the presence of the carboxylic acid as well as its position in the ligand structure in its activity.

## SUMMARY AND CONCLUSIONS

There is an unmet need in the field of amyloidosis to easily and accurately diagnose and monitor patients during treatment with different treatment modalities. TTR in TTR amyloidosis is a particularly interesting target due to its abundance of accessible blood samples and numerous successful available treatments. While this protein and its associated diseases have been researched for a long time, there is currently no clinically approved TTR-based biomarker for TTR amyloidosis in blood plasma or cerebrospinal fluid. One considerable issue is distinguishing misfolded TTR from native tetrameric TTR. Our work herein and previously on fluorescent amyloid ligands that can distinguish misfolded fibrillar TTR and native TTR is a development toward that end.<sup>20</sup> Furthermore, detailed knowledge of the binding modes of various small-molecule ligands toward the native TTR tetramer can lead to new kinetic stabilizers as alternatives to diflunisal and tafamidis currently approved as anti-TTR amyloid drugs.<sup>43</sup> The native TTR tetramer with its intrinsic symmetry poses a challenge for X-ray crystallography due to the partial occupancy necessarily observed due to the crystal lattice. In this work, we have therefore complemented our biophysical and structural work with molecular dynamics simulations to establish the most plausible ligand–TTR complex structure at atomic resolution of TTR–Py1SA. The work described in this study represents a

successful methodology for mitigating the issue of partial ligand occupancy.

## ■ ASSOCIATED CONTENT

### 📎 Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.3c02147>.

Experimental and computational details (Section S-1); interim refinement statistics for the analysis of the X-ray diffraction data (Section S-1.4); absorption and emission spectra as well as quantum efficiency plots for Py1SA in PBS in the presence of TTR (Section S-2); detailed analysis of the molecular dynamics simulations (Section S-3); detailed analysis of the umbrella sampling simulations (Section S-4); additional data for the computational part of the study including pdb files of the simulated structures of the two binding modes after the equilibration step can be found in ref 44 (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Per Hammarström** – Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden; [orcid.org/0000-0001-5827-3587](https://orcid.org/0000-0001-5827-3587); Email: [per.hammarstrom@liu.se](mailto:per.hammarstrom@liu.se)

**Carolin König** – Institute of Physical Chemistry and Electrochemistry, Leibniz University Hannover, 30167 Hannover, Germany; [orcid.org/0000-0001-8931-4337](https://orcid.org/0000-0001-8931-4337); Email: [carolin.koenig@pci.uni-hannover.de](mailto:carolin.koenig@pci.uni-hannover.de)

### Authors

**Nghia Nguyen Thi Minh** – Institute of Physical Chemistry and Electrochemistry, Leibniz University Hannover, 30167 Hannover, Germany

**Afshan Begum** – Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden; Present Address: Department of Cell and Molecular Biology, Molecular Biophysics, Uppsala University, 752 37 Uppsala, Sweden

**Jun Zhang** – Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden; Present Address: Department of Chemistry, Umeå University, 901 87 Umeå, Sweden.

**Petter Leira** – Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim, Norway; Present Address: Department of Applied Physics, Institute for Energy Technology (IFE), 1777 Halden, Norway

**Yogesh Tadarwal** – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden

**Mathieu Linares** – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden; Laboratory of Organic Electronics, ITN, Linköping University, PSE-581 83 Linköping, Sweden; Scientific Visualization Group, ITN, Linköping University, SE-581 83 Linköping, Sweden; [orcid.org/0000-0002-9720-5429](https://orcid.org/0000-0002-9720-5429)

**Patrick Norman** – Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of

Technology, SE-106 91 Stockholm, Sweden; [orcid.org/0000-0002-1191-4954](https://orcid.org/0000-0002-1191-4954)

**Dean Derbyshire** – Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden

**Eleonore von Castelmur** – Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University, 581 83 Linköping, Sweden; [orcid.org/0000-0001-7061-4890](https://orcid.org/0000-0001-7061-4890)

**Mikael Lindgren** – Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim, Norway; [orcid.org/0000-0001-6649-7871](https://orcid.org/0000-0001-6649-7871)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcc.3c02147>

### Author Contributions

N.N.T.M.: Molecular dynamics, umbrella sampling study: data curation, investigation, validation, visualization, writing part of the original draft; A.B., D.D., E.v.C.: Generated protein and solved the crystal structure, writing part of the original draft, visualization; J.Z.: Synthesis of Py1SA and binding studies; P.L., M.L.: Photophysical characterization, writing part of the original draft, visualization; Y.T.: Preparation of structure for molecular dynamics simulations, generation and validation of Py1SA force field; MaL, P.N.: Methodology and formal analysis of molecular dynamics simulations, review and editing of the manuscript; P.H.: Project initialization and writing part of the paper, review and editing; C.K.: Supervision and analysis of the molecular dynamics part, writing the original draft, review and editing, and project administration.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge funding by the German Research Foundation (DFG) through the Emmy Noether Young Group Leader Programme (CK, project KO 5423/1-1), the Swedish e-Science Research Centre (SeRC, MaL, PN), the Swedish Research Council (PN, Grant No. 2018-4343; PH, Grant No. 2019-04405), the Swedish Brain Foundation (PH, Grant No. ALZ2019-0004 and ALZ2022-0004), and Gustaf V and Drottning Victorias Stiftelse (PH). Computing resources were provided by the Swedish National Infrastructure for Computing (SNIC), and biophysical instrumentation was used at ProLinC core facility.

## ■ REFERENCES

- (1) Liz, M. A.; Coelho, T.; Bellotti, V.; Fernandez-Arias, M. I.; Mallaina, P.; Obici, L. A narrative review of the role of transthyretin in health and disease. *Neurol. Ther.* **2020**, *9*, 395–402.
- (2) Buxbaum, J. N.; Johansson, J. Transthyretin and BRICHOS: the paradox of amyloidogenic proteins with anti-amyloidogenic activity for A $\beta$  in the central nervous system. *Front. Neurosci.* **2017**, *11*, No. 119.
- (3) Wojtczak, A.; Cody, V.; Luft, J. R.; Pangborn, W. Structures of human transthyretin complexed with thyroxine at 2.0 Å resolution and 3', 5'-dinitro-N-acetyl-L-thyronine at 2.2 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1996**, *52*, 758–765.
- (4) Oza, V. B.; Smith, C.; Raman, P.; Koepf, E. K.; Lashuel, H. A.; Petrassi, H. M.; Chiang, K. P.; Powers, E. T.; Sachetti, J.; Kelly, J. W. Synthesis, structure, and activity of diclofenac analogues as transthyretin amyloid fibril formation inhibitors. *J. Med. Chem.* **2002**, *45*, 321–332.

