

UNIVERSITAT DE BARCELONA

Multiscale modelling of light harvesting, catalysis and ligand binding in biosystems

Beste Özaydin

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

MULTISCALE MODELLING OF LIGHT HARVESTING, CATALYSIS AND LIGAND BINDING IN BIOSYSTEMS

MODELITZACIÓ MULTIESCALA DE CAPTURA DE LLUM, CATÀLISI I UNIÓ DE LLIGANDS EN BIOSISTEMES

BESTE OZAYDIN

2025

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

DEPARTAMENT DE FARMÀCIA I TECNOLOGIA FARMACÈUTICA, I FISICOQUÍMICA

PROGRAMA DE DOCTORAT EN
RECERCA, DESENVOLUPAMENT I CONTROL DE MEDICAMENTS

MULTISCALE MODELLING OF LIGHT HARVESTING, CATALYSIS AND LIGAND BINDING IN BIOSYSTEMS

MODELITZACIÓ MULTIESCALA DE CAPTURA DE LLUM, CATÀLISI I UNIÓ DE LLIGANDS EN BIOSISTEMES

Memòria presentada per Beste Özaydin per optar al títol de doctora per la universitat de Barcelona

Carles Eduard Curutchet Barat – Director i tutor

Jordi Juárez Jiménez – Director

Beste Özaydin - Doctoranda

BESTE OZAYDIN
2025

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Javier Luque, who initially believed in me and opened my way to join the CBDD group. I will always remember him with gratitude.

I would like to thank my supervisor, Prof. Carles Curutchet, for everything he patiently did for me, everything he taught me, the trust he placed, and for being a reason to continue believing in goodness in people. His patient, hard-working, professional, and positive attitude, along with his impressive intelligence, will always inspire me.

I also express my gratitude to my supervisor, Dr. Jordi Juárez Jiménez, for bringing new opportunities into my life and brightening my path. His ability to create a positive and productive environment, combined with his guidance and everything he thought me, encouraged me to expand my knowledge and challenge myself, all while maintaining honest communications. As I promised, I will organize a big birthday party for him in 2044, stay tuned for the updates!

I also would like to thank FI scholarship from the Catalan Government, that supported me financially and made this work possible.

Dr. Xavier Barril, even though we didn't share a project, it was a pleasure to work in the same environment and have insightful discussions during the group meetings. His visionary and professional working style inspired me to go further in this field and imagine a brighter future. And, Dr. Carles Galdeano, I'm thankful to him for expanding my knowledge about the experimental part of our field with his expertise.

I would like to thank Prof. Maria João Ramos for contributing to my thesis and allowing me to do my secondment in her lab. She is an exceptional scientist and leader who makes the growth in many aspects possible while enjoying the process. I also thank Prof. Pedro Fernandes and Dr. Rui Neves for their kindness and support during my learning process in the University of Porto.

I am grateful to my previous supervisors, Prof. Viktorya Aviyente, for giving me the initial strength and ambition, together with Prof. Tereza Varnalı, and Prof. Şaron Çatak, for building the foundation that allowed me to grow in this field in Boğaziçi University. My gratitude also extends to Prof. Stephen Hashmi, who supported me during my internship in Germany and inspired me become a more disciplined person.

Of course, I want to thank my dear friends, who were always there for me emotionally and intellectually. Firstly, my group members from the Computational Photobiology Lab: Dani, our drama king, and my first friend here, who made the journey fun; Renato, who made my days with little surprises; Özge, who made me feel at home again; Jhonathan, who turned the office into a lively place; and Blanca, who amazed me with her clear mind and lovely heart. I am grateful for all the support and unforgettable moments we shared.

To the Computational Molecular Design group: Álvaro, for his friendship, help, inspiration and always positive attitude; my dear Varbina and Patricia, for their true friendship, brilliant advices and cheerful vibe; Roger, Carlos, Andrea, Begoña, Jordi for creating a warm and enjoyable environment and for the nice talks we had during the breaks; Salomé, for being a wonderful example of patience through her balance as both a great scientist and a devoted mother; and to Marina, Miriam, Serena, Juan, and Salvo; thank you all for the wonderful memories!

I also want to express heartfelt gratitude to my family: my mom, Handan, who sparked my curiosity for science and believed in me undoubtedly; my father Kamil, who passed me the courage to go further, Ümit, who supported me through every challenge; and my aunts, Nazan, Nalan, and Birsen who always made me feel loved. None of this would have been possible without all the members of my family.

To my friends, Cansel, Pınar, Serap, Yağmur, and Selim, thank you for your priceless friendship and all the deep conversations that helped me understand myself better. Caro, thanks for brightening both my good and bad days with your lovely friendship, and for countless unforgettable memories we had. Semra and Sezen, thank you for creating a safe space to be myself and always supporting me. Ersel, thanks for encouraging me in every obstacle, Robi and İpi, thank you for making life more fun and joyful during all those times.

And finally, Robin, congratulations on surviving my worst version in the last months. All those times your presence was incredibly valuable to me. I deeply want to thank you from my heart for being my source of strength, my partner, and my constant reminder of how beautiful life can be.

And last but not least,

I want to thank me...

I would rather have questions that can't be answered than answers that can't be questioned.

Richard Feynman

ABSTRACT

Biological processes are driven by dynamic mechanisms occurring across various timescales, from molecular interactions to large-scale conformational changes. This thesis explores how multiscale simulation techniques, integrating QM, MM, and QM/MM methods reveal the interplay of biological processes across different temporal and spatial scales, unraveling the dynamic mechanisms behind photosynthetic light harvesting, ligand-receptor interactions, and enzyme catalysis.

In chapter 4 we explore the excitonic properties of photosynthetic pigment-protein complexes, specifically the *Hemiselmis virescens* PC612 and *Chroomonas sp*. PC645, which exhibit open and closed quaternary structures, respectively. Using a dual molecular dynamics (MD) protocol that combines classical MD and Born-Oppenheimer QM/MM simulations, we analyze the influence of thermal dynamics on excitonic interactions and spectral properties. The findings highlight the critical role of thermal fluctuations in shaping the light-harvesting efficiency of PC645, while PC612 demonstrates minimal spectral sensitivity to such effects. These results emphasize the importance of incorporating thermal effects in studying photosynthetic complexes and pave the way for future investigations into their structural and functional dynamics.

In chapter 5 we address the structural and functional characterization of Imidazoline type 2 receptors (I2-IRs), which are emerging as potential therapeutic targets for neurodegenerative procesess like Alzheimer's and Parkinson's diseases. Despite their therapeutic promise, little is known about their structural biology and ligand-binding mechanisms. Combining MDmix simulations, molecular docking, and MD simulations, we identified six new putative binding sites for the canonical I2 ligand 2-BFI across seventeen proteins highlighted in previous proteomics studies. This comprehensive approach not only recovered known binding sites, such as that in monoamine oxidase B (MAO-B), but also revealed novel interaction

hotspots, offering novel insights into the structural biology of I2-IR and paving the ground for future drug development.

Finally in chapter 6 we focus on the catalytic mechanism of dopamine oxidation by MAO-B and its allosteric modulation by 2-BFI. Combining extended MD simulations and ONIOM QM/MM methodologies, we investigated the hydride transfer mechanisms of dopamine oxidation, direct and two-step pathways. The results highlight the critical role of dopamine's orientation, electrostatic interactions, and stabilization in facilitating reaction intermediates. Additionally, 2-BFI was found to alter dopamine's conformational preferences and potentially modulate the catalytic reaction through allosteric mechanisms. These findings contribute to understanding 2-BFI's functional role as both a modulator of MAO-B.

This thesis integrates computational approaches across various scales and biological systems to uncover critical insights into light harvesting, enzymatic catalysis, and ligand-receptor interactions, advancing the understanding of fundamental biological processes and their potential biomedical and biotechnological applications.

KEYWORDS: Light harvesting, enzymatic catalysis, protein-ligand complexes, allosterism, multiscale methods, QM/MM, molecular dynamics

RESUM

Des de les interaccions moleculars fins als canvis conformacionals a gran escala, els processos biològics son deguts a mecanismes dinàmics que ocorren en diverses escales temporals. Aquesta tesi explora com les tècniques de simulació multiescala, integrant mètodes QM, MM i QM/MM, poden descriure processos biològics a diferents escales temporals i espacials, desentranyant els mecanismes dinàmics darrere de la captació de llum fotosintètica, les interaccions lligand-receptor i la catàlisi enzimàtica.

Al capítol 4 s'exploren les propietats excitòniques dels complexos pigment-proteïna fotosintètics PC612 de *Hemiselmis virescens* i PC645 del gènere *Chroomonas*, que presenten estructures quaternàries obertes i tancades, respectivament. Utilitzant un protocol dual de dinàmica molecular (MD) que combina MD clàssica i simulacions QM/MM de Born-Oppenheimer, s'analitza la influència de la dinàmica associada a la temperatura en les interaccions excitòniques i les propietats espectrals. Els resultats contrasten el efecte de les fluctuacions tèrmiques en l'eficiència de captació de llum del PC645, amb la poca sensibilitat del PC612 a aquests efectes. En conjunt, els resultats remarquen la importància d'incorporar efectes tèrmics en l'estudi dels complexos fotosintètics i obren el camí per a futures investigacions sobre la dinàmica estructural i funcional d'aquests complexos.

Al capítol 5 abordem la caracterització estructural i funcional dels receptors d'Imidazolina tipus 2 (I2-IRs), que estan emergint com a possibles dianes terapèutiques per a processos neurodegeneratius com les malalties d'Alzheimer i Parkinson. Malgrat el seu potencial com eines terapèutiques, la seva biologia estructural i els mecanismes d'unió a lligands no son bén coneguts. Combinant simulacions de MDmix, *docking* molecular i simulacions de dinàmica molecular, es van identificar sis nous llocs d'unió d'alta afinitat per al lligand I2 canònic 2-BFI en disset proteïnes destacades anteriorment mitjançant estudis de proteòmica Aquest enfocament integral no només va recuperar llocs d'unió coneguts, com el

de la monoamina oxidasa B (MAO-B), sinó que també va revelar nous possibles punts d'interacció. En conjunt, els resultats d'aques capítol amplien el nostre coneixement sobre la biologia estructural dels I2-IRs i estableixen una base per al futur desenvolupament de fàrmacs.

Finalment, al capítol 6 es centra en el mecanisme catalític de l'oxidació de la dopamina per la MAO-B i la seva modulació al·lostèrica per 2-BFI. Combinant simulacions de dinàmica molecular i la metodologia ONIOM QM/MM, es van investigar els mecanismes de transferència d'hidrur de l'oxidació de la dopamina, mitjançant la via directa i la via en dues etapes. Els resultats destaquen el paper crític de l'orientació de la dopamina, les interaccions electrostàtiques i l'estabilització en la facilitació dels intermediaris de reacció. A més, es va trobar que el 2-BFI altera les preferències conformacionals de la dopamina en el centre actiu del enzim i potencialment modula la reacció catalítica a través de mecanismes al·lostèrics. Aquests resultats contribueixen a comprendre el paper funcional del 2-BFI com a modulador de la MAO-B.

En el seu conjunt, aquesta tesi integra enfocaments computacionals a diverses escales i en diversos sistemes biològics per descobrir nova informació sobre la captació de llum, la catàlisi enzimàtica i les interaccions lligand-receptor, avançant en la comprensió de processos biològics fonamentals i les seves possibles aplicacions biomèdiques i biotecnològiques.

PARAULES CLAU: Captura de llum, catàlisi enzimàtica, complexos proteïna-lligand, alosterisme, mètodes multiescala, QM/MM, dinàmica molecular

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 5
ABSTRACT 11
RESUM 15
ABBREVIATIONS 23
CHAPTER 1. INTRODUCTION 25
1.1 Timescales of biological processes 27
1.2 Overview in light harvesting 30
1.2.1 Light harvesting in cryptophyte algae 33
1.2.2 Exciton spectra and energy transfer in cryptopyte complexes 35
1.2.3 Multiscale modelling of the interplay between structure and exciton dyamics 37
1.3 Overview in ligand binding 38
1.3.1 Ligand protein interactions 40
1.3.1.1 Thermodynamic principles of protein-ligand binding 40
1.3.1.2 Non-covalent chemical interactions in biological systems 42
1.3.2 Imidazoline I2 receptors and ligands 44
1.3.2.1 Experimental studies on imidazoline I2 receptor ligands in neurodegenerative disease research 45
1.3.2.2 Difficulty in identifying I2-IR receptors 48
1.3.3 Current methods for characterizing protein-ligand complexes. 48
1.4 Overview in enzymatic catalysis 51
1.4.1 Catalytic efficiency and contributing factors 53
1.4.2 Current computational techniques to handle catalytic reactions 54
CHAPTER 2. OBJECTIVES 57
CHAPTER 3. METHODOLOGY 61
3.1 Classical methods for structural dynamics of biosystems 63
3.1.1 Pairwise additive force fields 63
3.1.2 Polarizable force fields 65

3.1.3 System minimization....69
3.1.4 Molecular Dynamics simulations....70
3.1.5 Methods to study ligand binding....73
3.1.5.1 Identification of binding sites with mixed-solvent MD simulations....74
3.1.5.2 Molecular docking....75
3.2 Multiscale methods for enzymatic catalysis and light harvesting ...79
3.2.1 General Approach to density functional theory....80
3.2.2 QM/MM methods: Additive and substractive schemes....85
3.2.2.1 Interactions between the layers in ONIOM method: Mechanical, electrostatic and polarizable embedding and boundary conditions....87
3.2.3 Determination of reaction free energy profiles in enzymes....90
3.2.3.1 Characterization of the stationary points....91
3.2.3.2 Establishing the reaction profile....92
3.2.4 QM/MM molecular dynamics....93
3.2.4.1 Born-Oppenheimer molecular dynamics....93
3.3 Excited states from QM/MM TD-DFT....94
3.3.1 QM/MMPol model....95
3.3.2 Excitonic properties from QM/MMPol....99
3.3.2.1 Excitation energies....99
3.3.2.2 Electronic couplings....101
3.3.2.3 Spectral densities....105
3.3.3 Excitonic spectra....109
CHAPTER 4. MULTISCALE MODELLING OF PHOTOSYNTHETIC COMPLEXES....111
4.1 Background....113
4.2 Computational details....114
4.2.1 Simulated systems, MD simulations and geometry optimizations....114
4.2.2 QM/MMPol calculations....116
4.2.3 Modelling of AD and CD pectra....117
4.3 Structural flexibility of open and closed Phycocyanins....118
4.4 Spectral tuning of bilins by the protein scaffold....120
4.5 Absorption and circular dichroism spectra....130

CHAPTER 5. COMPUTATIONAL SCREENING OF PUTATIVE I2 RECEPTORS 135
5.1 Background 137
5.2 Computational details 138
5.2.1 MDmix simulations using mixtures of water and organic solvents 141
5.2.2 Finding the pharmacophoric hotspots 143
5.2.3 Molecular docking 144
5.2.4 MD simulations 144
5.2.5 Protocol validation on the Monoamine Oxidase-B I2 ligand binding site 145
5.3 Virtual screening of putative I2-IR receptors 150
5.3.1 Identification of 2-BFI binding sites in Nischarin (Uniprot: Q80TM9) 152
5.3.2 Identification of 2-BFI binding sites in S-formylglutathione hydrolase (Uniprot: Q9R0P3) 157
5.3.3 Identification of 2-BFI binding sites in alpha-2 adrenergic receptor (Uniprot: Q01338) 161
5.3.4 Identification of 2-BFI binding sites in Heterogeneous nuclear ribonucleoprotein U (Uniprot: Q8VEK3) 165
5.3.5 Identification of 2-BFI binding sites in Hydroxymethylglutaryl-CoA lyase (Uniprot: P38060) 169
5.3.6 Identification of 2-BFI binding sites in creatine kinase – B 173
(Uniprot: Q04447) 173
CHAPTER 6. THE ALLOSTERIC EFFECT OF 2-BFI ON MAO-B 179
6.1 Background 181
6.1.1 Catalysis of dopamine deamination by MAO-B 184
6.2 Computational details 187
6.2.1 MD simulations 187
6.2.2 ONIOM QM/MM calculations 189
6.3 Investigating ligand interactions through MD simulations 192
6.4 Determination of reaction barrier for dopamine hydride transfer with FAD using ONIOM QM/MM methods 201
CHAPTER 7. CONCLUSIONS 213
BIBLIOGRAPHY 219
Appendix 1: Multiscale modelling of photosynthetic complexes 235
Appendix 2: Computational screening of putative I2 receptors 263

ABBREVIATIONS

2-BFI 2-(2-benzofuranyl)-2-imidazoline

AD Alzheimer's disease

BOMD Born-Oppenheimer molecular dynamics

B-CK Brain creatine kinase

CPMD Car–Parrinello molecular dynamics

CG Coarse-grained methods

CNS Central nervous system

Cryo-EM cryogenic electron microscopy

DBV 15,16-dihydrobiliverdin

DFT Density functional theory

EET Electronic energy transfer

EVB Empirical valence bond

FAD Flavin adenine dinucleotide

FCE Full Cumulant Expansion

FEP Free-energy perturbation

FF Force field

FMO Fenna–Matthews–Olson complex

HTS High-throughput screening

I2-IR Imidazoline 2 receptor

KIE Kinetic isotope effect

MAO B Monoamine oxidase B

MBV Mesobiliverdin

MD Molecular dynamics

MC Michaelis Complex

MM Molecular Mechanics

MSM Markov state models

NMR Nuclear magnetic resonance

PBC Periodic boundary conditions

PBP Phycobiliprotein

PBS Phycobilisome

PC Phycocyanin

PCB Phycocyanobilin

PDB Protein Data Bank

PE Phycoerythrin

PEB Phycoerythrobilin

PES Potential energy surface

PPC Pigment-protein complex

QM Quantum mechanics

QM/MM quantum mechanics/molecular mechanics

QM/MMPol Polarizable quantum/molecular mechanics

REMD replica exchange molecular dynamics

ROS Reactive oxygen species

SAXS Small-angle X-ray scattering

TD-DFT Time-dependent density functional theory

TS Transition state

CHAPTER 1. INTRODUCTION

1.1 Timescales of biological processes

Biological systems exhibit remarkable complexity and dynamism, and structural dynamics play a key role in a variety of processes such as enzyme catalysis, signal transduction, photosynthetic light harvesting, and protein-ligand or protein-protein binding. The dynamics of biosystems occur across an extensive spectrum of timescales, from the rapid vibrations of atoms to the slower, complex interactions of cellular processes. The motions influencing these processes include for example molecular vibrations (femtoseconds to picoseconds), side chain fluctuations (picoseconds to nanoseconds), protein conformational changes (nanoseconds to microseconds) or allosteric transitions (microseconds to milliseconds).[1] Understanding biological mechanisms often requires considering the dynamics in multiple timescales. For example, chemical reactions are modulated by slow protein conformational changes, and ultrafast photoinduced processes occurring in the initial steps of photosynthesis are also controlled by protein motions, which can be observed for example in single-molecule spectroscopic experiments. Therefore, understanding the motions in these varied timescales is essential for selecting the appropriate computational approach to study specific biological phenomena effectively.[2] In this thesis, multiscale simulation protocols combining different techniques are applied to investigate the role of dynamics in modulating different biological phenomena, including photosynthetic light harvesting, enzymatic catalysis and ligand-receptor interactions.

Current computational techniques to tackle biological processes

The description of processes like chemical transformations and electronic transitions taking place in the fastest timescales (femtoseconds to picoseconds) require using quantum mechanical (QM) or hybrid quantum mechanical/molecular

mechanical (QM/MM) descriptions. QM methods provide an accurate description of electronic structure and reactivity but are computationally intensive, limiting their application to small systems or short timescales. QM/MM approaches bridge this computational gap by treating the region of interest quantum mechanically while simulating the rest of the system with classical mechanics. These methods enable the study of enzymatic reactions and the photophysics and photochemistry of biological systems, offering insights into reaction mechanisms and electronic properties that are inaccessible to classical molecular dynamics (MD) simulations.[3]

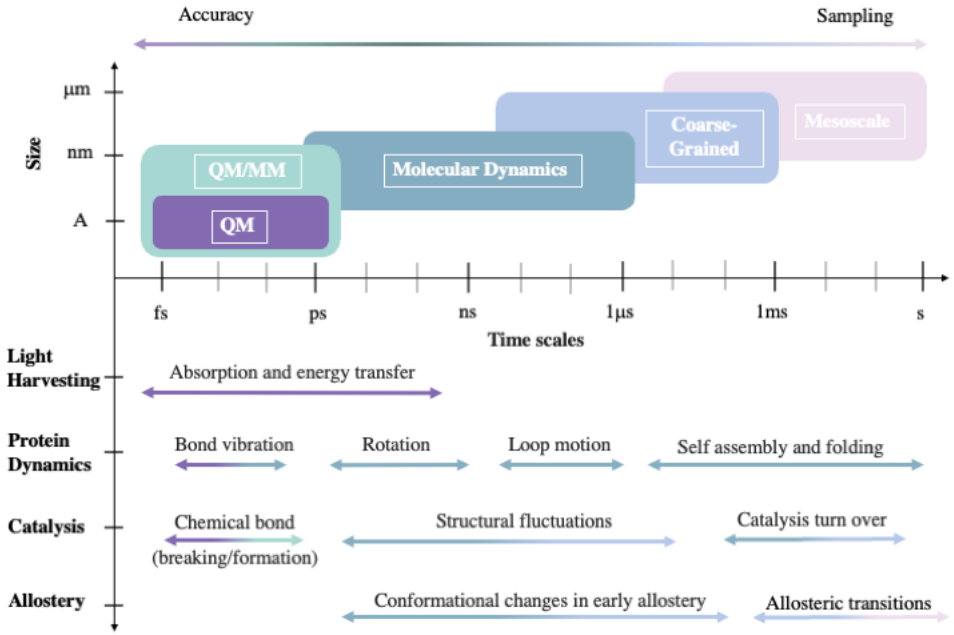


Figure 1.1 Timescales characterizing different biological events and hierarchy of computational methods suited for their study.

In proteins, the dynamics in the picosecond to nanosecond timescales involve for example local conformational changes or side chain rotations, which can be effectively captured by MD simulations based on classical potentials (Figure 1.1). As we progress to nanoseconds and microseconds, the investigation of protein

folding intermediates and ligand binding events necessitates extensive sampling in MD simulations. In microsecond to millisecond timescales, even larger conformational changes, allosteric transitions and enzyme catalysis often require the combination of enhanced sampling MD techniques and multiscale modeling approaches. At milliseconds to seconds, cellular processes and signaling pathways can be studied using coarse-grained simulations, which lack atomic resolution. For even longer timescales (seconds to minutes), metabolic processes and complex cellular responses are typically investigated using systems biology approaches and differential equation models.[4]

It is widely recognized that protein flexibility and conformational diversity are crucial to understand the behavior of pharmaceutically relevant molecules and enzymatic reactions. Experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (cryo-EM) provide essential structural insights and initial configurations for computational studies, offering a foundation for understanding biological processes across multiple temporal scales.[5] However, these methods often fail to capture short-lived or unstructured states that could lead to new drug classes or reveal protein conformations crucial for chemical reactions. MD simulations complement these experimental approaches by providing atomic-level detail and time resolution of conformational changes. MD simulations can theoretically span multiple timescales, from rapid local fluctuations (ps-ns) to slower, functionally relevant motions (ms and beyond). However, computational constraints often limit "brute force" simulations to the shorter timescales. Despite these limitations, brute force simulations offer comprehensive detail with minimal assumptions, requiring only a force field and an initial structure. Ongoing advancements in computing resources, particularly the use of GPUs, are steadily improving the viability and reach of these simulations.[6]

To address the limitations of brute force methods, enhanced sampling techniques have been developed. These approaches boost computational efficiency by introducing additional assumptions and leveraging the separation of timescales, focusing on rare transition events rather than extended waiting periods. Methods like Markov state models (MSMs)[7] and milestoning[8] combine data from shorter simulations to extract long-timescale information. Techniques such as replica exchange molecular dynamics (REMD)[9] allow systems to overcome energy barriers by running multiple simulations at different temperatures, while metadynamics biases specific reaction coordinates to explore free energy landscapes efficiently.[10] Accelerated molecular dynamics (aMD) flattens energy surfaces to encourage transitions between states,[11] and umbrella sampling applies restraining potentials to sample high-energy regions along a reaction coordinate.[12]

In addition, Monte Carlo (MC) methods efficiently sample equilibrium configurations by exploring possible states based on statistical probabilities, though they do not follow real-time dynamics.[13] Coarse-grained (CG) modeling reduces atomic detail by grouping atoms into larger units, allowing simulations of large biomolecular assemblies, such as proteins and lipid bilayers, over longer timescales than atomistic simulations can achieve.[14,15] Mesoscale approaches extend even further, capturing collective behaviors and structural dynamics at cellular scales, such as vesicle formation and cytoskeletal organization, over seconds to minutes.[16]

1.2 Overview in light harvesting

Photosynthesis, the process that sustains nearly all life on Earth, is exceptionally robust at converting sunlight into chemical energy. This process begins with light harvesting, a particularly efficient mechanism where pigment molecules such as chlorophyll, carotenoids and phycobilins absorb photons and transport the energy

from special pigment-protein complexes (PPCs) called antennae to the reaction centers Photosystem I (PSI) and Photosystem II (PSII). These photosystems, embedded in the thylakoid membranes of chloroplasts (in plants and algae) or cyanobacteria, operate sequentially and cooperatively during the light-dependent reactions. PSII captures light energy to extract electrons from water, splitting it into oxygen, protons, and electrons, while PSI further excites these electrons and transfers them to $NADP^+$, forming NADPH. Together, these systems drive the production of ATP and NADPH, the energy carriers required for the later stages of photosynthesis.[17, 18]

Photosynthesis can be categorized into two types: oxygenic and anoxygenic. Oxygenic photosynthesis, carried out by plants, algae, and cyanobacteria, involves three main phases. In the first and second phases, antenna proteins capture and transfer energy to the reaction centers, where it is used to produce ATP and NADPH, as described above. These molecules then power the third phase, which involves the reduction of CO_2 and carbon fixation, resulting in the synthesis of organic compounds essential for life. Anoxygenic photosynthesis, performed by certain bacteria, follows a similar energy transfer process but does not produce oxygen, as it uses electron donors other than water. This highlights the adaptability of photosynthetic mechanisms to diverse environmental conditions.[19,20]

In all photosynthetic organisms, light absorption is done by pigments that are chemically and structurally categorized into three major groups: chlorophylls, carotenoids, and phycobilins. Chlorophylls are essential for the reaction centers and constitute the primary light-harvesting antenna in plants and other organisms like green algae, while carotenoids serve a dual role in capturing light and protecting the photosynthetic apparatus from oxidative damage. Phycobilins, open linear tetrapyrroles found in aquatic organisms such as cyanobacteria, red algae and cryptophyte algae, broaden the range of absorbed light by harvesting wavelengths poorly captured by chlorophyll, such as the green, yellow, and orange regions of the spectrum.

In cyanobacteria and red algae, phycobilins are organized in large multiprotein complexes attached to the thylakoid membrane called phycobilisomes (PBS), which contain different phycobiliproteins (PBP). However, in cryptophyte algae, antenna proteins are located on the lumenal side of the thylakoid and arranged in a random matrix, with each cryptophyte species using only one type of PBP, such as phycocyanin (PC) or phycoerythrin (PE), depending on its habitat and light conditions. The variation in the absorption properties among species then arises from differences in the chromophore composition of their antenna, which have evolved to absorb light in the green, yellow, orange, or red parts of the visible spectrum. This complements the absorption range of Chl a, ensuring efficient light harvesting across a broader spectral window. Indeed, chlorophyll's effectiveness is limited due to its absorption properties. It primarily absorbs light in the red (around 680–700 nm) and blue (430–450 nm) regions, with poor absorption in the green spectrum, which is abundant underwater or in shaded environments. In contrast, phycobiliproteins complement chlorophyll by harvesting light in these less-effective wavelengths, ensuring efficient photosynthesis in underwater habitats, where green light penetrates more effectively, while terrestrial plants rely more on chlorophyll due to the abundance of red and blue light.[17,20]

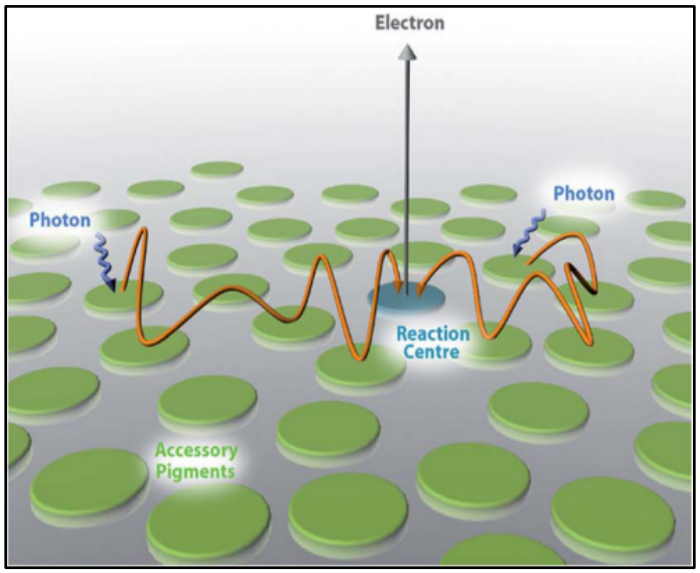


Figure 1.2: Energy transfer in photosynthetic complexes via phycobiliproteins (Figure adapted from [21]).

1.2.1 Light harvesting in cryptophyte algae

Cryptophyte algae live in marine and freshwater environments where the available light significantly drops, affecting the light absorption efficiency more than land plants.[22] To overcome this difficulty, these algae use PBPs to harvest light in the 450–640 nm spectrum, with these proteins containing eight bilins covalently attached to the protein scaffold via single or double cysteine bonds. The algae then adapt their light-harvesting capabilities to specific habitats by tuning their antennae through variations in their sequences and their bilin composition. In cyanobacteria and red algae, PBPs are arranged as $(\alpha\beta)_3$ trimers in the phycobilisomes, which contain different PBPs to absorb light at different wavelengths. In contrast, cryptophytes display greater spectral diversity despite that each species contains only a single PE or PC antenna. Another difference is that PBPs are located within the thylakoid lumen and organized as $\alpha_1\beta{\cdot}\alpha_2\beta$ homo or heterodimers composed of two identical 18–20 kDa β-subunits and two smaller α-subunits, weighing between

8–10 kDa.[23] Interestingly, the pigments in these antenna complexes exhibit rather large interpigment separations compared to antennae of other photosynthetic organisms, which still allows them to achieve optimal photosynthetic activity at low-light conditions.

After the first cryptophyte PBP structure was elucidated, the spectral and energy transfer properties became the central focus of numerous experimental and theoretical studies, which aimed to understand how the specific arrangement and interactions among pigments within the protein matrix influence light absorption and energy transfer dynamics, ultimately shedding light on the efficiency and functionality of these essential photosynthetic components.[20,21,24–27] The presence of multiple types of bilins indeed provides unique characteristics to this antennae, as modulation of the bilin protonation states and their flexible conformations by the protein environment allows additional ways to optimize light harvesting that those found in antennas like LH2 in anoxygenic purple bacteria and FMO in green sulfur bacteria, where the absorption spectra are primarily tuned through exciton pigment-pigment coupling or pigment-protein interactions, due to the presence of a single type of chlorophyll pigment.

In addition to their unique structural organization, cryptophyte PBPs exhibit quaternary structure variations. Recent discoveries in the Hemiselmis lineage revealed a novel open quaternary arrangement, characterized by the insertion of a single aspartic acid residue in the α subunit sequence.[28] This insertion causes a ~73° rotation of the two αβ protomers, significantly decreasing the electronic interaction between central bilin pigments. This structural shift has been linked to a decrease in observed quantum beatings in 2D electronic spectra, potentially affecting vibronic coherence in energy transfer mechanisms, which underscores the critical role of structure in modulating energy transfer pathways and dynamics.

Moreover, sequence variations in the α subunits contribute to additional spectral tuning mechanisms, allowing cryptophytes to adapt to diverse light conditions. X-ray crystallography and multiscale simulations have provided insights into how these adaptations optimize light harvesting.[29,30] However, uncertainties remain regarding the precise location of exciton states, particularly in closed phycocyanins like PC630 and PC645, which hinders quantitative studies on the impact of α subunit sequences and quaternary structures on the excitonic structure of cryptophyte PBPs.

This thesis specifically investigates the development of improved multiscale protocols to achieve an accurate description of the influence of pigment-protein interactions and thermal fluctuations in shaping the absorption properties of the PC612 and PC645 cryptophyte complexes. By studying the interrelation between exciton spectra and structural dynamics, it aims to deepen our understanding of how photosynthetic organisms have achieved such high energy transfer efficiency and adaptability, which provide a foundation for bioinspired approaches to artificial light-harvesting technologies.[31]

1.2.2 Exciton spectra and energy transfer in cryptopyte complexes

The light harvesting properties of PPCs have been a central focus of research since the high-resolution structure of the Fenna–Matthews–Olson (FMO) complex was determined 40 years ago.[20] Beyond their fundamental interest, PPCs also serve as ideal models to explore the physical principles governing energy transfer in multichromophoric systems due to the rich structural diversity of PPCs evolved to adapt the photosynthetic machinery to different habitats, reflecting their remarkable environmental adaptability.

The absorption of sunlight in these multichromophoric complexes generates excitons, collective excited electronic states generated by strong pigment-pigment

interactions that are highly sensitive to thermal fluctuations within the PPC. These excitons then migrate through the antenna protein network to reaction centers, where photochemical reactions occur. Understanding exciton dynamics, and how thermal structural fluctuations influence them, is essential to reveal the mechanisms behind photosynthetic efficiency. On short timescales (femtoseconds to picoseconds), pigment's intramolecular vibrations induce transient shifts in electronic states, altering the energy landscape and facilitating ultrafast equilibration, giving rise to dynamic disorder effects that lead to homogeneously broadened spectra. At longer timescales (nanoseconds to microseconds and beyond), protein motions give rise to static disorder effects, which means light harvesting and energy transfer vary across the microscopic ensemble of slowly interconverting protein conformations, leading to inhomogeneous broadening in their spectra. While the protein scaffold generally stabilizes and orients pigments for efficient energy transfer, these slower fluctuations may alter spatial organization, modulating electronic couplings and exciton migration pathways, as recently observed in single molecule experiments of cryptophyte complexes.[32] These dynamic adaptations ensure efficient energy funneling to reaction centers, even under varying conditions. To capture the impact of both dynamic and static disorder in excitonic spectra and dynamics, it is essential to adopt smart multiscale modeling approaches that combine computational techniques across different timescales.

In cryptophyte PBPs, ultrafast energy transfer occurs across distinct timescales: ultrafast exciton relaxation within strongly coupled bilins (<1 ps), intermediate energy transfer between neighboring bilins (1–8 ps), and final energy transfer to the lowest-energy bilins (20–50 ps), before funneling energy to Photosystem I and II.[33,34] As discussed before, sequence variations in α subunits and quaternary structural changes allow cryptophytes to fine-tune their light-harvesting capabilities. However, uncertainties still remain regarding the precise location of

exciton states, which poses a challenges for accurate quantitative studies of light harvesting.

1.2.3 Multiscale modelling of the interplay between structure and exciton dyamics

Thermal fluctuations span a broad range of timescales, from femtoseconds to milliseconds, as discussed in Section 1.1, each uniquely influencing exciton structure and dynamics. Among them, the rapid intramolecular vibrations experienced by the pigments on short timescales (femtoseconds to picoseconds) are key to accurately describe the energies of electronic excited states, but the later are also tuned by slow protein motions. Multiscale protocols in the field usually combine classical MD simulations with subsequent calculation of the excited state energies using QM/MM calculations along selected frames of the trajectories to tackle these effects.

Recent research from our group has attempted to improve such protocols. After clarifying the protonation properties of various bilins in cryptophyte complexes, including 15,16-dihydrobiliverdin (DBV), phycocyanobilin (PCB), mesobiliverdin (MBV), and phycoerythrobilin (PEB),[35] multiscale polarizable quantum/molecular mechanical (QM/MMPol) calculations where used to characterize the exciton structure in PC577, PC612, PC630 and PC645 by exploring different strategies to derive the input geometries used for excited state calculations.[29] Important inaccuracies where found when geometries are derived from classical MD simulations. Alternatively, QM/MM geometry optimization of the pigments in the complexes significantly improved results in PC577, PC612, PC630, and PC645, albeit at the cost of neglecting thermal effects. Discrepancies with experimental results remained, such as unbalanced descriptions of the relative intensity of low and high-energy bands in PC577 and PC612, or a strong underestimation of the MBV bands in PC630 and PC645, highlighting the impact

of thermal effects and the limitations of rationalizing spectral properties based on X-ray crystal structures.