- (5) Adams, D.; Koike, H.; Slama, M.; Coelho, T. Hereditary transthyretin amyloidosis: a model of medical progress for a fatal disease. *Nat. Rev. Neurol.* **2019**, *15*, 387–404.
- (6) Holmgren, G.; Steen, L.; Suhr, O.; Ericzon, B.-G.; Groth, C.-G.; Andersen, O.; Wallin, B. G.; Seymour, A.; Richardson, S.; Hawkins, P. N.; Pepys, M. B. Clinical improvement and amyloid regression after liver transplantation in hereditary transthyretin amyloidosis. *Lancet* **1993**, *341*, 1113–1116.
- (7) Aimo, A.; Castiglione, V.; Rapezzi, C.; Franzini, M.; Panichella, G.; Vergaro, G.; Gillmore, J.; Fontana, M.; Passino, C.; Emdin, M. RNA-targeting and gene editing therapies for transthyretin amyloidosis. *Nat. Rev. Cardiol.* **2022**, *19*, 655–667.
- (8) Burton, A.; Castaño, A.; Bruno, M.; Riley, S.; Schumacher, J.; Sultan, M. B.; Tai, S. S.; Judge, D. P.; Patel, J. K.; Kelly, J. W. Drug discovery and development in rare diseases: Taking a closer look at the Tafamidis story. *Drug Des., Dev. Ther.* **2021**, *15*, 1225–1243.
- (9) Gillmore, J. D.; Gane, E.; Taubel, J.; Kao, J.; Fontana, M.; Maitland, M. L.; Seitzer, J.; O'Connell, D.; Walsh, K. R.; Wood, K.; et al. CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* **2021**, *385*, 493–502.
- (10) Schonhoft, J. D.; Monteiro, C.; Plate, L.; Eisele, Y. S.; Kelly, J. M.; Boland, D.; Parker, C. G.; Cravatt, B. F.; Teruya, S.; Helmke, S.; et al. Peptide probes detect misfolded transthyretin oligomers in plasma of hereditary amyloidosis patients. *Sci. Transl. Med.* **2017**, *9*, No. eaam7621.
- (11) Kolstoe, S. E.; Mangione, P. P.; Bellotti, V.; Taylor, G. W.; Tennent, G. A.; Deroo, S.; Morrison, A. J.; Cobb, A. J. A.; Coyne, A.; McCammon, M. G.; et al. Trapping of palindromic ligands within native transthyretin prevents amyloid formation. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 20483–20488.
- (12) Takahashi, Y.; Ohashi, N.; Takasone, K.; Yoshinaga, T.; Yazaki, M.; Roberts, M.; Glidden, P. F.; Sekijima, Y. CSF/plasma levels, transthyretin stabilisation and safety of multiple doses of tetracycline in subjects with hereditary ATTR amyloidosis. *Amyloid* **2022**, *29*, 190–196.
- (13) Pinheiro, F.; Pallarès, I.; Peccati, F.; Sánchez-Morales, A.; Varejão, N.; Bezerra, F.; Ortega-Alarcon, D.; Gonzalez, D.; Osorio, M.; Navarro, S.; et al. Development of a highly potent transthyretin amyloidogenesis inhibitor: Design, synthesis, and evaluation. *J. Med. Chem.* **2022**, *65*, 14673–14691.
- (14) Tsai, F. J.; Nelson, L. T.; Kline, G. M.; Jäger, M.; Berk, J. L.; Sekijima, Y.; Powers, E. T.; Kelly, J. W. Characterising diflunisal as a transthyretin kinetic stabilizer at relevant concentrations in human plasma using subunit exchange. *Amyloid* **2022**, 220–224.
- (15) Lindgren, M.; Sörgjerd, K.; Hammarström, P. Detection and characterization of aggregates, prefibrillar amyloidogenic oligomers, and protofibrils using fluorescence spectroscopy. *Biophys. J.* **2005**, *88*, 4200–4212.
- (16) Sörgjerd, K.; Klingstedt, T.; Lindgren, M.; Kägedal, K.; Hammarström, P. Prefibrillar transthyretin oligomers and cold stored native tetrameric transthyretin are cytotoxic in cell culture. *Biochem. Biophys. Res. Commun.* **2008**, *377*, 1072–1078.
- (17) Blake, C.; Geisow, M.; Oatley, S.; Rerat, B.; Rerat, C. Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å. *J. Mol. Biol.* **1978**, *121*, 339–356.
- (18) Sun, F.; Liu, J.; Huang, Y.; Zhu, X.; Liu, Y.; Zhang, L.; Yan, J. A quinoline derived D-A-D type fluorescent probe for sensing tetrameric transthyretin. *Bioorg. Med. Chem. Lett.* **2021**, *52*, No. 128408.
- (19) Zhang, J.; Wang, J.; Sandberg, A.; Wu, X.; Nyström, S.; LeVine, H., III; Konradsson, P.; Hammarström, P.; Durbeej, B.; Lindgren, M. Intramolecular proton and charge transfer of pyrene-based trans-stilbene salicylic acids applied to detection of aggregated proteins. *ChemPhysChem* **2018**, *19*, 3001–3009.
- (20) Campos, R. L.; Wu, X.; Elgland, M.; Konradsson, P.; Hammarström, P. Novel trans-stilbene-based fluorophores as probes for spectral discrimination of native and protofibrillar transthyretin. *ACS Chem. Neurosci.* **2016**, *7*, 924–940.
- (21) Begum, A.; Zhang, J.; Derbyshire, D.; Wu, X.; Konradsson, P.; Hammarström, P.; von Castelmur, E. Transthyretin binding mode dichotomy of fluorescent trans-stilbene ligands. *ACS Chem. Neurosci.* **2023**, *14*, 820–828.
- (22) Gustafsson, C.; Shirani, H.; Leira, P.; Rehn, D. R.; Linares, M.; Nilsson, K. P. R.; Norman, P.; Lindgren, M. Deciphering the electronic transitions of thiophene-based donor-acceptor-donor pentameric ligands utilized for multimodal fluorescence microscopy of protein aggregates. *ChemPhysChem* **2021**, *22*, 323–335.
- (23) Arja, K.; Selegård, R.; Paloncýová, M.; Linares, M.; Lindgren, M.; Norman, P.; Aili, D.; Nilsson, K. P. R. Self-assembly of chiro-optical materials from nonchiral oligothiophene-porphyrin derivatives and random coil synthetic peptides. *ChemPlusChem* **2023**, *88*, No. e202200262.
- (24) Iakovleva, I.; Begum, A.; Pokrzywa, M.; Walfridsson, M.; Sauer-Eriksson, A. E.; Olofsson, A. The flavonoid luteolin, but not luteolin-7-O-glucoside, prevents a transthyretin mediated toxic response. *PLoS One* **2015**, *10*, No. e0128222.
- (25) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125–132.
- (26) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 235–242.
- (27) McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 32–41.
- (28) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355–367.
- (29) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.
- (30) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (31) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (33) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (35) Todorwal, Y.; Gustafsson, C.; Minh, N. N. T.; Ertzgaard, I.; Klingstedt, T.; Ghetti, B.; Vidal, R.; König, C.; Lindgren, M.; Nilsson, K. P. R.; et al. Tau protein binding modes in Alzheimer's disease for cationic luminescent ligands. *J. Phys. Chem. B* **2021**, *125*, 11628–11636.
- (36) Eastman, P. Openmm/pdbfixer: PDBFixer fixes problems in PDB files. <https://github.com/openmm/pdbfixer>.
- (37) Dennington, R.; Keith, T. A.; Millam, J. M. GaussView, Version 6; Semicem Inc.: Shawnee Mission KS, 2019.
- (38) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (39) Leach, A. *Molecular Modelling: Principles and Applications*, 2nd ed.; Pearson, 2001.
- (40) Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932–942.

- (41) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (42) Hamilton, J.; Benson, M. Transthyretin: a review from a structural perspective. *Cell. Mol. Life Sci.* **2001**, *58*, 1491–1521.
- (43) Müller, M. L.; Butler, J.; Heidecker, B. Emerging therapies in transthyretin amyloidosis—a new wave of hope after years of stagnancy? *Eur. J. Heart Failure* **2020**, *22*, 39–53.
- (44) Nguyen, T. M. N.; Todarwal, Y.; Linares, M.; Norman, P.; König, C. Computational data on binding of a pyrene-based fluorescent amyloid ligand (Py1SA) to transthyretin (TTR), 2023 <https://doi.org/10.5281/zenodo.7961594>.

## Recommended by ACS

### Structure–Selectivity Relationship Prediction of Tau Imaging Tracers Using Machine Learning-Assisted QSAR Models and Interaction Fingerprint Map

Maryam Gholampour, Amirhossein Sakhteman, *et al.*

APRIL 10, 2023

ACS CHEMICAL NEUROSCIENCE

READ 

### Mechanistic Insights into the Binding of Different Positron Emission Tomography Tracers to Chronic Traumatic Encephalopathy Tau Protofibrils

Bote Qi, Yu Zou, *et al.*

MARCH 31, 2023

ACS CHEMICAL NEUROSCIENCE

READ 

### Site-Specific C-Terminal Fluorescent Labeling of Tau Protein

Louise Bryan, Michael Mayer, *et al.*

DECEMBER 11, 2022

ACS OMEGA

READ 

### High-Affinity Fluorescent Probes for the Detection of Soluble and Insoluble A $\beta$ Deposits in Alzheimer's Disease

Rathnam Mallesh, Surajit Ghosh, *et al.*

APRIL 04, 2023

ACS CHEMICAL NEUROSCIENCE

READ 

Get More Suggestions >



**Supporting Information to**  
**Binding of pyrene-based fluorescent amyloid**  
**ligand to transthyretin: A combined**  
**crystallographic and molecular dynamics study**