In this thesis, we aim at providing an improved protocol based on combining classical MD simulations with multiple short QM/MM Born-Oppenheimer Molecular Dynamics (BOMD) simulations, which allows to capture both dynamic and static disorder effects in an efficient manner. In the context of phycocyanin complexes, such vibrations are critical as they influence the absorption spectra and the dynamics of exciton formation and decay, as they lead to transient changes in the electronic states of the pigments.

1.3 Overview in ligand binding

Protein regulation by small molecules occurs through two main types of binding: orthosteric and allosteric. Orthosteric binding happens at the protein's main functional site, where natural ligands and substrates attach and directly influence functions like enzymatic reactions. Orthosteric ligands compete with these substrates. On the other hand, allosteric binding occurs at secondary sites different from the primary site. Allosteric modulators cause conformational changes that indirectly alter protein activity (Figure 1.3). Since they don't compete with substrates, they offer a more precise regulation, with less risk of side effects. However, their effects can be harder to predict because of the subtle structural changes they induce, whereas orthosteric ligands face selectivity challenges due to the conserved nature of active sites.[36]

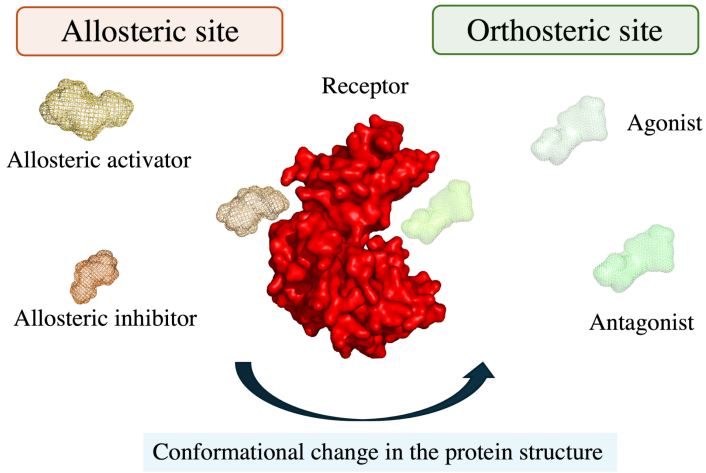


Figure 1.3 Illustration of the orthosteric site accommodating natural substrates, agonists, or inhibitors, and the allosteric site inducing conformational changes that enhance or inhibit protein function.

Three primary models are used to describe protein-ligand binding: lock-and-key, induced fit, and conformational selection. Fischer's lock-and-key model (1894) considers both the protein and ligand as rigid entities with complementary shapes, binding only when they match, a simplified approach that doesn't account for protein flexibility. Koshland's induced fit model (1958) introduces flexibility, suggesting the binding site adjusts to fit the ligand upon formation of the complex, creating stabilizing non-covalent interactions and explaining how proteins can accommodate buried ligands, though it doesn't fully capture large-scale conformational changes.[37] The conformational selection model, proposed in the 1990s, posits that proteins exist in multiple conformational states before binding, with each state's probability determined by its free energy. Ligands bind to pre-existing compatible conformations, shifting the population toward these states, reflecting both enthalpic and entropic contributions. This model underscores the dynamic nature of proteins, particularly in molecular recognition and allosteric modulation.

1.3.1 Ligand protein interactions

The interaction between proteins and ligands is often highly specific and is always mediated by a combination of hydrophobic and electrostatic interactions that can be described in the bases of thermodynamic principles.

1.3.1.1 Thermodynamic principles of protein-ligand binding

In aqueous solution and at constant pressure and temperature, the affinity of a ligand for a protein is described by the binding constant (K_a), which quantifies the equilibrium between the concentration of the protein-ligand complex ([PL]) and the unbound receptor ([P]) and ligand ([L]). The binding constant is expressed as:[38]

$$K_a = \frac{[\mathrm{P}L]}{[\mathrm{P}][L]} \qquad \text{Equation 1.1}$$

In drug design, the dissociation constant (K_d), defined as the inverse of K_a, is widely used to measure ligand affinity. It allows for comparisons between ligands and facilitates the derivation of structure-activity relationships (SAR). The dissociation constant is defined as:

$$K_d = \frac{\mathbf{1}}{K_a} = \frac{[\mathrm{P}][L]}{[\mathrm{PL}]} \qquad \text{Equation 1.2}$$

The association and dissociation processes are governed by kinetic parameters, specifically the on-rate constant (k_{on}) and the off-rate constant (k_{off}), with units of $M^{-1}s^{-1}$ and s^{-1}, respectively. These parameters define the equilibrium between binding ($P + L \rightarrow \mathrm{P}L$) and unbinding ($\mathrm{P}L \rightarrow P + L$) reactions. At equilibrium, the relationship is given by:

$$k_{on}[\mathrm{P}][\mathrm{L}] = k_{off}[\mathrm{PL}] \qquad \text{Equation 1.3}$$

and the binding constant can be expressed as:

$$K_a = \frac{k_{on}}{k_{off}} \qquad \text{Equation 1.4}$$

The dissociation constant (K_d) is related to the Gibbs free energy of binding (ΔG_{bin}) through the following equation:

$$\Delta G_{bin} = RTln\ \frac{K_d}{c^{\ominus}} \qquad \text{Equation 1.5}$$

Here, R is the ideal gas constant, T is the absolute temperature, and $c^{\ominus}$=1 mol/L is the standard reference concentration. The Gibbs free energy of binding can further be decomposed into enthalpic (ΔH_{bin}) and entropic (ΔS_{bin}) contributions, as expressed in the thermodynamic equation: [38]

$$\Delta G_{bin} = \Delta H_{bin} - T\Delta S_{bin} \qquad \text{Equation 1.6}$$

For spontaneous binding, ΔG_{bin} must be negative. ΔH_{bin} reflects the strength of non-covalent interactions like hydrogen bonds, van der Waals forces, salt bridges. A negative ΔH_{bin} indicates an exothermic process, where energy is released, and the interactions favor the formation of the complex. In contrast, a positive ΔH_{bin} reflects endothermic binding, where energy is absorbed, and interactions are unfavorable for the formation of the complex. On the other hand, entropy (ΔS_{bin}) measures the change in the number of possible microscopic configurations for a given macroscopic state that happen upon the formation of the ligand-protein complex. A positive ΔS_{bin} reflects an increase in the number of configurations and is favorable to binding. While both protein and ligand experience a reduction in the accessible microscopic configurations upon binding, which opposes the formation of the complex, it is often the case that this effect is counteracted by the release of water molecules from the solvation shells of the ligand and the protein, leading to an overall positive change in entropy.[37] Hence, binding can be enthalpy-

driven (dominated by strong interactions, resulting in negative ΔH_{bin} or entropy-driven (where the release of water molecules increases ΔS_{bin}. The concept of enthalpy-entropy compensation proposes that favorable enthalpy is often countered by unfavorable entropy, and vice versa. In cases where enthalpy is the primary driver, ligands are designed to form strong non-covalent interactions with the target protein, resulting in high specificity and stability. However, achieving such enthalpy-driven interactions is challenging, as they often involve compensatory entropy penalties.[37]

1.3.1.2 Non-covalent chemical interactions in biological systems

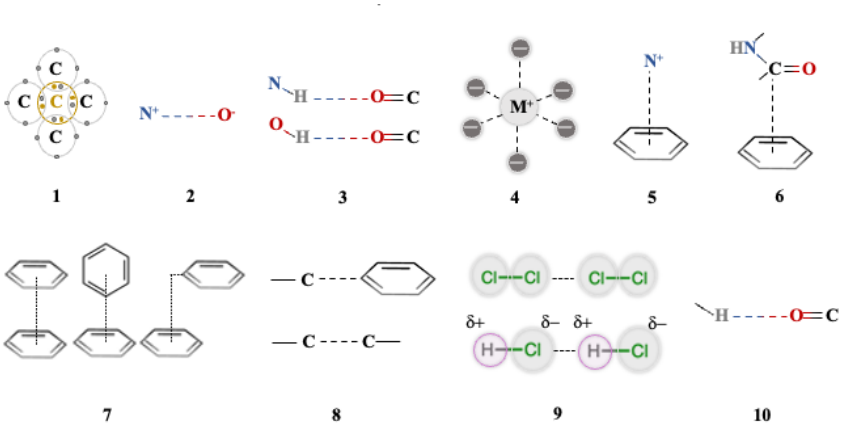


Figure 1.4 Chemical interactions that allow ligand binding to proteins: (1) Covalent bond, (2) ionic interaction, (3) hydrogen bond, (4) coordinate covalent bond (metal-ligand interaction), (5) π-cation interaction, (6) amide - π stacking, (7) π-π stacking, (8) hydrophobic interaction, (9) dipole-dipole (10) weak H bonds.

Non-covalent interactions, unlike (1) covalent bonds that involve the sharing of electron pairs, are governed by electrostatic and van der Waals forces (Figure1.4). These interactions are crucial for biological flexibility and specificity, including ionic bonds, hydrogen bonds, hydrophobic interactions, π-π, cation-π, and halogen bonds, enabling dynamic processes such as enzyme-substrate binding and protein-

ligand recognition.[39] **(2) Ionic bonds and salt bridges**, such as those between carboxylates and amino groups, stabilize protein structures and facilitate ligand binding. They contribute strongly to binding energies, with typical distances of 2.7–3.0 Å. Common amino acids forming salt bridges include Asp or Glu (negatively charged), and Arg, Lys or His (positively charged). **(3) Hydrogen bonds**, between electronegative atoms such as nitrogen and oxygen, are critical for stabilizing structures like α-helices, typically span 2.6–3.2 Å. **(4) Metal-ligand interactions** form through the donation of lone pair or π electrons from the ligand, with bond distances typically ranging from 1.8 to 3.0 Å, depending on the metal and ligand types. **(5)Cation-π interactions**, established between positive charges and aromatic rings, like lysine-phenylalanine, stabilize binding around 3.6 Å, especially in buried environments. (6)**Amide π stacking** is in the range of 3.2-3.7 Å. **(7) π-π interactions** between aromatic rings occur at distances of 3.4–3.8 Å, commonly involving residues like Phe, Tyr, His and Trp. These interactions typically adopt T-shaped (edge-to-face) or parallel-displaced stacking arrangements, enhancing stability and reducing repulsion. **(8) Hydrophobic interactions** stabilize protein folding and ligand binding, with van der Waals distances of 3.7–4.4 Å. **(9) Dipole-dipole interactions** involve electrostatic attractions between the partial positive (δ^+) and partial negative (δ^-) charges of molecules with permanent dipoles. Induced dipole interactions occur when the electron cloud of an apolar molecule is temporarily distorted by a nearby polar molecule or ion, creating weak van der Waals forces that depend on polarizability and act over short distances. **(10) Weak hydrogen bonds** (e.g., CH···O, 3.4–3.7 Å) subtly stabilize ligand binding. **Additionally, halogen bonds**, involving Cl, Br, or I, enhance binding specificity. Typical distances are 2.9-3.5 Å. **Sulfur interactions**, such as methionine-aryl π contacts (3.6–4.0 Å), also contribute to both hydrophobic and polar stabilization.[39]

1.3.2 Imidazoline I2 receptors and ligands

The imidazoline receptor family comprises three main subtypes—I1, I2, and I3—each associated with distinct physiological roles and potential therapeutic applications. I1 receptors (I1-IRs) are mainly found in the brainstem and peripheral tissues, where they contribute to blood pressure regulation.[40] Clinically, selective I1 ligands, such as moxonidine and rilmenidine, are used to lower blood pressure.[41] I3 receptors (I3-IRs), on the other hand, are present in pancreatic β-cells and play a role in insulin secretion, offering potential therapeutic applications for diabetes.[42] Imidazoline I2 receptors (I2-IRs) are unique nonadrenergic binding sites that are mostly identified by radioligands such as [3H]-idazoxan and [3H]-p-aminoclonidine, with [3H]-clonidine displaying the lowest affinity.[43] Autoradiography with [3H]-idazoxan has revealed higher densities of these receptors in the cortex, hippocampus, basal ganglia, and brainstem. Additionally, these receptors are located throughout the central and peripheral nervous systems, as well as in various organs and tissues.[44] Alterations in I2-IR expression are linked to a range of neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease, Huntington's disease, depression, and certain glial tumors. High-affinity I2-IR ligands have been shown to influence physiological and pathological processes like pain relief and inflammation. Notably, two I2-IR ligands, CR4056 and [13C]BU99008, are currently being evaluated in clinical trials, underscoring the therapeutic potential of these receptors. CR4056, the first-in-class analgesic targeting I2-IR, is in phase II trials for osteoarthritis and postoperative dental pain. Meanwhile, [13C]BU99008 has entered phase I trials for PET imaging in Alzheimer's disease patients.[45]

Although the therapeutic potential of I2-IRs, their pharmacological profile remains poorly defined, largely due to their structural diversity and limited detailed characterization. Insights into their pharmacological role are primarily derived

from studies involving I2-IR ligands and the physiological responses they trigger. At present, there are only a few known I2-IR ligands, most of which share a similar structural motif centered on the 4,5-nonsubstituted-2-heterocyclic-2-imidazoline nucleus (e.g., idazoxan, tracizoline, 2-BFI). Notably, CR4056 and LSL60101 diverge from this pattern, incorporating an imidazole motif (Figure 1.5).

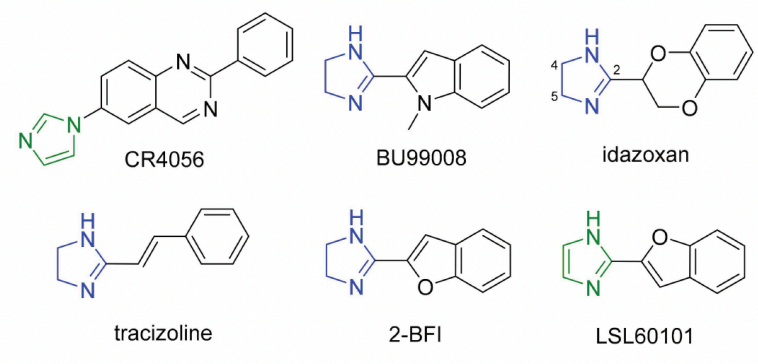


Figure1.5: Structure of known I2 ligands.[46]

The neuroprotective effects of I2 ligands are linked to their ability to reduce oxidative stress and neuroinflammation, which are key factors in neurodegeneration.[43] Ligands like 2-BFI and BU99008 may provide meaningful benefits in AD management by modulating neurotransmitters, reducing oxidative stress, and alleviating inflammation.[47]

1.3.2.1 Experimental studies on imidazoline I2 receptor ligands in neurodegenerative disease research

Experimental studies on I2 ligands have demonstrated their potential to reduce neurodegeneration, improve cognitive function, and modulate mood-related processes. The neuroprotective effect of I2 receptor ligands in brain cells in a model of Alzheimer's disease (AD) indicated that these ligands could promote cell survival in critical brain regions by activating protective pathways. Additionally,

they appeared to prevent processes linked to cell damage and cognitive decline.[43] Behavioral studies have shown that I2 ligands can improve cognitive performance in AD models, indicating that I2 ligands may mitigate cognitive decline by reducing biochemical markers of neurodegeneration.[48] In addition, species-specific differences in I2-IR behavior have also been explored, finding that certain I2 ligands interact with additional proteins in animal models, emphasizing the need for identification of the other receptors.[49]

Neuroprotective and neuroinflammatory effects of I2 receptor ligands

The I2-ligand 2-(2-benzofuranyl)-2-imidazoline (2-BFI) has been reported as an effective neuroprotector in several studies. It is reported that 2-BFI can reduce oxidative stress by upregulating antioxidant enzymes like superoxide dismutase (SOD1) and glutathione peroxidase (GPX-1) in mice, which protect neurons from oxidative damage linked to cognitive decline and neuronal death in diseases such as AD. By enhancing antioxidant activity, 2-BFI supports cellular resilience against oxidative stress, potentially slowing neurodegeneration.[47,50,51] 2-BFI also exhibits significant antinociceptive effects by directly targeting and modulating glial cell activity in the CNS. In a chronic constriction injury model of neuropathic pain, 2-BFI effectively reduced the activation of microglia and astrocytes in the spinal cord, which is known to play a key role in pain hypersensitivity (mechanical allodynia). Additionally, it lowered the elevated levels of TNF-α, a pro-inflammatory cytokine in the spinal cord, further supporting its role in reducing inflammation. These effects of 2-BFI were associated with decreased pain responses in CCI rats.[52,53] Complementary in vitro studies using mouse cortical astrocyte cultures confirmed that 2-BFI suppresses astrocyte activation triggered by inflammatory LPS-stimuli, significantly reducing TNF-α levels. These findings suggest that 2-BFI exerts its antinociceptive effects by modulating glial cell activity and reducing inflammation, highlighting its therapeutic potential for managing pain.[54]

Additionally, the imidazole analogue of 2-BFI, LSL60101 (garsevil) was tested in a 5xFAD familial AD mouse model. Garsevil reduced pro-apoptotic FADD protein levels and oxidative stress markers, suggesting its potential to decrease neurotoxicity through modulation of oxidative stress and apoptotic pathways. [47]

The neuroprotective effects of imidazoline ligands (I2) in ischemic brain injury models vary based on the compound and mechanism of action. Idazoxan reduced neuronal damage through hypothermia but also caused harm due to its interactions with multiple receptors. BU224 showed no neuroprotection in the rat, likely due to insufficient doses.[55] In contrast, 2-BFI selectively targeted I2 receptors, reducing NMDA receptor-mediated calcium influx, decreasing apoptosis, preserving cellular structures, and increasing Bcl-2 levels.[56,57] Other ligands, like tracizoline, produced similar behavioral outcomes to 2-BFI, such as anti-hyperalgesia and temperature reduction, suggesting additional mechanisms beyond biochemical changes.[58] These findings highlight the complexity of imidazoline ligands and their therapeutic potential.[40] Hypothermia has been linked to protective effects in brain injuries and neurological conditions like stroke and cerebral ischemia, where even a slight temperature drop can help protect brain cells and aid recovery. Studies have examined how specific I2-IR ligands such as idazoxan, BU224, 2-BFI, and LSL60101 have been shown to induce hypothermia in animal models.[43,47,49] Studies demonstrate that administering 2-BFI in mice leads to a controlled drop in body temperature, likely through interaction with I2 receptors affecting thermoregulatory pathways in the hypothalamus.[47,59] Hypothermia can lower metabolic demand and limit excitotoxicity, especially under ischemic or traumatic conditions. This hypothermic response is associated with neuroprotection, reducing neuroinflammatory damage, which may benefit neuronal health in neurodegenerative models.

1.3.2.2 Difficulty in identifying I2-IR receptors

Identifying the precise binding sites for I2-IRs is a challenging task due to their structural ambiguity and complex distribution across multiple proteins. Initially, I2-IRs were thought to be closely associated with monoamine oxidase B (MAO B). However, studies with MAO B-deficient mice have shown that certain I2-binding properties persist even in the absence of this enzyme, suggesting the existence of additional I2-binding proteins. This finding emphasizes the need to identify these alternate binding proteins to gain a full understanding of I2 receptor functionality and their pharmacological significance.[49,60] Further research has identified brain creatine kinase (B-CK) as a high-affinity I2-binding protein, supporting the hypothesis that I2-binding sites extend beyond MAO enzymes. B-CK's ability to bind I2 ligands implies roles in neuronal energy regulation, neurotransmitter release, and possibly neuroprotection.[60]

The presence of multiple I2-binding sites across different proteins highlights the need for highly selective ligands that can avoid off-target interactions, as many I2 ligands, including 2-BFI, also interact with α2-adrenoceptors. Identifying specific proteins associated with I2-binding would allow for the development of more targeted ligands, enhancing the precision of pharmacological applications and clarifying the roles of I2 receptors, particularly in neurodegenerative disorders where altered I2-IR levels may play a role.[61] Therefore, understanding the exact nature and the structure of I2-binding sites is essential for unlocking the full therapeutic potential of I2 ligands. As I2 receptors are implicated in complex brain functions, pinpointing their protein targets could lead to new therapeutic approaches. This potentially reduces side effects and broadens applications in treating neurodegenerative and psychiatric disorders.

1.3.3 Current methods for characterizing protein-ligand complexes.

Experimental techniques provide essential insights into protein-ligand binding by capturing structural, kinetic, and thermodynamic data. Methods like X-ray crystallography, NMR, Laue X-ray diffraction, small-angle X-ray scattering (SAXS), and cryo-EM offer high-resolution views of unbound proteins and protein-ligand complexes, revealing structural changes and binding dynamics. For example, X-ray crystallography provides enthalpic information through atomic interactions, with B-factors reflecting atomic displacements that indicate flexibility.[1] NMR measures dynamics across timescales from picoseconds to seconds, allowing indirect assessment of entropic contributions to binding.[62] Laue X-ray diffraction captures fast structural snapshots,[63] and cryo-EM offers lower-resolution views of protein conformations.[1] Additionally, single-molecule fluorescence spectroscopy[64] and time-resolved hydrogen-deuterium exchange mass spectrometry probe protein dynamics and binding mechanisms.[65]

Key methods for measuring binding affinity include isothermal titration calorimetry (ITC),[66,67] surface plasmon resonance (SPR),[68,69] and fluorescence polarization (FP).[70,71] ITC measures binding heat exchange directly, providing a complete thermodynamic profile, including binding free energy, enthalpy, entropy, and stoichiometry, though it requires large sample volumes and is not high throughput. SPR captures real-time binding kinetics via refractive index changes but requires immobilizing one binding partner; it can estimate enthalpy through van't Hoff analysis and has high-throughput potential. FP measures changes in fluorescence polarization between bound and unbound states, making it suitable for high-throughput screening due to its simplicity and low sample requirements. Together, these methods enable a comprehensive understanding of protein-ligand interactions, from structural changes to binding thermodynamics.

Theoretical and computational methods are crucial in protein-ligand binding studies, especially when experimental approaches are constrained by cost, time, or complexity. These techniques support high-throughput screening (HTS) of large

compound libraries in modern drug design and provide detailed insights into thermodynamic factors, such as enthalpic and entropic effects, which are challenging to analyze experimentally.

A key resource for these studies is the Protein Data Bank (PDB),[72,73] which offers three-dimensional protein structures derived from X-ray crystallography, NMR, and cryo-EM. Recently, AlphaFold, an AI-driven tool from DeepMind, has expanded available resources by accurately predicting protein structures, even in cases lacking experimental data. The AlphaFold Protein Structure Database has become a significant asset for computational studies.

Core computational tools like protein-ligand docking and binding free energy calculations are essential for studying molecular interactions. Docking predicts binding modes and affinities through search algorithms, which explore ligand poses, and scoring functions, which estimate binding affinity. Scoring functions fall into three categories: force-field-based (accurate but resource-intensive), empirical (faster but may overfit), and knowledge-based (fast and robust but data-limited). Often, consensus scoring is used to improve accuracy by combining results from multiple scoring functions. However, scoring methods rely on the additivity of pairwise interactions, treating them independently and missing the collective effects on binding free energy. This simplification overlooks critical entropic and enthalpic factors, including protein flexibility, solvent effects, and structural water roles.[74]

Unlike scoring functions, free energy calculations require extensive molecular dynamics (MD) or Monte Carlo (MC) simulations, making them computationally intensive but highly accurate. These methods account for both energetic and entropic contributions, offering quantitative insights into binding thermodynamics without needing case-specific parameters, though they demand significant computational resources for extensive conformational sampling.[13,75]

Additionally, mixed solvent MD methods like MDmix[76] provide insights into protein-ligand interactions with the protein from the analysis of the densities of selected solvents, which act as small fragments. This approach helps identify high-affinity binding spots where solvent molecules are retained. This pharmacophoric information allows to map solvent behavior within the binding pocket, providing a deeper understanding of key interaction sites in the protein.[76]

1.4 Overview in enzymatic catalysis

Without enzymes, the complex chemistry of life would be unmanageable, as reactions would lack the necessary speed and coordination to support health and survival.[77] Enzymes are vital biological catalysts that accelerate reaction rates by up to 10^6-10^{20} times.[77,78] They enable reactions to proceed under mild biological conditions of temperature and pH by lowering the activation energy.[79] This is achieved by several mechanisms like orienting substrates for bond formation or providing environments within active sites that stabilize transition states. They also exhibit high turnover numbers and frequencies, processing thousands of substrates per second.[1] Nearly all cellular processes rely on enzyme specificity and efficiency. Despite enormous experimental and theoretical advancements, fully understanding enzyme mechanisms remains challenging.[80,81]

In enzyme catalysis, reactions generally proceed by forming an enzyme-substrate complex (ES), which then allow transitions to the product (P), while the enzyme remains the same:

$$E + S \leftrightarrow ES \rightarrow P + E \qquad \text{Equation 1.7}$$

Two main theories, transition state (TS) stabilization theory and Michaelis Complex (MC) formation theory, explain how enzymes lower the activation energy and speed up the processes towards the product side of Equation 1.2. [78]

In TS stabilization theory, enzymes have a higher affinity for the TS than for the initial substrate. This results in a stronger binding free energy for the TS (ΔG^{TS}_{Bind}) compared to that of the MC (ΔG^{MC}_{Bind}), effectively lowering the activation free energy ($\Delta G^{\ddagger}_{cat}$) and facilitating a faster reaction. Linus Pauling proposed that an enzyme's active site complements the TS,[82] while Warshel suggested that the active site's electrostatic environment is preorganized to stabilize the TS's charge distribution,[83] avoiding the need for additional energy adjustments, unlike in aqueous solutions where the solvent must dynamically reorganize to stabilize the TS.

MC formation theory, on the other hand, emphasizes the role of the MC in catalysis. Upon binding, the substrate undergoes spatial rearrangement and stabilization, preparing it to form the TS. The binding free energy of the MC (ΔG^{MC}_{Bind}) reflects the energy changes associated with bringing substrate groups into proximity and proper orientation to react, which can include both enthalpic and entropic contributions. This process lowers the energy required for the substrate to progress to the TS, even if the enzyme shows a stronger affinity for the MC than for the TS.[78] Additionally, Bruice's near attack conformation (NAC) theory suggests that the enzyme's active site promotes TS-like substrate conformations, enhancing the likelihood of reaction.[84]

While TS stabilization theory (which highlights the enzyme's strong binding to stabilize the high-energy TS) and MC formation theory (which focuses on the enzyme positioning the substrate close to the transition state to lower activation energy) are the main models of enzyme catalysis, additional factors related to the dynamic nature of proteins have also been recognized as contributors to the catalytic process.[78]

1.4.1 Catalytic efficiency and contributing factors

Experimental techniques such as spectroscopy, mutagenesis, and kinetic analyses are widely used to investigate enzyme structure, conformation, and reaction kinetics. However, despite their contributions, these methods are not capable of providing a complete picture of the dynamic nature of enzymatic activity.[1,85] This limitation highlights the need for complementary computational approaches that can provide a more comprehensive understanding of enzyme catalysis with atomic detail. Computational methods highlight factors such as protein dynamics, solvent interactions, enthalpy-entropy balance, and electrostatic effects that influence enzyme catalysis.[86,87]

Enzymes contain networks of conserved residues that connect distant regions to active sites, facilitating energy transfer essential for catalysis and enabling allosteric regulation. Allosteric sites modulate enzyme activity through cellular signals, transmitting structural and energetic changes over long distances to influence the active site's properties.[88] The complex nature of enzyme active sites depends on catalytic residues, specific binding pockets, and dynamic networks working together to achieve catalysis. Rather than the simple 'lock and key' model, the 'induced fit' and 'conformational selection' models better represent dynamic enzyme behavior, where active sites use hydrogen bonds, electrostatic interactions, and structural flexibility to enhance catalysis.[81,87] The solvent environment around enzymes also plays a role in catalysis, forming a hydration shell that stabilizes specific conformational states. Solvent molecules interact with the enzyme's surface and active site. As evidenced by studies using terahertz spectroscopy, the enzyme activity is closely linked to solvent dynamics.[81,89] Enzyme catalysis also relies on enthalpy-entropy compensation[90] Electrostatic interactions within enzyme active sites are significant contributors to catalytic power. Electrostatic preorganization in the active site is thought to be a primary driver of catalytic efficiency. Studies show that specific arrangements of polar groups, like certain N-

H bonds, fine-tune electron distributions within reacting systems. Enzymes modify their electrostatic landscapes to optimize catalysis through organized interaction networks. Despite alternative models, such as reactant-state destabilization, mutational analyses suggest that electrostatic stabilization of TSs is central to enzyme catalysis. Studies using empirical valence bond (EVB) and hybrid QM/MM models support the role of electrostatic effects as key determinants of enzymatic catalytic power.[91]

1.4.2 Current computational techniques to handle catalytic reactions

Several computational methods are widely used in enzyme studies to capture different aspects of enzyme function and dynamics. Car–Parrinello molecular dynamics (CPMD) combines classical molecular dynamics with quantum mechanics using density functional theory (DFT) to simulate both nuclear and electronic movements. This technique, while computationally intensive, is particularly valuable for studying enzyme reactions involving complex electronic transitions, such as bond formation and breaking. In contrast, EVB methods offer a less computationally demanding approach by modeling reaction pathways through classical valence-bond theory, which represents a reaction as a combination of classical (rather than quantum) electronic states or intermediates.[92] Although EVB requires predefined intermediates, it provides insight into reaction mechanisms when combined with experimental data. Cluster modeling typically uses DFT to focus on specific enzyme regions, such as the catalytic site, allowing detailed analysis of the electronic structure and reactivity of these residues. However, cluster models may struggle to capture long-range interactions and dynamic effects, which are important for enzymes' full behavior.[93,94]

To explore reaction pathways, conformational changes, and rare events, advanced sampling techniques are also used. Umbrella sampling applies a bias to reaction coordinates, allowing exploration of high-energy states and calculation of free-energy profiles. Nudged elastic bands (NEB) identify minimum energy pathways

between states, such as reactants and products, clarifying how enzymes overcome energy barriers. When it comes to rare transitions in enzyme catalysis, metadynamics adds a time-dependent bias to selected degrees of freedom, helping the system explore complex energy landscapes by overcoming local minima. Additionally, free-energy perturbation (FEP) calculates free energy differences between states, useful for predicting the effects of mutations or ligand modifications on enzyme activity.

Among the computational techniques used in enzyme studies, hybrid QM/MM methods are particularly influential. QM/MM techniques, pioneered by Warshel and Levitt in the 1970s, divide the enzyme system into a QM region and an MM region. The QM region, often including the active site or key catalytic residues, is treated with quantum-level accuracy, while the MM region models the enzyme's surrounding environment. This division allows for quantum accuracy in critical areas without the prohibitive costs of full QM calculations. QM/MM techniques thus enable detailed studies of enzyme mechanisms, providing insights into catalytic residues, transition states, and reaction pathways that are challenging to capture through experiments alone. Although computationally expensive BOMD simulations take into account enzyme flexibility and reaction mechanisms by recalculating QM energies at each step along the MD simulation, assuming that the electronic state remains in its ground state as the nuclei move.[91]

CHAPTER 2. OBJECTIVES

In the first part of this thesis, we aim to investigate the exciton properties of *Hemiselmis virescens* PC612 and *Chroomonas sp*. PC645 antenna complexes using a dual-MD multiscale protocol combining classical MD and BOMD simulations based on DFT. QM/MM MD simulations accurately describe the spectral density of electron-vibrational coupling in photosynthetic complexes, allowing an accurate treatment of thermal effects. We compare this approach with an alternative protocol based on multiple QM/MM geometry optimizations along the trajectories. Additionally, we adopt the full Cumulant Expansion (FCE) formalism to describe the absorption and circular dichroism spectra, which includes non-Markovian and non-secular effects often neglected in modified Redfield theory. In this context, the specific objective of this project is:

1- **Assess the potential of the dual-MD multiscale protocol**
   We aim to assess the ability of the dual-MD strategy to characterize the excited states and light harvesting properties of multichromophoric pigment-protein complexes, and to provide accurate exciton Hamiltonians for PC612 and PC645 to be used in quantitative studies on their energy transfer mechanisms.

The second project of this thesis is conducted in collaboration with the experimental research groups of Prof. Carmen Escolano (Medicinal Chemistry), Prof. Mercè Pallàs, and Dr. Christian Griñán Ferré (Pharmacology) at the University of Barcelona, who develop I2 ligands with potential therapeutic applications. A critical step in this research is identifying target receptors for these ligands, for predicting their biological effects and therapeutic potential. By pinpointing these targets, we can gain insights into how I2 ligands may influence key cellular pathways and contribute to treating conditions such as neurodegenerative diseases and central nervous system disorders. And more importantly, by analyzing the pharmacophoric characteristics of the binding sites, we can facilitate the development of more targeted and effective drugs.

In this context, the specific objectives of this project are as follows:

1. **Identification of putative binding sites for I2 Ligands**
   Utilizing proteomics data for the potential I2-IR targets provided by our experimental collaborators, we aim to apply multistep molecular simulations to identify binding sites for the canonical I2 ligand 2-BFI.

2. **Investigation of 2-BFI's modulatory role on MAO-B**
   Given the known binding site of 2-BFI to MAO-B, our second objective is to deepen our understanding of how 2-BFI may influence MAO-B function. We aim to examine the potential impact of 2-BFI on the entry and exit pathways of MAO-B's natural substrate and product and to analyze the influence of 2-BFI on the energy profile of the rate limiting step of the catalytic degradation of MAO-B's substrates.

CHAPTER 3. METHODOLOGY

3.1 Classical methods for structural dynamics of biosystems

Biochemical systems exhibit motions that span a wide range of timescales, from femtoseconds to milliseconds. However, capturing these motions with QM simulations is challenging. While QM approaches are essential for accurately describing atomic-level processes, they are limited by their high computational cost, making them feasible only for very short timescales, typically up to a few hundred picoseconds. To address this limitations, classical empirical force fields (FFs) are employed, enabling the study of larger systems over extended timescales, reaching from nanoseconds to milliseconds.[95,96] Although these force fields ignore some electronic effects, such as charge transfer, polarization, subtle van der Waals dispersion contributions, and orbital interactions, and do not account for bond breaking and formation, they provide essential insights into the dynamics and properties of biochemical systems. This simplification allows FFs to bypass computationally expensive QM calculations, which makes simulations of larger systems over longer timescales feasible.[91,97]

3.1.1 Pairwise additive force fields

Empirical FFs are mathematical models used to estimate the potential energy of a system based on the positions of atoms using parameters derived from experimental data and QM calculations.[98] These FFs enable the calculation of forces acting on atoms in order to predict their motions over time. They simplify molecules into a collection of spherical particles linked by harmonic springs, which are defined by specific properties such as mass (m_i), charge (q_i) or van der Waals radius (R_i). The interactions between these particles are well described by classical mechanics, reducing the complex nature of molecular systems into more manageable mathematical forms.[99] In classical FFs, the total potential energy (U)

is expressed as a sum of different components, reflecting bonded intramolecular and non-bonded intermolecular interactions:

$$U(r) = E_{\text{bonded}} + E_{\text{non-bonded}} \qquad \text{Equation 3.1}$$

In the Amber familiy of FFs adopted in this thesis, the bonded term is expressed as the sum of several types of interactions;

$$E_{\text{bonded}} = \sum_{bonds} \mathrm{K_r}(\mathrm{r} - \mathrm{r_{eq}})^2 + \sum_{angles} \mathrm{K_\theta}(\theta - \theta_{\mathrm{eq}})^2 + \sum_{dihedrals} \mathrm{V_n}/2[1 + \cos(\mathrm{n}\emptyset - \gamma)] \qquad \text{Equation 3.2}$$

Harmonic potentials govern bond stretchings between two atoms (1-2 bonds) by Hooke's law. A similar expression is used for angle bendings between three atoms (1-3 angles), with r representing bond distances and θ the bond angles. Torsional movements involving four atoms are modeled with periodic functions; dihedral (1-4 dihedral) and improper torsions (1-4 improper). For dihedrals, $\mathrm{V_n}$ is the barrier height to rotation about an angle $\emptyset$, n indicates the number of minima in a full rotation, and γ is the phase angle corresponding to the lowest potential.