Nghia Nguyen Thi Minh,<sup>†</sup> Afshan Begum,<sup>‡</sup> Jun Zhang,<sup>‡</sup> Petter Leira,<sup>¶</sup>  
Yogesh Tadarwal,<sup>§</sup> Mathieu Linares,<sup>§,||,⊥</sup> Patrick Norman,<sup>§</sup> Dean Derbyshire,<sup>‡</sup>  
Eleonore von Castelmur,<sup>‡</sup> Mikael Lindgren,<sup>¶</sup> Per Hammarström,<sup>\*,‡</sup> and Carolin  
König<sup>\*,†</sup>

<sup>†</sup> *Institute of Physical Chemistry and Electrochemistry, Leibniz University Hannover, Callinstr.  
3A, 30167 Hannover, Germany.*

<sup>‡</sup> *Division of Chemistry Department of Physics, Chemistry and Biology, Linköping University,  
581 83, Linköping, Sweden*

<sup>¶</sup> *Department of Physics, Norwegian University of Science and Technology, 7491 Trondheim,  
Norway*

<sup>§</sup> *Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry,  
Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden*

<sup>||</sup> *Laboratory of Organic Electronics, ITN, Linköping University, SE-581 83 Linköping, Sweden*

<sup>⊥</sup> *Scientific Visualization Group, ITN, Linköping University, SE-581 83, Linköping, Sweden*

E-mail: per.hammarstrom@liu.se; carolin.koenig@pci.uni-hannover.de

## S-1 Methodology

### S-1.1 Fluorescence measurements of Py1SA ligand binding to TTR in solution

The photophysical properties in phosphate-buffered saline (PBS) buffer and with TTR present were examined using steady-state and time-resolved fluorescence. For the former, excitation and emission spectra were collected using typically 1  $\mu\text{M}$  TTR and Py1SA in buffer employing a PTI Quantamaster 8075-22 (Horiba Scientific) equipped with Double Mono 300 spectrometer chambers. Time-resolved fluorescence decays were recorded using an IBH time-correlated single photon counting (TC SPC) spectrometer system using an ns LED operating at 337 nm. For details on experimental procedures and signal processing see refs. 1,2; see caption to Figure 1 in the main text and S-3 for detailed settings. To see how the emission changed upon the binding between the Py1SA ligand and TTR, titrations to approximate the dissociation constant ( $K_d$ ) were performed with increasing concentrations of Py1SA to 0.5  $\mu\text{M}$  TTR tetramer using a fluorescence plate reader (Tecan Infinity M1000). Fitting was performed to the standard hyperbolic function using OriginPro.

### S-1.2 Recombinant expression and purification of TTR

Expression and purification of human TTR were carried out as described previously.<sup>3</sup> Competent *Escherichia coli* BL21 (DE3) cells were transformed with TTR-gene containing pET-3a plasmids and were grown overnight on LB agar plates containing 100  $\mu\text{g}/\text{mL}$  ampicillin. A few single colonies were transferred to LB media supplemented with 100  $\mu\text{g}/\text{mL}$  ampicillin and grown with orbital shaking at 37 °C. The cells were grown until an  $\text{OD}_{600}$  of 0.4 then the temperature was lowered to 20 °C and the cells were grown for another 30 min (to an  $\text{OD}_{600}$  of 0.6) followed by induction with 0.4 mM isopropyl thiogalactopyranoside (IPTG) at 20 °C.

After 18 hours of protein expression, the cell pellet was harvested by centrifugation

and was resuspended in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, (Buffer A) and lysed by sonication. The lysate was cleared by centrifugation at  $25,000 \times g$  for 30 min at 4 °C. The supernatant was heated to 60 °C for 30 min. After heating, the precipitated material was removed by centrifugation at  $14,000 \times g$  for 30 min at 4 °C followed by filtration through 0.45  $\mu\text{m}$  cellulose acetate membrane (Millipore) and applied to a Source-15Q 10/10 ion exchange chromatography column. The column was washed extensively with buffer A and elution was carried out with a linear gradient of 1M NaCl in buffer A. Sample containing TTR was further purified by a subsequent size-exclusion chromatography on a HiPrep 16/60 superdex 75 column (Cytiva) equilibrated with 10 mM Na-phosphate buffer, 100 mM KCl pH 7.6 at 20 °C. Fractions containing pure TTR were collected, pooled, and concentrated using an Amicon Ultra centrifugal filter device (Millipore, 3 kDa molecular-weight cutoff). Protein concentration was determined by using the absorption extinction coefficient  $73,156 \text{ M}^{-1}\text{cm}^{-1}$  at 280 nm applied for tetrameric TTR. Protein quality and purity were accessed by SDS-PAGE prior to experiments. Aliquots of purified TTR were flash-cooled in liquid nitrogen and stored at -80 °C until use.

### **S-1.3 Crystallization of the Py1SA–TTR complex**

The protein was crystallized as described previously.<sup>3</sup> The purified TTR was dialyzed against 10 mM Na-phosphate buffer with 100 mM KCl (pH 7.6) and concentrated to  $5.2 \text{ mg}\cdot\text{mL}^{-1}$  using an Amicon Ultra centrifugal filter device (Millipore, 3 kDa molecular-weight cutoff) and co-crystallized at room temperature with 500  $\mu\text{M}$  concentration of Py1SA added from DMSO stock solutions at 10 mM, using the vapor-diffusion hanging drop method. A drop containing 3  $\mu\text{L}$  protein solution was mixed with 3  $\mu\text{L}$  precipitant and equilibrated against 1 mL reservoir solution containing 1.3–1.6 M sodium citrate and 3.5 % *v/v* glycerol at pH 5.5 in 24-well Linbro-plates. Crystals grew to dimensions of  $0.1 \times 0.1 \times 0.4 \text{ mm}^3$  after 5-7 days. Once fully grown, the crystals were further transferred into a new equilibrated drop containing the same amount of ligand and were incubated for three days. The crystals were

cryo-protected with 12.5 % *v/v* glycerol and to avoid the possibility of the ligand washing out of the crystals during the brief cryo-protection step, the final cryo-solution always contained the same amount of ligand.

#### **S-1.4 X-ray data collection, integration, and processing**

The X-ray diffraction data of Py1SA-TTR were collected under cryogenic conditions at the MAX IV facility (MAXIV), Sweden, using PILATUS detectors at a wavelength of 0.97993 Å. These data were processed to a resolution of 1.4 Å using XDS<sup>4</sup> and AIMLESS from the CCP4 software suite.<sup>5</sup> Data collection statistics are summarized in Table S-I. Phasing was done by molecular replacement using Phaser<sup>6</sup> with a search model derived from the published coordinates 1F41. In short, residues 11–98 and 104–122 were included in the initial model omitting a known flexible region. The model was refined against all the diffraction data using REFMAC.<sup>7</sup> Manual map inspections were performed with COOT.<sup>8</sup> Ligands and solvent were placed in density after 1 to 2 rounds of rebuilding the protein model with COOT and refinement using REFMAC.

**Table S-I: Data collection and interim refinement statistics**

<b>Data collection*</b>		
Resolution	35.61 – 1.4 Å	
Space group	P2 <sub>1</sub> 22 <sub>1</sub>	
Cell parameters: a, b, c (Å)	43.018, 64.434, 85.441	
$\alpha, \beta, \chi$ (°)	90.00, 90.00, 90.00	
Completeness	98.8 (97.8)	
Redundancy	6.5 (6.8)	
Rmerge	0.059 (1.128)	
Rpim	0.025 (0.464)	
I/sigI	13.3 (1.5)	
CC1/2	0.998 (0.627)	
* Values in parentheses are for the highest-resolution shell		
<b>Current model</b> (2 rounds of COOT & REFMAC post ligand placement)	TTR + Py1SA (forward)	TTR + Py1SA (reverse)
Protein	2 chains: 116, 114 residues	2 chains: 116, 114 residues
Waters	158 waters	158 waters
	1 Py1SA molecule	1 Py1SA molecule
R factor	0.1453	0.1447
'free' R factor	0.1865	0.1847
Real-space correlation coefficient	0.966	0.956
G values:		
Dihedrals	-0.18	-0.19
Covalent	0.31	0.31
Overall	0.03	0.02
Estimated coordinate error (DPI)	0.0544	0.0539
Ligand validation (individual):		
Real-space R factor	- / 0.246	- / 0.172
Real-space correlation coefficient	- / 0.766	- / 0.888