Similarly, the non-bonded potential energy is calculated considering atom pairs separated by more than three bonds (1–4 interactions), as shorter interactions are already handled by bonded terms:

$$E_{\text{non-bonded}} = \sum_{i<j} \frac{q_{\mathrm{i}}\, q_{\mathrm{j}}}{\varepsilon R_{\mathrm{ij}}} + \sum_{i<j} \left[\frac{A_{\mathrm{ij}}}{R_{\mathrm{ij}}^{12}} - \frac{B_{\mathrm{ij}}}{R_{\mathrm{ij}}^{6}}\right] \qquad \text{Equation 3.3}$$

Here, the first term represents the electrostatic interaction between atoms i and j, calculated using Coulomb's law, where q_{i} and q_{j} are the charges of the atoms and

R_{ij} is the interparticle distance. In FFs, the dielectric constant ε is set to 1, representing vacuum conditions, so it is typically omitted from the equations. The second term is the Lennard-Jones potential, capturing van der Waals forces, where A_{ij} and B_{ij} are the coefficients for the attractive and repulsive interactions, respectively. While the Lennard-Jones potential is widely used, some force fields employ an alternative exponential repulsion term, which provides a more accurate representation of short-range interactions. However, this approach is computationally more demanding.

3.1.2 Polarizable force fields

Polarization is the redistribution of a molecule's electron density in response to an electric field from surrounding molecules.[100] When multiple molecules are present, this leads to **nonadditivity**, where the interaction between any two molecules is altered by the influence of others. Nonadditivity is essential for accurately modeling complex systems, as each molecule's response depends on its environment.[101]

Polarizable force fields improve on traditional non-polarizable force fields by accounting for electron redistribution in response to varying environments. This enables more accurate simulations of molecular properties, such as solvation energies, binding affinities, and thermodynamic properties. By allowing atoms to respond dynamically to changes in their dielectric environment, polarizable force fields capture interactions more realistically, which is important for studying complex biological systems and material properties.[100] There are different approaches to implementing polarization in force fields. Some of the key models include:

Fluctuating Charge Model is based on electronegativity equalization, this model allows charges to redistribute among atoms until their electronegativities are

balanced, providing a means to account for polarization effects. Examples include the Universal Force Field (UFF)[102] and CHARMM.[103]

Drude Oscillator Model, also known as the shell model, represents each atom as a two-particle system (a charged core and a charged shell) connected by a harmonic spring. Polarization occurs through the relative displacement of the core and shell in response to an external field. Force fields such as CHARMM[104] and GROMOS[105] have implemented Drude oscillator methods.

Induced Dipole Model calculates polarization by inducing dipole moments in atoms in response to the local electrostatic field.[106,107] Force fields incorporating this method include AMBER,[108–110] OPLS/PFF,[111] and AMOEBA.[112,113]

In this thesis, we use the **induced dipole model** for excited state QM/MM calculations in photosynthetic complexes, in which an atom's induced dipole moment, μ_i is directly related to its polarizability and the surrounding electrostatic field. This polarization energy, represented as V_{pol} is added to the system's total potential energy, which also includes bonded and nonbonded contributions.

In AMBER's polarizable force fields, the total potential energy of a system is still given as the sum of contributions from bonded and nonbonded terms (equation 3.2 and 3.3), but the polarization energy is added to the nonbonded contribution. The polarization energy, noted as V_{pol}, is specifically determined through an induced dipole model that accounts for how the electron clouds of atoms and molecules respond to their electrostatic environment. Here, μ_i is the induced dipole moment for the ith atom, and E_i^0 is the local electric field acting on atom i due to all other charges.

$$V_{pol} = -\frac{1}{2}\sum_i \mu_i E_i^0 \qquad \text{Equation 3.4}$$

Here, μ_i is the induced dipole moment for the ith atom, and E_i^0 is the local electric field acting on atom i due to all other charges.

$$\mu_i = \alpha_i E_i \qquad \text{Equation 3.5}$$

The induced dipole moment, μ_i, is the product of the atomic polarizability, α_i, and the electric field, E_i, at atom i (Eq 3.5).

$$E_i = E_i^0 - \sum_{j \neq i} T_{ij}\, \mu_j \qquad \text{Equation 3.6}$$

The total electric field, E_i, acting on atom i is determined by combining the external field E_i^0, which originates from the static charges, with the contributions from the induced dipoles of surrounding atoms. This cumulative effect is adjusted by the dipole-dipole interaction tensor T_{ij}, which modulates the influence of each neighboring atom's dipole moment μ_j. on i. In this formulation, the dipole contributions adds to the total field, as E_i^0 represents only the charge-based field, and the term $-\sum_{j \neq i} T_{ij}\, \mu_j$ accounts for the additional field from surrounding induced dipoles (Eq 3.6).

$$E_i^0 = \sum_{j \neq i} \frac{q_j \mathfrak{r}_{ij}}{{r_{ij}}^3} \qquad \text{Equation 3.7}$$

The external electrostatic field, E_i^0, is the sum of contributions from all other point charges, q_j, with $\mathfrak{r}_{ij}$ being the distance vector between atoms i and j. (Eq 3.7)

$$T_{ij} = \frac{f_e}{{r_{ij}}^3} I - \frac{3 f_t}{{r_{ij}}^5} \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix} \qquad \text{Equation 3.8}$$

The iterative calculation of induced dipoles and the resulting dipole-dipole interactions in a system is managed through a self-consistent field approach. To prevent excessive polarization (overpolarization) at short distances, screening functions *fe* and *ft* are applied. These Thole damping functions modulate interactions by adjusting the dipole-dipole tensor, which reduces the intensity of dipole interactions at close ranges. Additionally, to further minimize short-distance polarization issues, certain force fields exclude direct 1-2 and 1-3 interactions, while 1-4 interactions are scaled according to the specific polarizable force field used.[114,115]

The dipole-dipole interaction tensor T_{ij}, adjusts the electric field at one atom due to neighboring dipoles, capturing the directional (anisotropic) nature of dipole interactions (Eq 3.8). This tensor incorporates the x, y, and z components of the distance vector r_{ij}, which links the i-th and j-th atoms or dipoles, with I being the unit matrix for isotropic interactions. When applied, T_{ij} precisely modulates the effect of an induced dipole at one location on the field at another, accounting for spatial arrangement. This calculation is crucial for determining polarization energy, as it captures how each dipole's induction affects others, within or between molecules.

In this thesis, the Amber pol12 AL (Thole Linear) parameters are used to describe MMPol region in Chapter 4. In the Amber pol12 AL force field, short-range polarization is managed by excluding 1-2 and 1-3 interactions from the polarization energy calculation, as these interactions are already addressed by bonded terms. 1-4 interactions (atoms separated by three bonds) are included but scaled down by a factor of 1.2, reducing their strength to prevent overpolarization. Additionally, Thole-linear screening is applied to dampen short-range dipole-dipole interactions, stabilizing the induced dipoles and ensuring a realistic representation of polarization across the molecule.[114,115]

3.1.3 System minimization

Minimization in empirical force fields is a critical refinement step to determine coordinates, r_{min}, that correspond to the system's minimum potential energy, U_{min}. This process involves iteratively adjusting atomic positions to reduce potential energy, moving the system toward a stable, low-energy state. Two main algorithms are commonly used for energy minimization: **steepest descent**[116] and **conjugate gradient.**[117] In both algorithms, an initial configuration r_0, with potential U_0 is set up, and a new configuration r_i with potential U_i is then extrapolated. The gradient of the potential energy is utilized as a search vector ∇U_i for iterative optimization, where each step involves adding a weighted contribution from the previous step to refine the current configuration until the change in potential energy is minimized.

$$r_{i+1} = r_i + g_i \nabla U_i \qquad \text{Equation 3.9}$$

Here, g_i determines how far along the gradient's direction the current position should move to get the new position and signifies the weight of the gradient's influence on the next configuration. The way of constructing new search vector differentiates steepest descent and conjugate gradients methods. Steepest descent requires only that each new gradient be orthogonal to the previous one, iteratively moves towards the minimum energy by following the steepest downhill path, which can be beneficial when starting far from the minimum. Meanwhile the conjugate gradient method constructs each new search vector to be orthogonal to all previous gradients, enabling a more efficient and non-redundant exploration of the search space. In conjugate gradient method, this is achieved by a repetitive operation where each new gradient ∇U_{i+1} is a linear combination of the previous gradients, scaled by a weight g. This weight adjusts the influence of the previous gradient on the new direction. Conjugate gradients tend to converge more slowly

when distant from the energy minimum but offer greater stability close to it by avoiding oscillatory behaviors. Both methods, however, assume a constant second derivative, which can lead to numerical instabilities with increasing iterations, underlining the complexity of energy minimization in these simulations.[118]

3.1.4 Molecular Dynamics simulations

MD simulations using empirical force fields provide a dynamic view of molecular systems, enabling the exploration of their structural fluctuations and conformational changes over time. Before starting MD, identifying the minimum energy configuration is essential to avoid bad contacts or unrealistic geometries, which could destabilize the system during simulation. This minimization step helps ensure the system's stability at the start of the MD simulation, but it represents only one of the many conformations the molecular system might explore. This is because any system can exist in a variety of configurations, influenced by thermodynamic variables such as temperature (T), pressure (P), and volume (V). To understand macroscopic properties, it's essential to analyze ensembles of configurations reflective of different conditions. MD simulations allow for this by computing the time-evolved trajectories of atoms and molecules, exploring the system's phase space through Newton's laws of motion.[75,85]

$$-\frac{\nabla U(r)}{\mathrm{m}} = \frac{\mathrm{d}^2}{\mathrm{dt}^2} r(t) \qquad \text{Equation 3.10}$$

This process is governed by this second order differential equation, where the acceleration of an atom of mass *m* in a configuration *r* is determined by the forces acting on it, which is the negative gradient of the potential energy, $-\nabla U$. This second-order differential equation, due to its polyatomic complexity, must be solved numerically within MD simulations. This allows iteratively approximate

molecular motions over incremental time steps, providing a dynamic view of molecular behavior at a microscopic level.

In MD simulations, trajectories are typically computed using integration algorithms like Verlet, leapfrog, and velocity-Verlet. These methods update atomic positions and velocities to simulate motion. The Verlet algorithm, for example, estimates positions based on current and previous positions and the current acceleration:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \qquad \text{Equation 3.11}$$

The Verlet algorithm calculates the new position $r(t + \Delta t)$ using current $r(t)$ and previous positions $r(t - \Delta t)$, as well as the current acceleration a. This approach avoids directly computing velocities, which can be useful for energy conservation. However, the lack of velocity information can limit precise temperature control. To address this, the velocity-Verlet method expands upon the Verlet algorithm by also updating velocities, making it easier to maintain temperature while preserving stability and energy conservation in the simulation.[98]

The integration time step Δt is crucial in MD simulations, which balances efficiency and accuracy. A smaller Δt (0.5–1 fs for flexible molecules) captures fast motions, such as the 10 femtosecond C-H bond vibrations, but increases computational demand. For rigid molecules, larger steps (around 5 fs) improve efficiency but risk energy fluctuations or instability. Checking energy conservation in a microcanonical ensemble helps ensure an appropriate Δt selection.[119,120] Methods like the SHAKE algorithm permit larger Δt by constraining rapid bond movements involving hydrogens. In this way it balances the computational efficiency and accurate dynamics and enables a broad exploration of the biomolecular phase space.[121] On the other hand, other protocols like the hydrogen

mass repartitioning (HMR) method adjusts the masses of atoms within the simulation.[122] It increases the mass of hydrogen atoms and decreases the mass of the atoms to which they are bonded. The redistribution of mass slows down the vibrations of hydrogen atoms, allowing for larger time steps without the need for constraints like those imposed by the SHAKE algorithm.[123]

The choice of phase space ensemble, on the other hand, has a crucial impact on conducting meaningful biomolecular dynamics studies under specific thermodynamic conditions. The selection ranges from the Microcanonical (NVE) ensemble, ideal for isolated systems where the number of particles (N), volume (V), and total energy (E) are constant, to the Canonical (NVT) ensemble, which allows for energy exchange with an external heat bath, maintaining stable conditions suitable for biological studies through constant N, V, and T, achieved with thermostats like Langevin[124] or Nose-Hoover. The Isothermal-Isobaric (NPT) ensemble further extends the simulation's realism by allowing volume and energy fluctuations under constant N, P, and T, using barostats such as Berendsen[125] or Parrinello-Rahman, thereby closely mirroring laboratory conditions ideal for constant pressure and temperature simulations. In this thesis, NVT and NPT ensembles are used for exploring the phase space of biomolecular systems, where explicit solvation effects are significant.

Molecular simulations require a balance between computational efficiency and accurate long-range interaction modeling, especially for large biomolecular systems like proteins and lipids in extensive water environments. Techniques such as periodic boundary conditions (PBC) and cutoff distances help manage these challenges. PBC replicates the simulation cell to simulate bulk systems, allowing particles exiting on one side to re-enter on the opposite side, while a cutoff distance limits interactions to nearby neighbors. To reduce electrostatic errors, which decay more slowly than Lennard-Jones interactions, the Ewald summation technique adds Gaussian distributions to modulate electrostatic potentials beyond the cutoff.

Although implicit solvation models simplify simulations by using a dielectric constant, explicit solvation models are often preferred for capturing the detailed dynamics essential to biomolecular accuracy.[126]

Several leading simulation software such as CHARMM,[127] AMBER,[128] GROMACS,[129] and OpenMM[130] support MD studies. Common force fields like AMBER,[99,131] CHARMM,[132] GROMOS,[133] and OPLS[134] differ in their parametrization strategies but are generally effective in accurately reproducing experimental data. In this thesis the AMBER force fields and software have been used.

3.1.5 Methods to study ligand binding

Computational methods provide valuable thermodynamic insights and enable high-throughput compound screening for ligand binding, particularly when experimental options are limited by cost or complexity. Protein-ligand docking is widely used to predict binding modes and affinities by sampling ligand poses within a binding site and scoring their potential interactions.[135,136] However, traditional scoring functions often simplify binding interactions, overlooking critical factors like protein flexibility and solvent effects. To address these limitations, binding free energy calculations offer a more precise analysis, incorporating both energetic and entropic contributions through MD or Monte Carlo simulations. Although computationally intensive, these methods provide quantitative insights into binding that are more accurate than those from docking alone.

In this thesis, we employ a dual approach using MDmix[76,137] and docking (rdock)[138]. MDmix integrates MD simulations with mixed solvents to identify high-affinity binding spots. By analyzing solvent behavior within the binding

pocket, MDmix maps dynamic interaction sites, adding depth to traditional docking calculations. This pharmacophoric mapping highlights solvent-protein interactions that play critical roles in ligand binding and specificity.

3.1.5.1 Identification of binding sites with mixed-solvent MD simulations

Over the years, various computational methods have been developed to identify druggable sites on proteins, each with unique strengths. MixMD, developed by Carlson's group,[139] utilizes protein flexibility and solvent competition to map binding hotspots, capturing both active and allosteric sites effectively. SILCS, from MacKerell's group,[140] uses all-atom explicit-solvent MD to create 3D probability maps for fragments, enabling accurate binding free energy calculations, though it is computationally intensive. JEDI, introduced by the Michel's group,[141] combines empirical descriptors during MD simulations to evaluate druggability on-the-fly without using organic probes or protein restraints, offering a fast and continuous method that can also bias sampling to detect cryptic sites. SWISH-X, developed by research group of Gervasio[142] focusing on enhanced sampling, combines SWISH with OPES MultiThermal to identifying cryptic pockets at protein-protein interfaces and individual proteins by lowering kinetic barriers and effectively sampling complex conformational changes. The MDmix technique, introduced by Barril's lab in 2009, applies molecular dynamics simulations using aqueous/organic solvent mixtures to identify and evaluate protein binding sites for their druggability.[76,137] This method enables the identification of energetically favorable hotspots for potential drug-like molecules to bind, without the need for the actual ligands themselves. It also allows identifying hidden pockets and enhances the accuracy of docking simulations. MDmix, builds on previous work by Goodford [143] and others, utilizes MD simulations with mixed solvents to provide highly accurate interaction maps and water displaceability predictions. It outperforms methods like GRID from Goodford's group[144] by explicitly

accounting for solvation and protein flexibility, making it a powerful tool in structure-based drug discovery, particularly in industrial settings where quick turnaround times are essential.

In the MDmix approach, proteins are placed in a solvent mixture, often using different solvent mixtures such as 20% (v/v) ethanol/water, 20% (v/v) isopropanol/water, and 1% (v/v) pyridine/water solutions. The concentrations of the solvent mixtures are determined carefully to prevent phase separation or self-aggregation.

$$\Delta G_{bind} = -k_b T.ln\left(\frac{N_i}{N_0}\right) \qquad \text{Equation 3.12}$$

After the MDmix production phase, the system is analyzed by mapping solvent interactions onto spatial grids around the protein. This approach enables measurement of local solvent densities, identifying high-affinity interaction sites. These densities are compared against expected standards and converted into binding free energy (ΔG_{bind}) through the inverse Boltzmann equation, linking density variations to energy (Equation 3.12). Areas with significantly negative ΔG_{bind}, highlighting the interaction with the probe, yield the most energetically favorable results and are identified as "hotspots". Those areas suggest potential sites for strong ligand binding.[76]

3.1.5.2 Molecular docking

Molecular docking calculations aim at predicting ligand and target molecule interactions in a high-throughput manner. The first component of any docking program is a search algorithm (SA) that can generate different ligand poses within a defined binding site of the target protein. Different docking programs use different approaches to implement the search algorithm: For example, Metropolis-Monte Carlo based methods randomly adjust the ligand's position and shape,

accepting changes based on energy levels to avoid local energy minima, suitable for large conformational spaces. Alternatively, genetic algorithms apply natural selection concepts, starting with varied ligand poses and selecting the most efficient binders. Through crossover and mutation, they evolve new generations, producing optimal docking poses. A third kind of searching algorithm is based on systematic or grid searches, which evaluate energy interactions across a grid, ensuring thoroughness at the cost of computational resources, used when precision is paramount. Though ideally, search algorithms would cover all binding modes, the extensive computational resources needed make this impractical, particularly for larger molecules. Thus, sampling is limited to feasible conformational spaces, with constraints and approximations focusing on probable minimum energy regions.

The second component of a molecular docking programs is the scoring function (SF) that evaluates and ranks generated ligand poses by their likelihood to be true binding modes, differentiating various interactions to find the most promising ones. SFs are categorized into three main classes: force-field based, knowledge-based, and empirical. Force-field SFs, like AMBER and CHARMM, compute binding energy from energy terms including van der Waals, electrostatics, and hydrogen bonds, demanding extensive computational resources. The DOCK program, utilizing Amber force fields, exemplifies force-field SF in assessing molecular interactions. Knowledge-based SFs, as in DrugScore and PMF (Potential of Mean Force) program, use statistical analyses from large databases like the PDB to emulate experimental molecular structures. However, their precision relies on database size, and they may overlook less common interactions.[74] Empirical docking methods, favored for speed and suitability in high-throughput screening, fit their scoring functions from experimental data, which may more accurately reflect biochemical behaviors in certain systems. These functions use a weighted sum of parameters to estimate binding free energy. For instance, AutoDock utilizes empirical scoring for automated docking, while

Glide, part of the Schrödinger suite, is noted for high-throughput virtual screening due to a highly refined scoring function, GlideScore, that allows for faster, more accurate screening of large compound libraries.[145] Additionally, AI-driven advancements, such as deep learning and generative models, have significantly improved docking accuracy. Tools like Gnina and DiffDock use Convolutional Neural Network (CNNs)[146], a type of deep learning algorithm, and diffusion models to predict diverse ligand poses, enhancing binding predictions. Platforms like DockStream in REINVENT[147] enable complex protein-ligand studies, while Deep Docking uses QSAR models to accelerate virtual screening across large databases.[148]

rDock

In this thesis, Rdock[138] was used to perform the docking calculations. The standard docking protocol in rDock consists of four stages: First it needs to delineate the region of the protein in which the docking calculation will take place. There are two alternative methods for cavity definition within rDock. On the one hand, in the **two sphere method** the docking cavity is defined using two spheres of different sizes. The cavity is only accessible by smaller spheres, representing the ligand, but not by larger spheres preventing the consideration of highly solvent accesible surfaces. On the other hand, in the **reference ligand method** a docking volume of a specified size is created around a known ligand, using spheres of small radius to avoid sterical clashes with the protein residues. Subsequently, rDock leverages a genetic algorithm to explore the conformational space of the ligand within the defined cavity. Then, a low temperature Metropolis-Monte Carlo algorithm is applied to further refine the ligand poses. Finally, a minimization stage using conjugate gradient allows to optimize the interaction distances and angles, therefore minimizing the energy of the system.

The scoring function of rDock is an empirical SF constructed using weighting factors (*W*) and score components (*S*). According to equation 3.13, the output of the scoring function (S^{total}) is a weighted sum of intermolecular (S^{inter}), ligand intramolecular (S^{intra}), site intramolecular (S^{site}), and external restraints terms ($S^{restraint}$). $S^{inter}$is the main term of interest as it represents the protein-ligand interaction score. S^{intra} reports the change in energy of the ligand relative to the input ligand conformation. Similarly, S^{site} represents the relative energy of the flexible regions of the active site. In the current implementation, the only flexible bonds in the active site are terminal OH and NH3+ bonds. $S^{restraint}$ are a collection of non-physical restraint functions that can be used to bias the docking calculation in several useful ways, for example evaluating if the molecule complies with a set of previously determined pharmacophoric traits. S^{inter}, S^{intra}, and S^{site} are built from a common set of constituent potentials. Briefly, they mainly consist of a van der Waals potential (vdW), an empirical term for attractive and repulsive polar interactions, and an optional desolvation potential that combines a weighted solvent accessible surface area approach with a rapid probabilistic approximation to the calculation of solvent accessible surface areas for computational efficiency.

$$S^{total} = S^{inter} + S^{intra} + S^{site} + S^{restraint} \qquad \text{Equation 3.13}$$

Docking programs face limitations that impact their accuracy and reliability, particularly due to oversimplifications in scoring functions. Scoring methods often rely on the concept of additivity, where pairwise interactions are treated independently to estimate binding free energy, neglecting the complexity of protein-ligand binding. This approach fails to capture key factors like dynamic nature of intermolecular interactions, protein flexibility, solvent effects, and the role of structural waters, all of which contribute to the system's entropy and enthalpy. Additionally, docking models frequently assume a rigid target structure, overlooking essential protein flexibility and induced fit effects. These

simplifications limit the ability to accurately predict binding affinities, highlighting the need for complementary methods to achieve reliable results in ligand binding.[145]

3.2 Multiscale methods for enzymatic catalysis and light harvesting

QM offers a rigorous approach to modeling biochemical systems at the atomic and subatomic levels, where the classical laws of physics do not apply. QM methods allow understanding the electron behavior and chemical bonding within molecules, enabling the study of processes such as electronic transitions, chemical reactivity, and bond formation or breakage. Rigorous ab initio QM methods based on the solution of the Schrödinger equation using wavefunctions are the golden standard to determine molecular properties, but due to their high computational cost, methods based on DFT have emerged as an attractive choice for larger systems up to a few hunderd of atoms, offering a good compromise between accuracy and cost.

MM methods, on the other hand, approximate molecular interactions using force fields based on classical physics. While MM methods are computationally more efficient and can handle larger systems over longer timescales, they have important limitations like the unability to describe bond breakage or electronic properties because electrons are only included implicitly. This has motivated the development of focused hybrid approaches like QM/MM, which combine the precision of QM for a region of interest where the process of interest occurs, with the efficiency of classical MM for the remaining environment. This enables the study of much larger systems and longer timescales while maintaining a detailed description of the process under scrutiny.

3.2.1 General Approach to density functional theory

DFT is a QM method for modeling the electronic structure of atoms and molecules. It provides an efficient way to approximate the Schrödinger equation for electron-rich systems by focusing on total electronic energy and electron density distribution, rather than the more complex wavefunctions of the electrons. Introduced by Hohenberg and Kohn in 1964, the foundational theorem of DFT states that a system's ground-state properties can be obtained from its electron density, reducing the variables from 3N for N electrons to just three spatial dimensions. This simplification, incorporating exchange and correlation effects via functionals of the electron density, makes DFT more computationally manageable than traditional ab initio methods like Hartree-Fock, which deal with electron wavefunctions and electron correlation with higher computational demands.
In DFT, the electron density is defined as:

$$\rho(\mathrm{r})=\mathrm{N}\int\cdots\int|\Psi(\mathrm{r}_1,\mathrm{r}_2,\cdots,\mathrm{r}_n)|^2\mathrm{dr}_1\mathrm{dr}_2\cdots\mathrm{dr}_n \qquad \text{Equation 3.14}$$

where *r* represents both spin and spatial coordinates for a system with N electrons indicated by the *n* index. Integrating the square of the many body wavefunction over all possible positions of the electrons results in the electron density ($\rho(\mathrm{r})$), which represents the probability density of finding electrons at any particular point within the system.[118]

The energy of the interacting electron system is modeled as a functional of $\rho(\mathrm{r})$, decomposed into:

$$E[\rho(\mathrm{r})]=\int \mathrm{V}_{\mathrm{ext}}(\mathrm{r})\,\rho(\mathrm{r})\mathrm{dr}+\mathrm{F}[\rho(\mathrm{r})] \qquad \text{Equation 3.15}$$

The term $V_{ext}(r)\,\rho(r)dr$ represents how electrons interact with an external potential, $V_{ext}(r)$, typically the Coulomb potential from the atomic nuclei.

Meanwhile, $F[\rho(r)]$ expresses the kinetic energy of electrons and electron-electron interactions. Kohn and Sham's pivotal contribution in 1965 refined DFT by specifying $F[\rho(r)]$ as a sum of the kinetic energy of non-interacting electrons ($E_{KE}[\rho(r)]$), the electrostatic energy between electrons ($J[\rho(r)]$), and the exchange-correlation energy ($E_{XC}[\rho(r)]$), which accounts for both exchange interactions and correlations among electrons, including the kinetic energy difference between non-interacting and interacting electrons:[149]

$$F[\rho(r)]=E_{KE}[\rho(r)]+J[\rho(r)]+E_{XC}[\rho(r)] \qquad \text{Equation 3.16}$$

The Kohn-Sham approach, resembling the one-electron Schrödinger equation, treats electrons in a system as if they were non-interacting and moving in an effective potential by introducing a set Kohn-Sham orbitals that generate the same electron density as the real, interacting system.

The ground-state solution, derived from Kohn-Sham equations, uses the Kohn-Sham wavefunction—a single Slater determinant formed from independent orbitals Ψi

$$\left[-\frac{1}{2}\nabla^2+V_{KS}\right]\Psi_i=\mathcal{E}_i\Psi_i \qquad \text{Equation 3.17}$$

Here, the first term represents the kinetic energy operator of an electron, V_{KS} represents the Kohn-Sham potential, which combines the external potential (V_{ext}) by the nuclei, the Hartree electrostatic potential (V_H) of electrons, and the exchange-correlation potential (V_{XC}), and the eigenvalue, $\mathcal{E}_i$, represents the Kohn-Sham orbital energy.

The electron density in Kohn Sham DFT is then defined as:

$$\rho(r)=\sum_i^N|\Psi_i|^2 \qquad \text{Equation 3.18}$$

This equation sums the densities from all occupied Kohn-Sham orbitals, Ψ_i, to get the total electron density, $\rho(r)$ at position r, and N is the number of electrons.

In the self-consistent field (SCF) method, there is a continuous feedback loop between Kohn-Sham orbitals (Ψ_i) and electron density ($\rho(r)$). The process starts with an initial guess of $\rho(r)$, which is used to construct the Kohn-Sham potential (V_{KS}). Following this, the Kohn-Sham equation is solved for each electron orbital, refining the electron density with each iteration. The total energy $F[\rho(r)]$ of the system can be calculated using the updated density. This iterative cycle continues until the changes in electron density between successive iterations provide energy difference below the treshold, indicating that the orbitals and electron density are synchronized, accurately representing the system's ground state.

Separation of Exchange and Correlation Effects

The Kohn-Sham potential is defined to yield the same electron density as the interacting system by non-interacting electrons. The exchange-correlation component of the Kohn-Sham potential is critical for mimicking electron-electron interactions, reflecting the repulsion between the electrons with the same spin due to the Pauli exclusion principle (exchange) and the dynamic adjustment of electrons to minimize the energy by avoiding proximity due to their mutual electrostatic repulsion (correlation). The precise form of the exchange-correlation functional is yet to be determined, leading to continuous efforts in developing approximations to accurately represent this crucial element.[120,149–151]

In practical implementations of DFT, it is common to treat exchange and correlation interactions separately, which allows for the development of different functionals that can target the specific characteristics of exchange and correlation effects. However, this separation, while convenient, is not strictly accurate on a physical level for several reasons and it oversimplifies the underlying complexity. First, these interactions are coupled; the way electrons avoid each other due to the Pauli exclusion principle (exchange) is inherently related to how their movements

correlate to minimize energy (correlation). Second, the Hartree-Fock method provides a clear definition of the exchange interaction based on the antisymmetrization of the wavefunction, leading to the exact exchange operator. When DFT attempts to describe exchange based on density alone (as in the local density approximation, LDA), it uses a model derived from a completely different system: the homogeneous electron gas. In his model the volume of N electrons is characterized on a regular grid where their density is numerically evaluated, it struggles when accounting for valence electrons due to a mismatch between the actual electron density and the homogeneous electron gas approximation. Thus, it does not capture the effect of the exchange interaction as accurately as the HF approach.[161–163]

The recognition of the limitations due to the uniform electron gas assumption in the systems where the electron density has significant variations, has led to the development of more sophisticated functionals. For example, generalized gradient approximations (GGA) and meta-GGAs include density gradients and higher-order terms to better account for the non-uniformity of real electron distributions. Hybrid functionals mix DFT exchange with HF exchange to overcome some of the inaccuracies inherent in pure DFT approximations.[152]

$$E_{XC}^{hyb} = \frac{X}{100} E_X^{HF} + \left(1 - \frac{X}{100}\right) E_X^{DFT} + E_C^{DFT} \qquad \text{Equation 3.19}$$

Here, E_{XC}^{hyb} is the hybrid exchange-correlation energy, *X* is the percentage of Hartree-Fock exact exchange energy, E_X^{HF}, which is included in the hybrid functional, and E_X^{DFT} and E_C^{DFT} are exchange and correlation energies from DFT functionals respectively.

In this thesis, for example, the B3LYP and M06-2X hybrid functionals have been used, in which_X is defined as 20 and 54, respectively, indicating the percentage of Hartree-Fock exchange.[152,153]

Basis sets in DFT consist of one-particle functions that are essential for representing the electronic wavefunction of atoms and molecules. Through the LCAO-MO approximation, molecular orbitals (ϕ_i) are defined as linear combinations of atomic orbitals (f_μ), with coefficients ($c_{\mu i}$) and K representing the total atomic functions.

$$\phi_i = \sum_{\mu=1}^{K} c_{\mu i} f_\mu \qquad \text{Equation 3.20}$$

Introduced by Pople, split-valence basis sets like 6-31G, 6-31+G, or 6-311G(d) represent popular basis sets where atomic orbitals are split to balance precision and computational demand.[154] The digit before the dash in each set (6) signifies the number of Gaussian functions modeling the core electrons, on the other side the numbers indicate the number of Gaussians used for valence electrons: "31" means a double-ζ basis set for valence electrons, which each atomic orbital constructed as a combination of 3 or 1 Gaussian functions, enhancing accuracy and flexibility for valence electrons involved for example in chemical bonding, without a steep increase in computational cost. Similarly, "311" represents triple-ζ valence electron treatments, employing three sets of Gaussian functions. The "+" and "d" symbols in 6-31+G and 6-311G(d) denote the inclusion of diffuse and polarization functions, respectively. Diffuse functions allow for a more extensive spatial description of electron clouds, vital for accurately capturing behaviors in anions and polar molecules ("+" for heavy atoms' extended orbitals and an additional "+" for hydrogen), while polarization functions allow the basis set to depict changes in the electron cloud shape by adding orbitals with higher angular momentum polarization (denoted by "d" for d orbitals in heavy atoms and "d,p" for d orbitals in heavy atoms and p orbitals in hydrogen), which is important for depicting molecular geometries and reactivities more accurately.[155]

3.2.2 QM/MM methods: Additive and substractive schemes

While modern DFT can handle QM calculations for hundreds of atoms, scaling up to thousands requires strategic approximations. Hybrid QM/MM methodologies address this by partitioning systems into regions. Those regions are modeled according to its role in the overall chemical process, this approach is especially beneficial for studying enzymatic reactions, where the QM region includes substrates, key active-site residues, and sometimes cofactors, all influenced by the surrounding MM region, which may comprise the rest of the enzyme and the solvent (Figure 3.1). The main challenges with QM/MM include the selection of the scheme used for energy calculation (additive or subtractive), treatment of the boundary region, interaction modeling between the QM and MM regions, and the selection criteria for theoretical methods applicable to each region.[91]

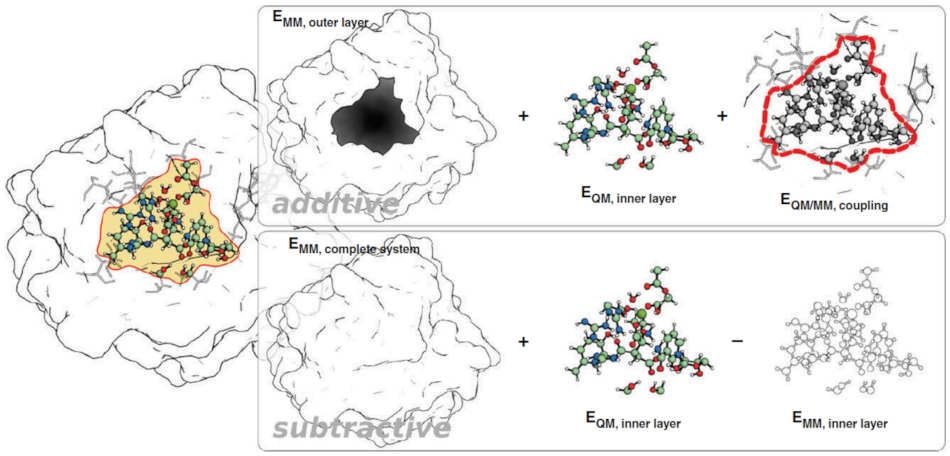


Figure 3.1 Illustration of additive and subtractive coupling approaches in the QM/MM methodology. (Figure adapted from [91])

In additive schemes, the overall system's Hamiltonian, is a sum of the Hamiltonians for each layer, alongside a coupling term that accounts for interactions between these regions. While this additive QM/MM scheme simplifies

modeling large molecular systems by employing empirical biomolecular force fields, it introduces challenges, especially in the presence of link atoms and electrostatic perturbations in the QM Hamiltonian.

$$E_{QM/MM_allsystem} = E_{QM_innersystem} + E_{MM_outersystem} + E_{QM/MM_coupling} \quad \text{Equation 3.21}$$

In subtractive schemes, independent calculations are conducted on different system parts using different theoretical levels, which are then combined through addition and subtraction to yield a corrected energy value.

$$E_{QM/MM_allsystem} = E_{MM_allsystem} - E_{MM_innersystem} + E_{QM_innersystem} \quad \text{Equation 3.22}$$

This approach facilitates the standard application of QM and MM techniques without significant modifications. Typically, three calculations are performed: the MM energy of the entire system and the energy of the inner (QM) layer calculated with both QM and MM methods to prevent double counting. Morokuma and colleagues have developed influential subtractive methods like the Integrated MO + MM (IMOMM), Integrated MO + MO (IMOMO), and the ONIOM method, enabling the integration of multiple layers and widely used in practice. These schemes eliminate the need for parameterizing inter-region interactions and mitigate systematic errors associated with lower theory levels since they are canceled out. However, defining parameters at the interface between regions, especially near polar or unsaturated contacts, poses challenges. Some software tools, like Gaussian, simplify this process by estimating MM parameters for the QM region. While subtractive methods are effective for mostly nonpolar MM regions, systems with highly polar MM regions require careful consideration due to potential issues with electronic structure influence. Methods like ONIOM

address these concerns through electronic-embedding techniques to ensure result accuracy.[156]

3.2.2.1 Interactions between the layers in ONIOM method: Mechanical, electrostatic and polarizable embedding and boundary conditions

In QM/MM calculations, accurately addressing the interaction between the QM and MM regions is essential due to the long-range nature of Coulomb interactions. While explicit consideration of every electron-electron interaction is impractical, hybrid calculations are required to approximate solutions effectively. One critical aspect is how electrons in the QM region perceive the electron charge in the MM layer, which varies from being unseen to resembling a punctual charge or a spherical smeared electron density. Conversely, in the MM Hamiltonian, the MM charge distribution interacts with electron densities rather than individual QM electrons. This perception influences the development of three main schemes: mechanical embedding, electrostatic embedding, and polarizable embedding. The mechanical embedding scheme, the simplest, considers the QM region embedded in vacuum, while the electrostatic embedding scheme introduces the polarization induced by MM charges on the QM electron density. The most complex, polarizable embedding scheme involves full mutual polarization among QM and MM regions. These schemes aim to reasonably account for polarization effects between the QM and MM regions, with computational complexity increasing from mechanical to polarizable embedding.