### S-1.5 Forcefield parameters for Py1SA

The geometry optimization using the Gaussian (version 16.B.01) program<sup>9</sup> was initially performed at the B3LYP level of theory in combination with the 6-31G(d,p) basis set to identify the lowest energy conformers. Next, the initial forcefield was generated by deriving the RESP charges for the most stable conformer (Conf3 in Fig. S-1) with B3LYP/6-31G\* (as recommended for RESP charges<sup>10</sup>) and incorporating the remaining parameters from the General Amber Force Field (GAFF).<sup>11,12</sup> The equilibrium bond distance and bond angle parameters were further improved based on an optimized structure (B3LYP/6-31G(d,p)). The dihedral potentials of  $\phi_1$  and  $\phi_2$  (marked in Fig. S-1) calculated from Molecular Mechanics (MM) were fitted to the DFT potential (see Fig. S-2). For a detailed procedure about force-field parametrization, refer to supporting information of our previous work.<sup>13</sup>

To validate the forcefield, the ground state energies of conformers computed using MM and DFT methods were compared (see Table S-II). The highest error between the two meth-

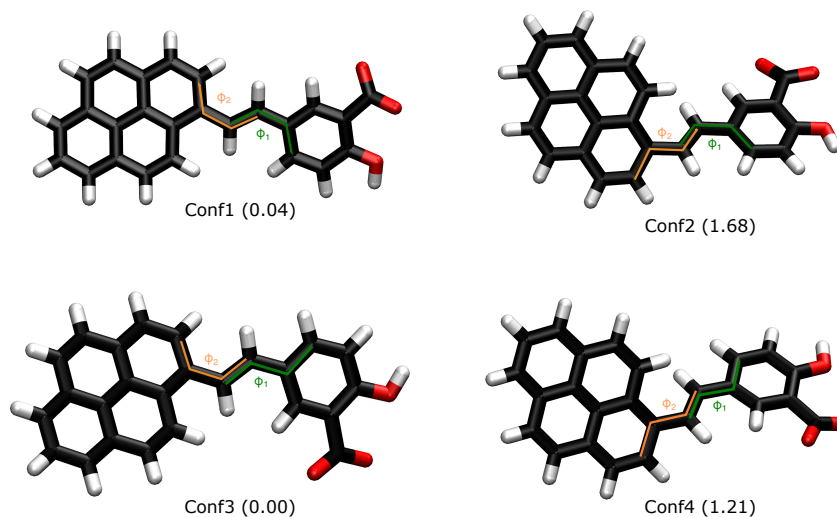


Figure S-1: Optimized molecular structure of Py1SA conformations with labels indicating important dihedrals and relative energy in kcal/mol with respect to the most stable conformer (B3LYP/6-31G(d,p)).

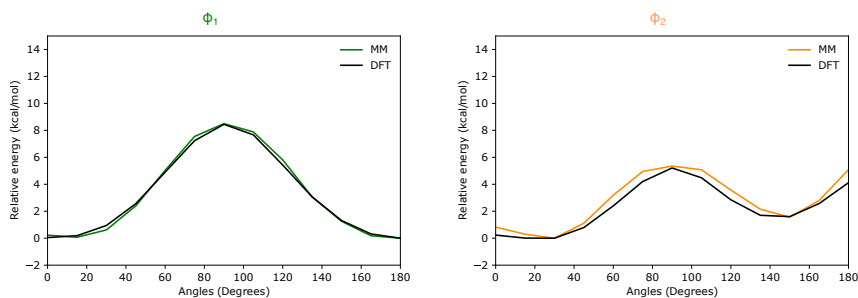


Figure S-2: Energy profiles for relaxed scans varying the dihedral angles  $\phi_1$  (left) and  $\phi_2$  (right) of Py1SA marked in Fig. S-1 with the fitted force field (MM) and B3LYP/6-31G(d,p) (DFT), respectively.

**Table S-II:** Comparison of DFT vs. MM relative ground-state energy of Py1SA conformers optimized with the respective method.

	conf1	conf2	conf3	conf4
MM Relative Energy(kcal/mol)	-0.31	1.70	0.00	1.64
DFT Relative Energy(kcal/mol)	0.04	1.68	0.00	1.21
Error (kcal/mol)	0.35	0.02	0.00	0.43

ods was found to be 0.43 kcal/mol for conf4, which falls within the error range of the DFT method.

### S-1.6 Molecular dynamics simulations

All molecular dynamics (MD) simulations were performed using the Gromacs version 2019.3<sup>14</sup> with the Amber ff14SB force field<sup>15</sup> for the TTR protein and re-parametrized General Amber Force Field (GAFF)<sup>11,12</sup> for Py1SA (see Section S-1.5).

The starting structures of the TTR protein with Py1SA ligand existing in two configurations with respect to protein were based on the X-ray crystallography data from the present study (see Figure 3 in the main text). As a result of electron density being averaged along the AA' BB' symmetry axis, crystallography could not differentiate between four potential binding models for the ligand, i.e., forward-B, forward-B', reverse-B, and reverse-B'. All models were presented as possible starting points for modeling. All initial structures from

the X-ray analysis were cleaned using the module `pdbfixer`,<sup>16</sup> Gromacs<sup>14</sup> utilities, and the Gauss view<sup>17</sup> program by removing water molecules, adding H atoms to the protein and the ligand, and adding three missing amino acid residues to the two chains of the TTR protein.

We used Gromacs tools to solvate the protein–ligand system in a TIP3P<sup>18</sup> water box of size  $7 \times 7 \times 9 \text{ nm}^3$  with a total of approximately 12,000 water molecules. 21  $\text{Na}^+$  ions were then added to obtain system charge neutrality. The long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) method<sup>19</sup> with a long-range cutoff of 1 nm which also is the cutoff of the short-range van der Waals interactions. Default settings were used for Fourier spacing (0.12 nm) and PME order (4). Periodic boundary conditions were applied to all three directions of the simulation box. The energy minimization was then performed using the steepest decent algorithm with a maximum step size of 0.001 nm and a maximum force of  $10 \text{ kJ mol}^{-1}\text{nm}^{-1}$ . Subsequently, all protein atoms except the ones that are within 4 Å from the ligand (4 Å pocket) were under the harmonic constraints with the strength of  $1,000 \text{ kJ mol}^{-1}\text{nm}^{-2}$ . In addition, the LINCS algorithm<sup>20</sup> was used to constrain all bonds. We first performed short 100 ps equilibration in the *NVT* 300 K ensemble using velocity rescaling, followed by a 100 ps equilibration in the *NPT* ensemble at 1 atm and 300 K. The coupling time used is 0.2 ps for the temperature and 1.0 ps for the pressure. Subsequently, the MD simulations were performed in the *NVT* ensemble for 1  $\mu\text{s}$  simulation time. Processing of the different raw data from MD production was done using Gromacs utilities.

### S-1.7 Umbrella sampling simulations

We computed the potential of mean force (PMF) surface using umbrella sampling.<sup>21–23</sup> MD simulations were employed to extract the initial coordinates for binding free energy calculations. We applied the same computational settings as in the molecular dynamics simulations but using Gromacs version 2021.3<sup>14</sup> and except the box size was set to  $7 \times 11 \times 9 \text{ nm}^3$ . We further only constrained the backbone throughout the protein in the US simulations.



Starting from the last snapshot of each initial structure from the equilibration phase, Py1SA was pulled away from its binding site about 2.5 nm; we used a force of 5,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the first (B) and 8,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the second structure (B') of reverse mode, 7,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the first (B) and 8,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for the other structure (B') of forward mode. The Py1SA was pulled at the rate of 0.005 nm/ps. The spring constant as well as the pulling rate in each simulation were carefully selected based on our empirical study so that large overlaps of the histograms were obtained (see Section S-4). From these pulling trajectories, snapshots with equivalent distance of 0.025 nm were set as a starting configuration for each umbrella sampling simulation, which was independently simulated by performing an *NPT* equilibration for 100 ps followed by a 1 ns *NVT* trajectory. The force constant of the umbrella potential is set to 4,000 kJ mol<sup>-1</sup>nm<sup>-2</sup> for all trajectories. Each simulation consists of a total of 101 umbrella windows. Additionally, we added 17 extra windows between 0.2 nm and 0.4 nm for each trajectory of these modes. Thus in total, for these modes, each simulation consists of 118 windows.

Finally, the weighted histogram analysis method (WHAM)<sup>24,25</sup> was used to combine several windows into a PMF curve to estimate the binding free energy. For each initial structure, two pulling simulations followed by umbrella sampling were conducted to get the average PMF curve.

Molecular graphics were prepared using the VMD program.<sup>26</sup>

## S-2 Spectroscopic evidence for Py1SA ligand binding to TTR in solution

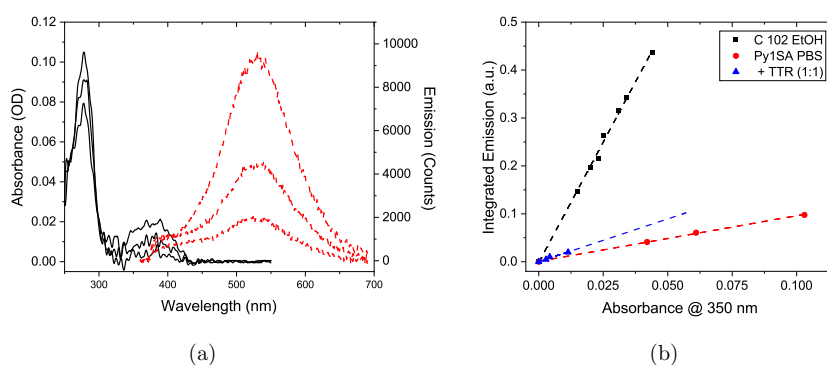


Figure S-3: Absorption and emission spectra and quantum efficiency plots of the Py1SA fluorescent ligand in PBS in the presence of TTR. a) Representative absorption (black) and emission (red) spectra for the Py1SA:TTR system. b) Plots of total emission vs. absorption giving the quantum efficiency from the slope values. Here also data of Py1SA in PBS and the quantum efficiency standard Coumarin 102 (C102) in ethanol, are shown.<sup>27</sup>

### S-3 Molecular dynamics simulations

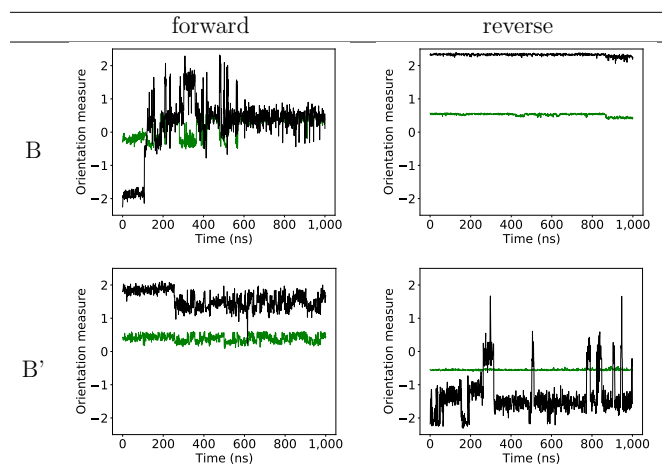


Figure S-4: Projection of vector 1 ( $v_1$ , green, showing the orientation of the salicylic acid group) and vector 2 ( $v_2$ , black, showing the orientation of the pyrene group) on the respective reference vectors of reverse-B conformer during the simulation time as a measure of orientation. For the definition of  $v_1$  and  $v_2$  see Figure 1 (a) in the main text. In none of the trajectories, we observe a sign change in the orientation of both, the salicylic acid group and the pyrene group. That means we observe no rotation of the ligand inside the pocket. The relatively large change in orientation measures in the forward-B trajectory is accompanied by an unbinding/binding process. The observed rotations of the pyrene group in the reverse-B' trajectory represent a rotation of this group at the edge of the pocket, while the orientation of the more buried salicylic acid group is very stable in the binding pocket.

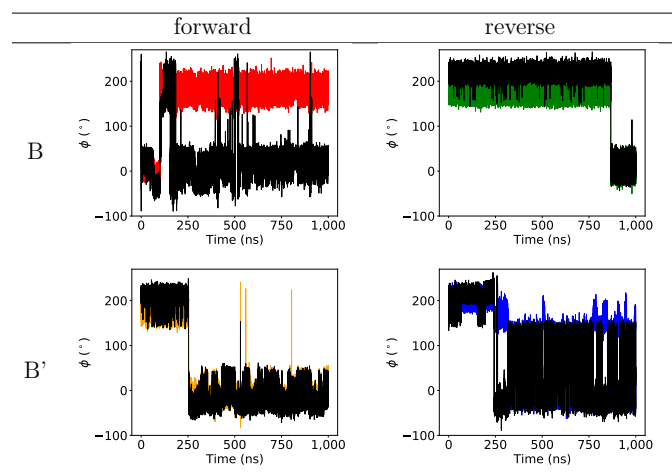


Figure S-5: Dihedral  $\phi_1$  (color) and  $\phi_2$  (black) (for the definition of  $\phi_1$  and  $\phi_2$  see Figure 1 (a) in the main text) in the Py1SA the simulation time.  $\phi_1$  and  $\phi_2$  exhibit few changes over the simulation time. Notably, we observe a tendency to simultaneous switches of both dihedral angles for instance for forward-B' at around 250 ns and for reverse-B and about 850 ns. This simultaneous switch may be favored over individual switches as the molecular shape remains similar during this procedure.

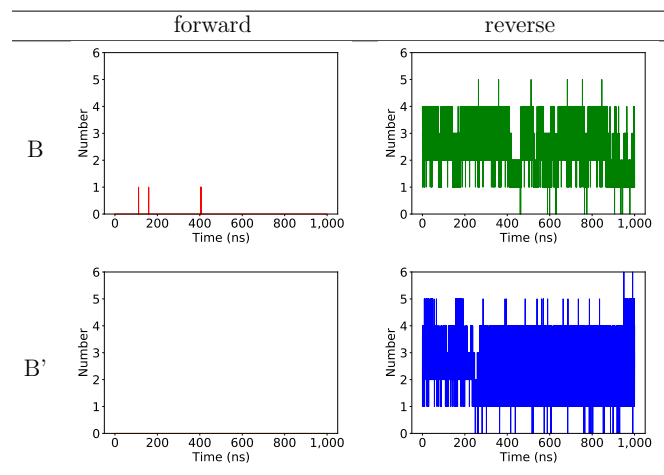


Figure S-6: Number of hydrogen bonds during 1  $\mu$ s simulation (red for forward-B mode, green for reverse-B mode, orange for forward-B' mode, and blue for reverse-B' mode), as defined by the Gromacs utilities. We observe significantly more hydrogen bonds between Py1SA and the 4 Å-pocket for the reverse mode than for the forward mode.