In mechanical embedding QM/MM calculations, the QM region is treated independently from the electrostatic effects of the MM environment. It is calculated in the gas phase, while the MM region employs classical point charge models and Lennard-Jones potentials. However, accurately representing MM parameters for the QM region poses challenges in terms of enzymatic reactions

involving charge-transfer processes. To address this, manual updating of MM charges or interpolation schemes along the potential energy surface (PES) can be employed, despite potential discontinuities in the PES. Each layer is optimized independently, with the QM layer 'blind' to the MM environment, which means the MM layer doesn't dynamically adjust based on changes in the QM atomic charges, but rather maintains a fixed response throughout the calculation. While suitable for minimum energy states, this approach may be inadequate for studying chemical events, especially near transition states. Improving accuracy by updating QM layer charges during PES exploration or considering environmental polarizing effects sacrifices computational efficiency but provides more accuracy.[91]

In electrostatic embedding schemes the polarization of the QM region by the MM charge distribution is already included in the QM calculation. This polarization is achieved through the inclusion of specific one-electron terms in the effective QM Hamiltonian, which are constructed from pre-parameterized MM point charges obtained from available MM FFs. While this improvement enhances accuracy, it also introduces complexities in implementation and leads to increased computational costs. However, the results obtained from electrostatic embedding schemes are notably better, offering several advantages over mechanical embedding. For instance, the electronic structure of the inner region can adapt to changes in the charged distribution of the environment and is automatically polarized by it, resulting in more accurate energy calculations. Furthermore, during the MM optimization step, the environment is influenced by the new charges in the QM region, contributing to improved accuracy in describing the system's behavior. Additionally, since the MM charges do not significantly change during the reaction process, as they are not directly involved in bond breaking and formation, there is no strict need for charge updates in the QM region. Despite these advantages, it's important to consider that electrostatic embedding methods still overlook the less pronounced polarization of the MM system by the QM region, which can occasionally impact results significantly. Therefore, while

electrostatic embedding represents a significant advancement in QM/MM calculations, there are still limitations to be addressed, particularly concerning polarization effects and computational efficiency. However, due to the balance between accuracy and computational cost, the optimizations of the reactant, TS and product using QM/MM methods in Chapter 6 have been performed using the electrostatic embedding approach.[91]

In polarizable embedding schemes, the mutual polarization among the QM and MM regions is included until a self-consistent charge distribution is achieved. This requires a polarizable MM FF that can respond to perturbations from the QM region. Challenges in this method include handling the QM/MM boundary and the increased computational cost due to the iterations for achieving self-consistency. In this thesis, the polarizable embedding method has been used for the excited state calculations in Chapter 4. Detailed explanations are provided in the following section under the title QM/MMPol model.[156]

Boundary schemes in QM/MM methods are crucial for handling covalent bonds between atoms in the QM and MM regions, particularly in enzymatic reactions. Two common strategies are employed: the link atom approach and the frozen orbital approach, both offering similar levels of accuracy when used appropriately. The frozen orbital approach employs localized orbitals to describe the interface between QM and MM regions. Meanwhile in the link atom approach, as it is used in this thesis, atoms in the QM region bonded to the MM part are filled to their coordination number with "link atoms," typically hydrogen atoms. These link atoms are treated explicitly in QM calculations, while in the MM calculation the charge of the MM atom in the QM-MM boundary is zeroed and its charge distributed into neighboring atoms. The frozen orbital approach employs localized orbitals to describe the interface between QM and MM regions.[157,158]

3.2.3 Determination of reaction free energy profiles in enzymes

Free energy profiles can be efficiently computed using single-conformation QM/MM calculations. This approach allows the study of reaction pathways from a single enzyme conformation, typically obtained by experimental data such as X-ray or NMR. The ONIOM methodology within this framework include the enzymatic environment, even at a lower level of theory. This can provide valuable insights into short and long-range interactions that stabilize the TS in a chemical reaction.Studies show that enzyme conformation can impact the kinetic profiles obtained from QM/MM schemes since the enzyme's conformational space is complex and time-scale dependent. As a result, a single-conformation approach can yield biased results. Despite this limitation, single-conformation QM/MM methods continue to provide valuable results in enzymatic reactions. [159,160]

In this model, QM calculations depend on the internal coordinates. The initial guess structures can be obtained from MM minimizations or MD simulations, which provide similar but slightly different conformations for the QM layer. At that stage, an adequate QM layer must represent the reactive substrate and catalytic residues, as well as other intermolecular interactions, with a special focus on the electrophiles, nucleophiles, and hydrogen-bonding residues, with careful evaluation of conserved waters. If the enzyme structure lacks the substrate, cofactors, or their analogues, docking methodologies may be necessary. At each reaction step, all involved states should be freely optimized. DFT is the preferred level of theory for QM calculations in moderately large systems, with the B3LYP or other common functionals used for general geometry calculations and exploring the PES of reaction coordinates. However, single-point energy calculations are sensitive to the choice of DFT functional.[91]

3.2.3.1 Characterization of the stationary points

Linear transit scan (LTS) can be used to identify possible intermediates, transition states, and products in a chemical reaction. Starting from an optimized enzyme-complex guess, linear transit scans are performed to explore the PES, commonly using a polarized double zeta basis set, such as 6-31G(d). After determining the initial and final values of the reaction coordinate based on the expected reaction, ONIOM calculations are performed at incremental steps of 0.05–0.10 Å. This approach helps to elucidate the key changes occurring in the system. A proper reaction coordinate's linear transit scan will show a maximum in energy along the reaction coordinate, indicating proximity to the TS. The identification of a TS is achieved by optimizing the system and analyzing the nuclear normal modes derived from the Hessian matrix of nuclear displacements, which should exhibit only one imaginary frequency.

Intrinsic reaction coordinate (IRC) scans can be utilized to trace the entire path from the TS to the reactants (in the backward direction) and to the products (in the forward direction), producing an energy profile along the reaction path. It follows the steepest descent path on the potential energy surface (PES).[161] The calculation starts at the TS and the Hessian matrix (the second derivative of the energy with respect to coordinates) is calculated to obtain the initial force constants. Mass-weighted coordinates are determined by scaling each Cartesian coordinate by the square root of its mass m_i

$$q_i = \sqrt{m_i R_i} \qquad \text{Equation 3.23}$$

The scaling ensures that the reaction path accurately reflects the inertia of the atoms. Then, the gradient vector g at the TS is calculated as the first derivative of the potential energy V with respect to the mass-weighted coordinates q:

$$g = \Delta V(q) \qquad \text{Equation 3.24}$$

The direction of the reaction path is determined by the normalized gradient vector v(s), which is: $v = -g/|g|$ for $s > 0$ (forward direction) and $v = g/|g|$ for $s < 0$ (backward direction). The path is traced by solving the differential equation:

$$\frac{dq(s)}{ds} = v(s) \qquad \text{Equation 3.25}$$

This equation is integrated numerically, meaning small steps Δs are taken along the reaction coordinate s, updating the coordinates q and recalculating the gradient g at each step.

As the path progresses, the gradient g is recalculated at each step to ensure that the path follows the correct direction of steepest descent on the PES. By focusing on the gradient and force constants to follow the path, IRC calculations can efficiently map the reaction mechanism without the need for full optimization at every point. Although IRC scans provide detailed profiles and can safely derive reagents and products from a previously optimized TS, they are more computationally intensive than standard geometry minimizations and linear transit scans. This is because IRC scans require calculating analytical force constants repeatedly to ensure that the calculation stays on track and converges to the correct pathway, which demands significant computational resources.

3.2.3.2 Establishing the reaction profile

The final step in QM/MM enzymatic studies involves the thermodynamic characterization of stationary states in the chemical space, aiming to generate free-energy profiles comparable with experimental data. This involves calculating the

electronic energy for the reaction, zero-point energy, and entropy at the desired temperature. Frequency calculations are performed at all reaction stationary points to account for thermal and entropic corrections, often using the same level of theory as in exploring the chemical space, such as B3LYP/6-31G(d).[152]

The choice of basis set and density functional is crucial for calculating the electronic single point energy. Typically, more complete basis sets are used, such as triple-split valence Pople basis sets with diffuse and polarization functions, e.g. 6-311+G(2d,2p). Over-polarization of the wave function should be managed carefully due to the MM environment's interaction. Empirical dispersion corrections may be introduced to improve the inaccurate description of dispersion in DFT.[152]

3.2.4 QM/MM molecular dynamics

Another challenge of hybrid QM/MM methods is the representation of molecular conformational diversity, which is essential for accurately capturing the system's behavior. Traditional QM/MM applications often lack time-dependence, which limits their ability to account for dynamic processes like structural fluctuations during chemical reactions or excited state non-adiabatic processes. To address this, approaches combining QM/MM with MD simulations or employing time-dependent potentials have emerged. These methods, such as multi-conformation QM/MM, QM/MM Car-Parrinello MD[162,163], and Born-Oppenheimer MD,[164] improve the statistical representability of the system's phase space and allow for exploration of non-adiabatic regimes.

3.2.4.1 Born-Oppenheimer molecular dynamics

The dynamics of molecular systems in BOMD is simulated based on the Born-Oppenheimer approximation, which separates nuclear and electronic motions

within molecules. This approximation suggests that electronic motion occurs much more rapidly than nuclear motion, allowing the electronic configuration of a molecule to adjust instantly to changes in nuclear positions, resulting in the total energy being expressed solely as a function of nuclear coordinates.

In a BOMD simulation, the molecular system begins with defined atomic positions and velocities. The total energy is computed using a QM method, typically DFT or Hartree-Fock, to account for electronic structure. Classical equations of motion, such as Newton's equations, are then solved for the atomic nuclei, guided by forces derived from the PES obtained from the QM calculation. By integrating these equations over time, the simulation predicts the temporal evolution of the system's atomic positions and velocities. BOMD simulations offer insights into atomic-level structural changes and the quantum mechanical effects on the system's dynamics, which are often challenging to observe experimentally.

3.3 Excited states from QM/MM TD-DFT

Time-dependent density functional theory (TD-DFT) extends traditional DFT to the QM analysis of excited states by calculating excitation energies and allowing simulation of absorption spectra and other excited state properties. It’s crucial for investigating electronic dynamics in such states. TD-DFT applies a time-dependent Hamiltonian to the time-dependent Schrödinger equation, following the Runge-Gross theorem that links time-varying potentials with electron densities. TD-DFT examine dynamic properties and reactions to changing fields, unlike the static nature of ground-state DFT. TD-DFT is specialized in modeling transitions and includes a time-dependent exchange-correlation potential, key for precise excitation descriptions, and in calculating properties of excited states like dipole moments and emission geometries.[156]

The linear response (LR) formulation of TD-DFT is one of the most common methods for predicting electronic excitations and examining optical properties, offering a computationally efficient way to understand complex systems. It uses the ground state wavefunction to assess the system's response to small, time-variant perturbations and identifies transition energies using frequency-dependent response functions. Implementing LR in TD-DFT involves first solving the Kohn-Sham equations for the ground state electron density, then introducing a perturbation, and calculating the modified response to include electron interactions. Solving Casida equations then yields excitation energies and transition amplitudes, which facilitate physical property calculations like oscillator strengths and excited state geometries, thus providing insights into molecular excited state properties. [165]

TD-DFT has been adapted to include solvent effects using continuum solvation models, a development that has been widely implemented and is commonly used in spectral analysis. However, explicit descriptions of the environment in QM/MM TD-DFT implementations are more appropriate to study excited states in biosystems, as they allow to include the impact of specific interactions with the environment, which are key to understand the function of biochromophores.[156]

3.3.1 QM/MMPol model

QM/MM models are further refined when they account for the mutual polarization between the quantum and molecular mechanics subsystems. This advancement is achieved through the adoption of a polarizable force field. In this enhancement, the conventional approach primarily relies on point charges to describe electrostatic interactions which is supplemented with a polarization component, thus providing a more comprehensive representation of intermolecular forces. The polarizable QM/MM (QM/MMPol) model that is used in our group, developed in

collaboration with Mennucci's group at the University of Pisa,[156] incorporates this intricate aspect of mutual polarization.[20,166–168]

Within the QM/MMPol framework, an explicit term representing the solute-solvent interaction is introduced into the effective Hamiltonian (H_{eff}). This model stands apart from other approaches wherein the solvent is typically depicted as a continuous polarizable medium, defined only by its macroscopic dielectric properties. Instead, the QM/MMPol approach describes the MM system with a classical polarizable force field based on the induced dipole model. This key distinction allows for a more detailed and dynamic interaction between the solute and solvent, reflecting the discrete and anisotropic nature of the molecular environment. But, more importantly, it allows to include the ultrafast electronic response of the environment polarization to electronic excitations.

In particular, electrostatic forces within the MM system are characterized by atomic partial charges. These charges form the baseline for interaction forces but do not account for the reaction of the system to changes in the electronic state of a solute. To capture this response, isotropic polarizabilities are explicitly added to the MM atoms, which introduces a level of dynamism into the MM system. This allows for the modeling of how the environment electronic cloud can be distorted, thus accounting for the polarization effects that are critical for a realistic depiction of molecular interactions.

$$\widehat{H}_{eff}|\Psi\rangle = (\widehat{H}_0 + \widehat{H}_{QMM/MM} + \widehat{H}_{MM})|\Psi\rangle. = E|\Psi\rangle. \qquad \text{Equation 3.26}$$

The gas phase Hamiltonian of the QM solute system is denoted as H_0, capturing the inherent quantum mechanical behavior of the solute in the absence of environment interactions. The further refinements include the terms $H_{QM/MM}$ and

H_{MM} to encapsulate the interactions and energies that arise from the MM environment.

$$\widehat{H}_{QM/MM} = \widehat{H}^{el}_{QM/MM} + \widehat{H}^{pol}_{QM/MM}$$
$$= \sum_m q_m \hat{V}(r_m) - \frac{1}{2} \sum_a \mu_a^{ind} \hat{\mathrm{E}}_a^{solute}(r_a) \quad \text{Equation 3.27}$$

$$\widehat{H}_{MM} = \widehat{H}^{el}_{MM} + \widehat{H}^{pol}_{MM}$$
$$= \sum_m \sum_{n>m} \frac{q_m q_n}{r_{mn}} - \frac{1}{2} \sum_a \mu_a^{ind} \sum_m \frac{q_m (r_a - r_m)}{|r_a - r_m|^3} \quad \text{Equation 3.28}$$

Where $\hat{V}(r_m)$ represents the electrostatic potential operator, and $\hat{\mathrm{E}}_a^{solute}(r_a)$ is the electric field operator at the MM sites due to the electrons and nuclei of the QM system. Together, these operators model the influence of the QM region on the classical MM region.

The indices *m* and *n* run over the MM charges, denoted as q_m and q_n, and the index *a* runs over the induced dipoles, μ_a^{ind}, which are situated at their respective positions, r_m and r_a. These induced dipoles arise as a response to the electrostatic potential and field of the QM system, embodying the mutual polarization effects within the MM region.

In the effective Hamiltonian, $\widehat{H}^{el}_{QM/MM}$ and $\widehat{H}^{pol}_{QM/MM}$ encapsulate the interactions of the QM system with the MM charges and the induced dipoles, respectively. Meanwhile, $\widehat{H}^{el}_{MM}$ indicates the electrostatic self-energy of the MM region's charges, which is a fixed energetic constant for a given conformation. The $\widehat{H}^{pol}_{MM}$ term, covering polarization interactions occuring between these charges and the induced dipoles, is specifically included in the Fock operator due to the dependency of induced dipoles on the QM wave function, allowing for dynamic reflection of the QM region's electronic state.

The induced dipoles, situated on each polarizable atom of the MM region, are defined by the equation.

$$\mu_a^{ind} = \alpha_a(\mathrm{E}_a^{solute} + \mathrm{E}_a^{solvent}\{q\,;\mu^{ind}\} \qquad \text{Equation 3.29}$$

where a linear approximation is assumed and an isotropic polarizability is used for each point in the MM system, denoted as α_a. This takes into account the total electric field from the environment at each site in the MM region, neglecting any magnetic contributions. The electric field, $\mathrm{E}_a^{solvent}$, includes the sum of contributions from the point charges and the induced dipoles within the system, establishing that each induced dipole moment is dependent on all other induced dipoles in the solvent.

The application of this methodology requires Eq. 3.29 to be solved within each self-consistent field (SCF) iteration to account for the described polarization interactions. The calculation only includes the polarization effects of the atoms in the MM region up to a certain distance limit relative to the QM region, as the influence of polarization diminishes beyond 15-20 Å, having a negligible impact on the energy and properties of the QM system, thereby significantly reducing computational costs.

The creation of a polarizable hybrid QM/MM model has significantly advanced the study of excited states and electronic energy transfer processes in varied environments.[20,166–171] A key contribution of this model is the possibility to compute of electronic couplings mediating energy transfer, thus allowing the account for dielectric screening effects that are otherwise absent in electrostatic embedding. It also allows for changes in solvatochromic shifts in chromophores due to polarization.

3.3.2 Excitonic properties from QM/MMPol

This section describes the methods to estimate excitonic parameters in photosynthetic pigment-protein complexes using QM/MMPol.[20,168,171] These parameters include the excitation energies, which are the energy levels associated with the absorption of photons by the pigments; the electronic couplings, which refer to the interactions between electronic states of different pigments facilitating energy transfer and exciton delocalization; and the spectral density of vibronic coupling, which characterizes how electronic transitions are influenced by vibrations within the nuclear framework of the system. To accurately determine these parameters, a combination of experimental techniques such as spectroscopy, alongside computational methods including QM and MD simulations, are employed. These methods provide a detailed understanding of the energetics and dynamics of exciton states and energy transfer processes, which are crucial for the efficiency of photosynthesis.[20,168,171]

3.3.2.1 Excitation energies

Site energies, the uncoupled excitation energies of the pigments, define the energy levels where chromophores absorb light, determined by their electronic structure and environmental interactions. These energies are crucial in exciton models, as they define the states that characterize light harvesting processes in photosynthesis or in artificial light harvesting systems. However, calculating site energies is challenging due to the large size of the pigments involved (sometimes over 90 atoms) which complicates using precise QM methods. Moreover, site energies are profoundly affected by the surrounding molecular environment. Therefore, in the QM method it is important to integrate models that account for environmental interactions. To achieve this, a linear response theory within the TD-DFT framework is utilized. The method enables the determination of excitation energies through the analysis of system responses to perturbative external fields. Central to

this methodology is the Hessian matrix, composed of matrices A and B, which reflect the second-order derivatives of the electronic energy with respect to changes in electron density. The excitation energies can be obtained through the following expression:

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix}\begin{pmatrix} X_n \\ Y_n \end{pmatrix} = \omega_n \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \qquad \text{Equation 3.30}$$

The transition vectors, (Xn, Yn), correspond to the collective eigenmodes of the density matrix, with eigenfrequencies ω_n indicating the natural oscillations of the electronic structure in response to excitation.