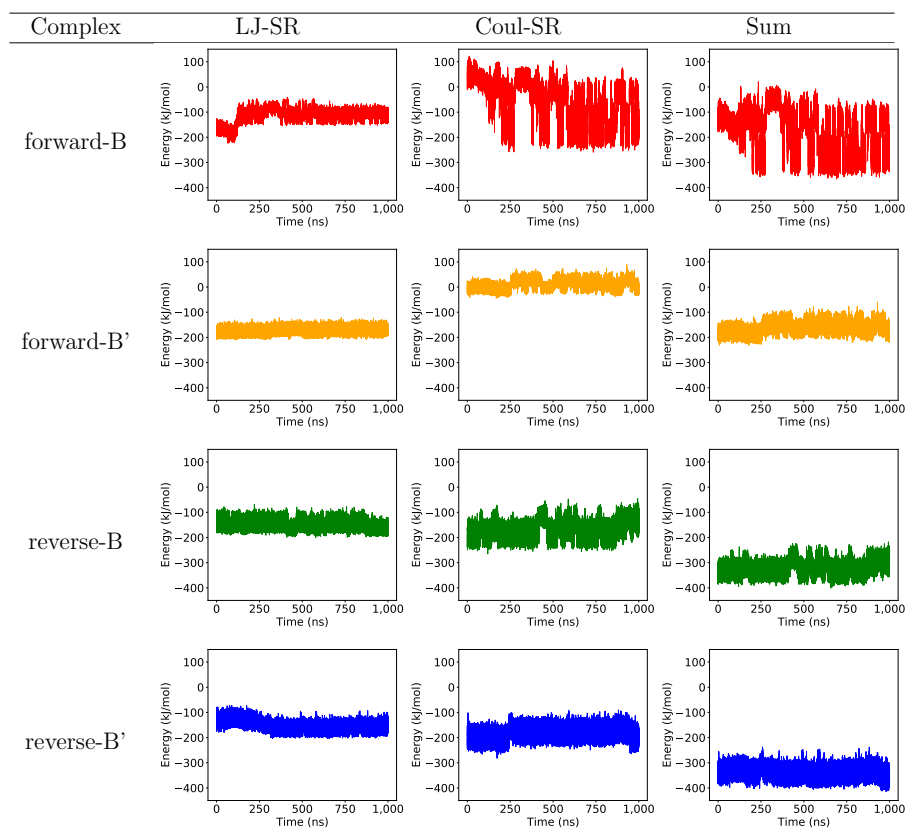


Figure S-7: Short-range interaction (Lennard–Jones short–range (LJ–SR), Coulombic short–range (Coul–SR)) energy between ligand and protein during 1  $\mu$ s MD simulations. While the average attractive contribution to the interaction energy of LJ–SR in forward modes (-116 kJ/mol for forward-B, -175 kJ/mol for forward-B') is greater than Coul–SR potential (-57 kJ/mol and 10 kJ/mol, respectively) the contribution of these interactions in reverse mode is very similar. In general, the overall interaction energy for the forward mode is higher than that for the reverse one by about 160–170 kJ/mol.

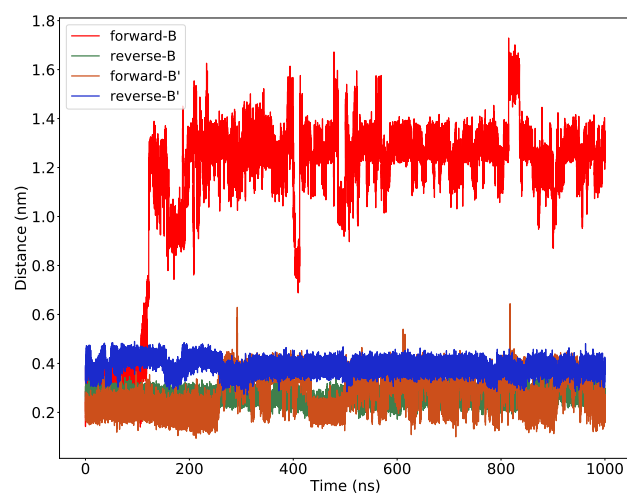


Figure S-8: Distance of the center of mass (COM) of the pocket and the COM of Py1SA along the MD trajectories. In the forward-B mode, there is a noticeable increase in the distance between the pocket's COM and the ligand's COM, which also fluctuates during the MD simulation.

## S-4 Umbrella sampling

We have obtained two PMF trajectories for every starting structure. The results are shown in Figures S-9 – S-12. In our analysis, we focus on the results from the forward-B' and reverse-B'. We disregard forward-B due to rather pronounced Coulombic and hydrogen-bond interaction in the pulling trajectories at distances up to more than 1.5 nm from the original binding pocket (see the first trajectory of forward-B in Figure S-10). The second trajectory of forward-B exhibits Lennard–Jones interaction at about the same distance. In both cases, the PMF shows significantly smaller energetic differences between the minimum and endpoint than for the trajectories for the forward-B' (Figure S-9). These observations suggest that the unbinding process in these two trajectories is not fully finished so we cannot conclude on a free binding energy in case of the pulling trajectories obtained from the forward-B initial structure. Similar Coulombic interaction is also obtained for the second pulling trajectory of the reverse-B' initial structure and first pulling trajectory of reverse-B initial structure (see Figure S-12). In contrast to the first case, the latter case is also accompanied by a smaller energetic difference of the minimum with the endpoint. This is why we disregarded this trajectory.

For the reverse-B' trajectories, we observe a shift in the  $\phi_1$  angle close to the minimum for the pulling trajectories which we do not observe for the remaining pulling trajectory of reverse-B (*c.f.* Figure S-12). However, due to the small number of pulling trajectories, there is no clear evidence for another minor binding mode driven by the  $\phi_1$  angle. If we include the PMF for reverse-B in the bootstrap for the reverse mode, we obtain a binding free energy of  $80\pm 4$  kJ/mol and if we do treat speculative possible minor modes separately, we obtain binding free energies of  $83\pm 5$  and  $78\pm 4$  kJ/mol for the reverse modes. All reverse mode binding free energies are, hence, clearly larger than the  $67\pm 4$  kJ/mol obtained for the forward modes. This leaves the conclusion unchanged that the reverse mode is dominant. In the main text, we only show the PMFs obtained from the forward-B' and reverse-B' initial structures.



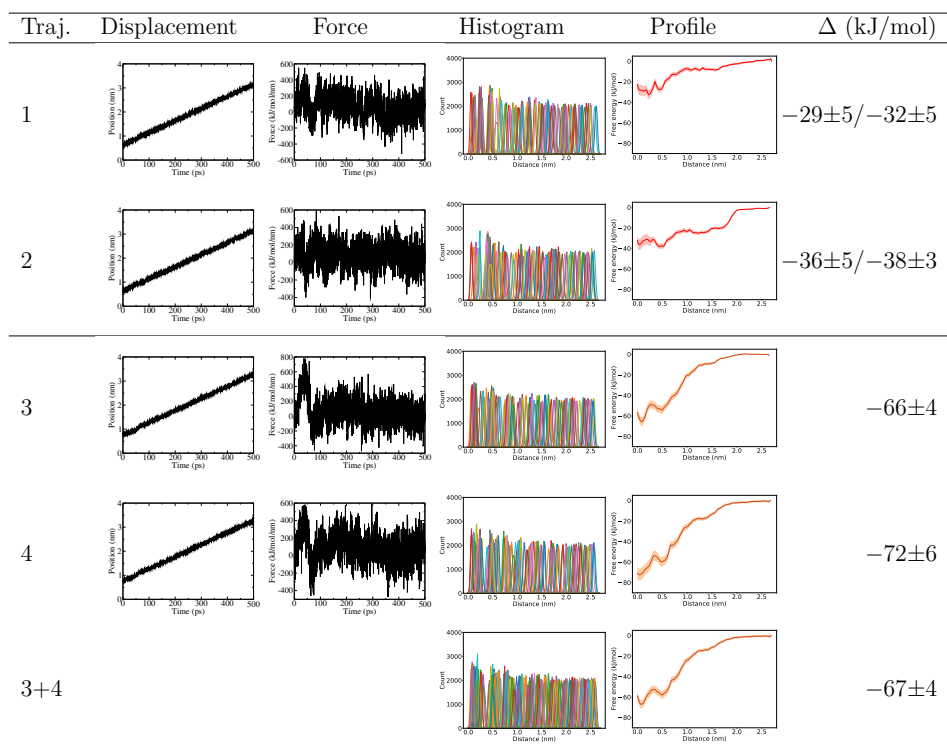


Figure S-9: Displacement, force on the spring over time, bootstrap profile, histogram, and binding free energy ( $\Delta$ , in kJ/mol) of two independent trajectories (Traj.) derived from 118 fully independent sets of umbrella simulations for forward-B (trajectory 1, 2) and B' (trajectory 3, 4) initial structures (pulling force =  $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , pulling rate =  $0.005 \text{ nm/ps}$ , pulling distance =  $2.5 \text{ nm}$ ).

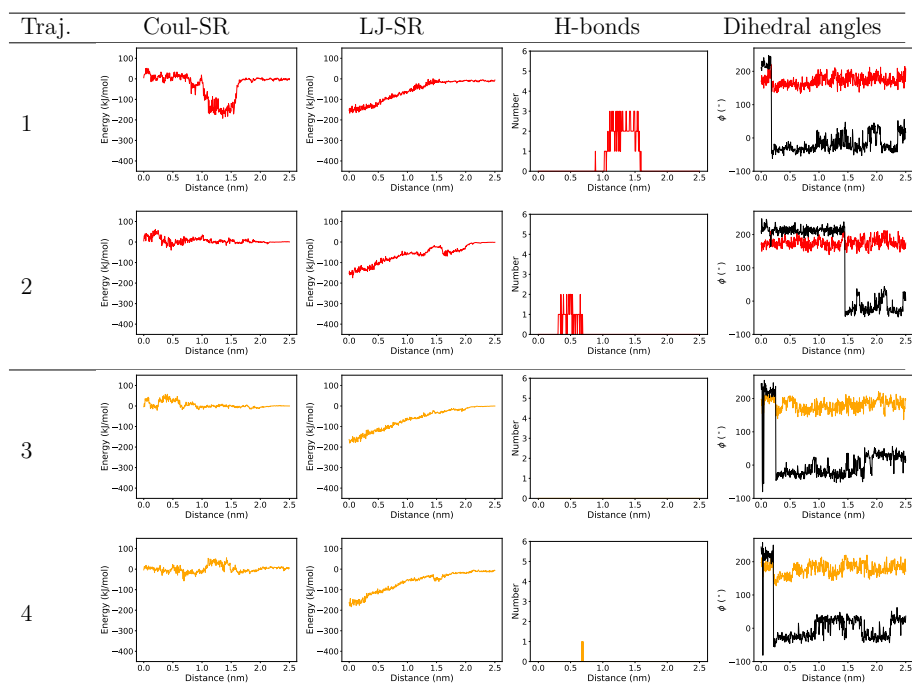


Figure S-10: Coulomb and Lennard–Jones interactions, number of hydrogen bonds between the ligand and protein, dihedral angles  $\phi_1$  (black) and  $\phi_2$  (red) of 2 independent pulling trajectories (Traj.) of forward-B (trajectory 1, 2) and B' (trajectory 3, 4) initial structures (pulling force =  $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , pulling rate =  $0.005 \text{ nm/ps}$ , pulling distance =  $2.5 \text{ nm}$ ).

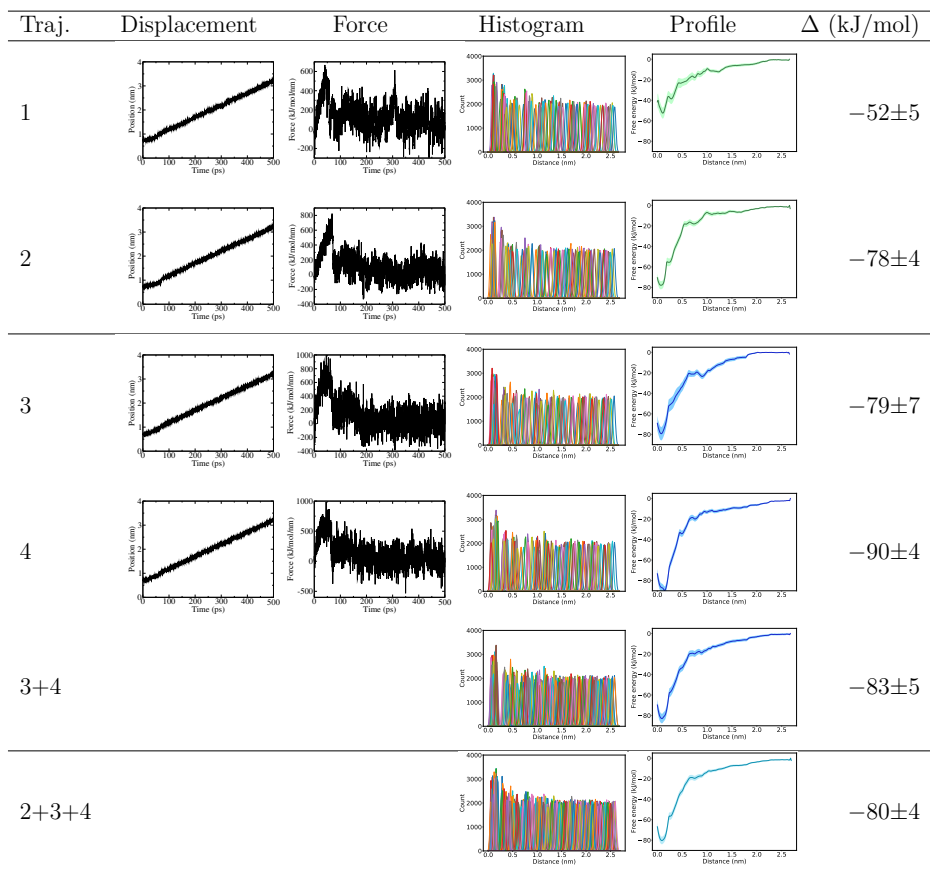


Figure S-11: Displacement, force on the spring over time, bootstrap profile, histogram and binding free energy ( $\Delta$ , in kJ/mol) of two independent trajectories (Traj.) derived from 118 fully independent sets of umbrella simulations for reverse-B (trajectory 1, 2) and B' (trajectory 3, 4) initial structures (pulling force =  $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , pulling rate =  $0.005 \text{ nm/ps}$ , pulling distance =  $2.5 \text{ nm}$ ).

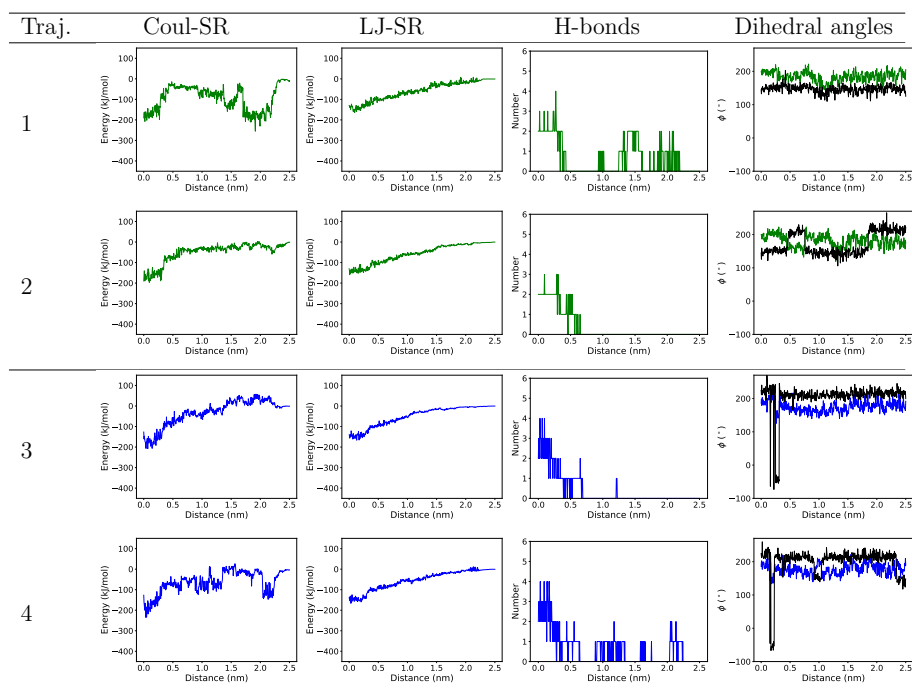


Figure S-12: Coulomb and Lennard–Jones interactions, number of hydrogen bonds between the ligand and protein, dihedral angles  $\phi_1$  (black) and  $\phi_2$  (orange) of two independent pulling trajectories (Traj.) of reverse-B (trajectory 1, 2) and B' (trajectory 3, 4) initial structures (pulling force =  $4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ , pulling rate =  $0.005 \text{ nm/ps}$ , pulling distance =  $2.5 \text{ nm}$ ).

## References

- (1) Gustafsson, C.; Shirani, H.; Leira, P.; Rehn, D. R.; Linares, M.; Nilsson, K. P. R.; Norman, P.; Lindgren, M. Deciphering the electronic transitions of thiophene-based donor-acceptor-donor pentameric ligands utilized for multimodal fluorescence microscopy of protein aggregates. *ChemPhysChem* **2021**, *22*, 323–335.
- (2) Arja, K.; Selegård, R.; Palonc'ová, M.; Linares, M.; Lindgren, M.; Norman, P.; Aili, D.; Nilsson, K. P. R. Self-assembly of chiro-optical materials from nonchiral oligothiopheneporphyrin derivatives and random coil synthetic peptides. *ChemPlusChem* **2023**, *88*, e202200262.
- (3) Iakovleva, I.; Begum, A.; Pokrzywa, M.; Walfridsson, M.; Sauer-Eriksson, A. E.; Olofsson, A. The flavonoid luteolin, but not luteolin-7-O-glucoside, prevents a transthyretin mediated toxic response. *PLOS ONE* **2015**, *10*, e0128222.
- (4) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.
- (5) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242.
- (6) McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 32–41.
- (7) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367.

- (8) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126–2132.
- (9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (10) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (11) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph* **2006**, *25*, 247–260.
- (12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (13) Tadarwal, Y.; Gustafsson, C.; Thi Minh, N. N.; Ertzgaard, I.; Klingstedt, T.; Ghetti, B.; Vidal, R.; König, C.; Lindgren, M.; Nilsson, K. P. R. et al. Tau protein binding modes in Alzheimer’s disease for cationic luminescent ligands. *J. Phys. Chem. B* . **2021**, *125*, 11628–11636.
- (14) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (15) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (16) Eastman, P. Openmm/pdbfixer: PDBFixer fixes problems in PDB files. <https://github.com/openmm/pdbfixer>.

- (17) Dennington, R.; Keith, T. A.; Millam, J. M. GaussView Version 6. 2019; Semichem Inc. Shawnee Mission KS.
- (18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. **1983**, 79, 926–935.
- (19) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. J. Chem. Phys. **1995**, 103, 8577–8593.
- (20) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. **1997**, 18, 1463–1472.
- (21) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comput. Phys. **1977**, 23, 187–199.
- (22) Leach, A. Molecular Modelling: Principles and Applications, 2nd ed.; Pearson, 2001.
- (23) Kästner, J. Umbrella sampling. Wiley Interdiscip. Rev. Comput. Mol. Sci. **2011**, 1, 932–942.
- (24) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J. Comput. Chem. **1992**, 13, 1011–1021.
- (25) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g-wham-A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J. Chem. Theory Comput. **2010**, 6, 3713–3720.
- (26) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. **1996**, 14, 33–38.
- (27) Rurack, K.; Spieles, M. Fluorescence quantum yields of a series of red and near-infrared dyes emitting at 600- 1000 nm. Anal. Chem. **2011**, 83, 1232–1242.





## CURRICULUM VITAE

---

# Thi Minh Nghia Nguyen

## Working Experience

- Since 11/2019 **Teaching assistant**, *Leibniz Universität Hannover, Germany*
- Assisting and correcting the Physical Chemistry laboratory's reports
  - Supervision of students in research internships
  - Tutoring in practical course "Computational Spectroscopy"
- 06/2016 – 09/2017 **Pharmaceutical Scientist**, *Danapha Pharmaceutical JSC, Vietnam*
- Developing Chinese Drugs of Plant Origin in different scales
  - Providing the supervision and performance management of drug production
- 09/2014 – 12/2015 **Research Assistant**, *Universidad Central "Marta Abreu" de Las Villas, Cuba*
- Supervising students in conducting pharmacognosical and pharmacological experiments
  - Performing *in vivo* pharmacological studies and writing scientific reports
- 04/2014 – 05/2014 **Pharmaceutical Research Intern**, *Empresa Laboratorios Farmacéuticos MEDSOL, Cuba*
- Analyzing and assisting all manufacturing activities according to GMP guidelines

## Education

- 11/2019–08/2023 **Ph.D. candidate**, *Leibniz Universität Hannover, Germany*
- Proposing, developing, and implementing novel approaches for the liable calculation and prediction of vibrationally resolved emission spectra of pathological biomarkers
  - Thesis title: Understanding Amyloid Biomarkers by Computational Chemistry
  - Supervisor: Prof. Dr. rer. nat. Carolin König
- 10/2017–07/2019 **The European Master in Theoretical Chemistry and Computational Modelling (8.574 out of 10.0)<sup>1</sup>**, *Universidad de Valladolid, Spain*
- Achieving theoretical and practical foundations of computational chemistry
  - Thesis title: Antioxidant Activity of Serotonin: A Quantum Chemical and Computational Thermodynamics Study (10.0 out of 10.0)<sup>1</sup>
  - Supervisors: Prof. Dr. María Carmen Barrientos, Prof. Dr. Pilar Redondo Cristóbal
- 11/2013–12/2015 **Master in Production of Drugs from Natural Origins (5.0 out of 5.0)<sup>2</sup>**, *Universidad Central "Marta Abreu" de Las Villas, Cuba*
- Focusing on the extraction, isolation, and identification of natural products for use in modern medicines
  - Thesis title: Phytochemical and pharmacological evaluation of the ethanolic extract of *Eugenia clarensis Britton & P.Wilson* (5.0 out of 5.0)<sup>2</sup>
  - Supervisors: Dr. Liliana Vicet Muro and Dr. Dany Siverio Mota
- 09/2010–07/2015 **Bachelor of Science in Pharmaceutical Science (5.39 out of 5.0)<sup>2</sup>**, *Universidad Central "Marta Abreu" de Las Villas, Cuba*
- Achieving theoretical and practical foundations of Pharmaceutical Sciences.
  - Thesis title: Pharmacological evaluation of ethanolic extract of *Eugenia clarensis Britton & P.Wilson* (5.0 out of 5.0)<sup>2</sup>
  - Supervisors: Dr. Liliana Vicet Muro and Dr. Dany Siverio Mota

<sup>1</sup>Grading Scales: 9-10: Outstanding; 7-8.9: Remarkable; 5-6.9: Pass; 0-4.9: Failed.

<sup>2</sup>Grading Scales: 5.0: Excellent; 4: Very Good; 3: Pass; 2: Failed.

---

## Academic Honours

- 12/2019 ○ Master's Degree Extraordinary Award for academic year 2018-2019, *Universidad de Valladolid, Spain*
- 10/2017-06/2019 ○ Full Master Scholarship Programa de Becas Iberoamérica + Asia Banco Santander, *Universidad de Valladolid, Spain*
- 2014 ○ Bronze medal at the National Olympic of Pharmacology, *Cuba*
- 2014 ○ Outstanding Student Awards, *Universidad Central "Marta Abreu" de Las Villas, Cuba*

---

## Professional training

- 09/2022 ○ European Summer School of Quantum Chemistry *Palermo, Italy*
- 12/2021 ○ QM/MM best practices workshop *HPC2N, Sweden*
- 10/2021 ○ Molecular Response Properties Winter School *Stockholm, Sweden*
- 04-07/2020 ○ Scientific writing in English *Kiel University, Germany*
- 12/2019 ○ High-Performance Computing *Kiel University, Germany*

---

## Language and professional skills

### Language

**Vietnamese, native**                      **English, fluent**  
**Spanish, fluent**                              **German, basic**

### Software

**LaTeX, good**                                  **MidasCPP, good**                                  **VMD, good**  
**Turbomole, good**                              **Gromacs, good**                                  **Hyperchem, basic**  
**ORCA, good**                                      **NAMD, basic**  
**Gaussian, good**                                **Autodock, basic**

### Programming

**Python, good**                                  **Bash, good**    **Fotran, basic**

---

## Publications

- Binding of a pyrene-based fluorescent amyloid ligand to transthyretin: A combined crystallographic and molecular dynamics study, *J. Phys. Chem. B*, 2023.
- Book-chapter: Quantum-derived embedding schemes for local excitations, *Chemical Modelling*, 2022.
- Tailored anharmonic-harmonic vibrational profiles for fluorescent biomarkers, *Phys. Chem. Chem. Phys.*, 2022
- Tau protein binding modes in alzheimer's disease for cationic luminescent ligands, *J. Phys. Chem. B*, 2021
- Tamizaje fitoquímico y evaluación de la actividad sobre el sistema nervioso central del extracto etanólico de *Eugenia clarensis* Britton & P. Wilson, *Journal of Pharmacy & Pharmacognosy Research*, 2016

---

## Extra-Curricular Activities

- 2021 –2022 ○ Promotion plus+ qualifiziert (Ph.D. and beyond) program, *Leibniz Universität Hannover, Germany*
  - Project management
  - Corporate entrepreneurship
  - Agile working
  - Team and career development