The effect of the polarizable environment is incorporated by considering the MM-induced dipoles in the computation of the A and B matrices. This accounts for the environment's influence on electronic transitions, an effect neglected in standard TD-DFT computations based on QM/MM electrostatic embedding schemes. The modification of these matrices involves adding the contributions from the coupling matrix $K_{ai,bj}$ and the polarizable MM matrix $C^{pol}_{ai,bj}$, which mediate the interactions within the molecule and between the molecule and its environment, following this form;

$$A_{ai,bj} = \delta_{ab}\delta_{ij}(\varepsilon_a - \varepsilon_i) + K_{ai,bj} + C^{pol}_{ai,bj} \qquad \text{Equation 3.31}$$

$$B_{ai,bj} = K_{ai,bj} + C^{pol}_{ai,bj} \qquad \text{Equation 3.32}$$

where ε are the energies of the orbitals, with i,j representing occupied molecular orbitals and a,b the virtual ones. $K_{ai,bj}$ and $C^{pol}_{ai,bj}$ are the coupling matrix and the polarizable MM matrix, respectively.
These matrices are further defined as:

$$K_{ai,bj} = \int dr \int dr' \, \phi_i(r') \phi_i^*(r') \left(\frac{1}{|r'-r|} + g_{XC}(r', r) \right) \phi_j(r) \phi_b^*(r) \quad \text{Eq. 3.33}$$

$$C_{ai,bj}^{pol} = -\sum_k \left(\int dr \, \phi_i(r) \, \phi_a^*(r) \frac{(r_k - r)}{|r_k - r|^3} \right) \mu_k^{ind} \, (\phi_j \, \phi_b^*) \quad \text{Eq. 3.34}$$

The term $K_{ai,bj}$, representing the electronic coupling within the molecule, is crucial for understanding how electronic energy is shared between different parts of the molecule. The integral expression calculates the interactions of molecular orbitals with each other, considering both the direct Coulomb interaction and the exchange-correlation interaction. The term $C_{ai,bj}^{pol}$, the polarization response of the environment, involves the integral over the induced dipoles μ_k^{ind} in the environment caused by the electronic transitions within the molecule. The index k in equation $C_{ai,bj}^{pol}$ represents the total number of polarizable atoms in the MM region.

3.3.2.2 Electronic couplings

Electronic coupling is key for simulating exciton states and energy transfer in biological systems.[20,171] The magnitude of this coupling, relative to vibronic coupling (interaction with environmental vibrations), determines whether the system's states are localized or delocalized, forming excitons. Vibronic coupling involves the interaction between electronic states and the vibrational modes of the environment. Strong vibronic coupling keeps energy states localized within specific molecules, while strong electronic coupling delocalizes energy states across molecules, leading to excitons.

Formally, electronic coupling describes the matrix element between the initial and final states of the process:

$$V_{DA} = \langle D^*A \mid \widehat{H}_{el} \mid DA^* \rangle \qquad \text{Equation 3.35}$$

where $\widehat{H}_{el}$ is the electronic Hamiltonian of the system, while D^*A and DA^* are the localized excited states in D (donor) and A (acceptor). This term can be formulated based on a perturbative expression as the sum of three contributions:

$$V_{DA} = V_{coul} + V_{exch} + V_{ovlp} \qquad \text{Equation 3.36}$$

where V_{coul} is the Coulombic interaction between the transition densities, V_{exch} is the exchange interaction, and V_{ovlp} is proportional to the overlap between the transition densities.

The transition density (ρ^T) measures the distribution of electronic charge during the transition from the ground state to an excited state. The Coulombic interaction V_{coul} representing the electrostatic interaction between the transition densities of the donor (D) and acceptor (A) molecules, is given by:

$$V_{DA}^{coul} = \int dr \int dr' \, \rho_D^{T^*}(r') \frac{1}{|r-r'|} \rho_A^T(r) \qquad \text{Equation 3.37}$$

where $\rho_D^{T^*}$ is the complex conjugate of the transition density of the donor at position r', and $\frac{1}{|r-r'|}$ represents the Coulombic interaction potential between charges at positions r and r'. The transition density $\rho_A^T(r)$ corresponds to the diagonal part of the density matrix constructed from the wave functions of the ground state Ψ_{GS} and the excited state (Ψ_{ES}) by integrating over all other coordinates, and is given by;

$$\gamma_{ES-GS}^T(r,r') = \mathrm{N} \int .. \int \Psi_{ES}^*(r, r_2 \ldots r_N) \Psi_{GS}(r', r_2 \ldots r_N) dr_2 dr_3 .. dr_N \qquad \text{Eq 3.38}$$

where $\gamma^{T}_{ES-GS}(r,r')$ represents the transition density matrix element between positions r and r' and N is the normalization constant. The transition density $\rho^{T}_{ES-GS}(r)$ is then obtained by setting $r = r'$:

$$\rho^{T}_{ES-GS}(r) = \gamma^{T}_{ES-GS}(r,r) \qquad \text{Equation 3.39}$$

The exchange interaction (V^{exch}_{DA}) describes the short-range interaction based on wavefunction overlap, significant for Dexter's energy transfer mechanism. Dexter's mechanism involves the simultaneous transfer of an electron from the donor to the acceptor and vice versa, making it a short-range interaction. Lastly, the overlap term (V_{ovlp}) is proportional to the direct spatial overlap of the transition densities of the donor and acceptor.

For energy transfers between singlet states (Förster mechanism), the electronic coupling is dominated by its Coulombic component. In forbidden excitations or those involving triplet states, the exchange component (Dexter interaction) becomes significant. For precise estimation of the coupling in such cases, calculations of excited states on the supermolecule D+A are necessary. In symmetric systems, the electronic coupling corresponds to half the excitonic splitting between the resonant states involved:

$$|V| = \frac{|E_1 - E_2|}{2} \qquad \text{Equation 3.40}$$

Other methods involve diabatization of the states and direct calculation of their coupling. This thesis focuses on excitonic coupling between singlet states in cryptophyte photosynthetic complexes, where the Coulombic term dominates, allowing short-range terms to be neglected.

The thesis calculates these terms using TD-DFT QM/MMPol calculations, which allow to overcome the limits of Förster point dipole approximation (PDA) commonly used in energy transfer studies, which gives the R^{-6} dependency of the transfer rate in Förster theory. The calculation of Coulombic couplings with the PDA approximation are given by:

$$V_{DA}^{PDA} = \frac{\kappa \mu_D \mu_A}{R_{DA}^3} \qquad \text{Equation 3.41}$$

$$\kappa = \hat{\mu}_D \,.\hat{\mu}_A - 3\left(\hat{\mu}_D.\hat{R}_{DA}\right)\left(\hat{\mu}_A.\hat{R}_{DA}\right) \qquad \text{Equation 3.42}$$

where $\hat{\mu}_D$ and $\hat{\mu}_A$ are unit vectors of the transition dipoles, $R_{DA}$is the distance between the dipoles, and $\varkappa$ represents their relative orientation.

Furthermore, the environment significantly affects electronic coupling by modifying geometric and electronic structures, impacting excitation energies, densities, and transition dipoles. It induces strong screening on the Coulombic interaction between donor (D) and acceptor (A). Förster's theory describes this with a screening factor $s = 1/\varepsilon_{opt}$, where $\varepsilon_{opt} = n^2$ (n is the index of refraction). Typically, $\varepsilon_{opt} = 2$ is used in biomolecules, reducing energy transfer rates by a factor of 4. In some studies, other screening factors are used based on models of two dipoles located in spherical cavities inside the dielectric, given by $s = 3/(2\varepsilon_{opt} + 1)$. Hybrid QM/classical methods like QM/MMPol have shown however that screening also depends on molecular shape, orientation, and distance.

The QM/MMPol model adopts a polarizable atomistic description of the classical MM environment, expressing coupling as the sum of the Coulombic term and an explicit term mediated by the MM environment:

$$V_{DA} = V_{DA}^{coul} + V_{DA}^{MMPol} \qquad \text{Equation 3.43}$$

$$V_{DA}^{MMPol} = -\sum_k \left(\int dr\, \rho_D^{T*}(r') \frac{(r' - r_k)}{|r' - r_k|^3} \right) \mu_k^{ind}(\rho_A^T) \qquad \text{Equation 3.44}$$

where the MMPol term (V_{DA}^{MMPol}) accounts for the interaction mediated by the induced dipoles (μ_k^{ind}) in the environment due to the presence of donor and acceptor molecules.

The screening factor (s), quantifies the effect of the environment on the electronic coupling:

$$s = \frac{V_{DA}}{V_{DA}^{coul}} = \frac{V_{DA}^{coul} + V_{DA}^{MMPol}}{V_{DA}^{coul}} \qquad \text{Equation 3.45}$$

This ratio shows how much the coupling is altered by the polarizable environment compared to the Coulombic interaction alone.

3.3.2.3 Spectral densities

The dynamic properties of the environment play a crucial role in light harvesting and electronic energy transfer (EET) processes, where interactions between electronic states and nuclear vibrations, such as those in proteins and solvents, cause significant fluctuations in the system's energy levels.[20,168,171,172] These fluctuations can lead to exciton localization, where collective exciton states dynamically localize in specific chromophores. The spectral density function $J_i(\omega)$ captures these dynamic properties by quantifying the coupling strength between the system and environmental vibrations across different frequencies. It is generally divided into two components:

$$J_i(\omega) = J_{0,i}(\omega) + J_{vib,i}(\omega) \qquad \text{Equation 3.46}$$

Here, $J_{0,i}(\omega)$ accounts for the coupling of electronic transitions to a continuum of low-frequency vibrational modes from the protein and solvent, while $J_{vib,i}(\omega)$ pertains to the coupling with high-frequency intramolecular vibrations. Spectroscopic techniques like spectral hole-burning and fluorescence line-narrowing can measure the spectral density function experimentally, though these methods typically yield data only for the lowest energy exciton or an averaged state.

Computational techniques for estimating spectral density are primarily categorized into methods utilizing MD simulations and those based on normal mode analysis (NMA).[171] MD-based methods involve simulating the ground state using classical MD while computing the excited states with QM/MM calculations along the trajectory. This MD-QM/MM approach captures the temporal evolution of excitation energies by performing QM/MM calculations at every 2-5 femtosecond interval, which tracks the rapid vibrational oscillations of heavy atom bonds. However, this method is computationally intensive, limiting its practical application.

MD-based techniques enable the simultaneous determination of the two components $J_{0,i}(\omega)$ and $J_{vib,i}(\omega)$ of the spectral density function, derived from the Fourier transform of the autocorrelation function of the fluctuations in excitation energies:

$$J_i(\omega) = \frac{\beta\omega}{\pi}\int_0^\infty C_i^{cl}(t)\cos(\omega t)\,dt \qquad \text{Equation 3.47}$$

Where $\beta = \frac{1}{K_B}T$ and $C_i^{cl}(t)$ is the classical autocorrelation function of the fluctuations in excitation energies:

$$C_i^{cl}(t_j) = \frac{1}{N-j}\sum_{k=1}^{N-j}\Delta E_i(t_j + t_k)\,\Delta E_i(t_k) \qquad \text{Equation 3.48}$$

A limitation of MD-based methods is their reliance on classical force fields, which can overestimate high-frequency vibration couplings. This can be mitigated by developing specialized force fields or using QM/MM MD simulations that model the system at a QM level.

Alternatively, the vertical gradient (VG) method[29] incorporating NMA offers a computationally efficient approach to calculate spectral densities. The VG method assumes that both ground and excited state PES are harmonic but with shifted equilibrium positions. This enables the straightforward extraction of reorganization energy from the differences between vertical and adiabatic excitation energies. NMA involves computing normal modes from optimized geometries and diagonalizing the Hessian matrix to understand atomic displacements around equilibrium positions. This technique has been used to model the intramolecular contributions to the spectral density function, $J_{vib,i}(\omega)$, in this thesis Meanwhile, the continuous part of the function, $J_{0,i}(\omega)$, is introduced via an overdamped Brownian oscillator model.

The PES of the ground (H_g) and excited (H_e) states are given by:

$$H_g = \sum_{i=1}^{M}\left(\frac{p_i^2}{2m_i} + \frac{1}{2}\,m_i\omega_i^2 q_i^2\right) \qquad \text{Equation 3.49}$$

$$H_e = \omega_{eg}^0 + \sum_{i=1}^{M}\left(\frac{p_i^2}{2m_i} + \frac{1}{2}\,m_i\omega_i^2 (q_i - d_i)^2\right) \qquad \text{Equation 3.50}$$

The energy gap gradient U, derived from the PES difference, is expressed as:

$$U(\tilde{q}) = H_e - H_g = \omega_{eg}^0 - \sum_i \omega_i^2\,\tilde{d}_i\tilde{q}_i + \frac{1}{2}\,\omega_i^2\tilde{d}_i^2 \qquad \text{Equation 3.51}$$

Here, $d_i = -\frac{f_i}{\omega_i^2}$, is the displacement for mode i, where f_i is the vertical gradient in normal coordinates, and ω_{eg}^0 is the frequency of the vibrational mode. These gradients, determined from Cartesian coordinates and the nuclear mass matrix, yield the Huang-Rhys factors, which quantify the coupling strength between electronic transitions and vibrational mode, and contribute to the spectral density function.

The Huang-Rhys factor S_i for each mode *i* is calculated as

$$S_i = \frac{\tilde{f}_i^2}{2\hbar\omega_i^3} \qquad \text{Equation 3.52}$$

The vibrational contributions to the spectral density function $(J_{vib,i}(\omega))$ are then given by this expression:

$$\tilde{C}''(\omega) = \pi\omega \sum_{k=0} S_i \omega_i \qquad \text{Equation 3.53}$$

Protein fluctuations can be included by performing multiple QM/MM frequency calculations along and MD trajectory, where the chromophore is placed in a frozen environment, and only the vibrations of the chromophore are explicitly computed in the NMA analysis.

The continuous contribution to the spectral density $(J_{0,i}(\omega))$ due to slow environmental motions can be modeled using an overdamped Brownian oscillator:

$$\tilde{C}_{OBO}''(\omega) = \pi\lambda\omega \frac{\gamma}{\omega^2+\gamma^2} \qquad \text{Equation 3.54}$$

Here, λ represents the total reorganization energy needed to reorganize nuclear positions due to changes in electronic states, while γ is the attenuation constant,

indicating the decay rate of these fluctuations. This model accurately reflects the experimental Stokes shift between absorption and emission spectra. One challenge of these methods is the approximation to a single minimum in the normal mode analysis, which is particularly problematic for biomolecules characterized by PES with multiple minima.

3.3.3 Excitonic spectra

Simulations of absorption, circular dichroism, and fluorescence spectra in this thesis were conducted using the site energies, electronic couplings, and spectral densities as described in the previous sections using the EXAT[173] and the FCE[174,175] codes. EXAT simulations were based on modified Redfield theory, whereas FCE ones include non-Markovian and non-secular effects. Static disorder was modeled by averaging realizations of the spectra over a random distribution of the site energies, characterized by a given standard deviation σ, which was adjusted to reproduce the broadening of the experimental emission lineshapes for each complex.

The following excitonic Hamiltonian was used to describe the multichromophoric system:

$$\hat{\mathrm{H}} = \sum_{n=1}^{N} \epsilon_n \mid n\rangle \langle n \mid + \sum_{n \neq m}^{N} V_{nm} \mid n\rangle \langle m \mid \qquad \text{Equation 3.55}$$

where N is the number of interacting chromophores, ϵ_n is the excitation energy of the n^{th} chromophore, and V_{nm} is the coupling between the n^{th} and m^{th} chromophores. Exciton states $|k\rangle$ were then obtained from the diagonalization of the excitonic Hamiltonian:

$$\hat{\mathrm{H}} = \sum_{k=1}^{N} \epsilon_k \mid k\rangle \langle k \mid ; \mid k\rangle = \sum_{n=1}^{N} C_n^k \mid n\rangle \qquad \text{Equation 3.56}$$

where ϵ_k is the energy of the kth exciton state and C_n^k describes the participation of the nth chromophore in the kth exciton state.

This formalism allows for the detailed simulation of various spectroscopic properties of multichromophoric systems by considering both the electronic structure and the dynamic interactions between chromophores.

CHAPTER 4.
MULTISCALE MODELLING OF PHOTOSYNTHETIC COMPLEXES

The contents of this chapter were reported in the following publication:

Unraveling the role of thermal fluctuations on the exciton structure of the cryptophyte PC612 and PC645 photosynthetic antenna complexes
B. Ozaydin and C. Curutchet
Frontiers in Molecular Biosciences **2023**, *10*, 1268278

Author contribution: The author of this thesis performed all the simulations reported in the article and in this chapter, performed the analysis of the results and was the leading author in the writing of the manuscript.

4.1 Background

As discussed in Section 1.2.3, we recently investigated the exciton structure in PC577, PC612, PC630, and PC645 cryptophyte complexes using multiscale QM/MMPol calculations.[29,35] However, inaccuracies related to the choice of input geometries for excited state calculations severely hamper these protocols.[29,171,176,177] Inclusion of thermal effects was done by performing classical MD simulations of the protein complex, followed by post-processing to compute excited states with QM/MM models.[178] Alternatively, QM/MM geometry optimization of the pigments were done, significantly improving results in PC577, PC612, PC630, and PC645 complexes, although at the expense of neglecting thermal effects.[29] however, important discrepancies with experimental data remained, such as unbalanced descriptions of the relative intensity of low and high-energy bands in PC577 and PC612, or a strong underestimation of the MBV bands in PC630 and PC645. This highlights the impact of thermal effects and the limits of rationalizing spectral properties based on X-ray crystal structure.[29,30,178]

In this Chapter, we investigate the exciton properties of the Hemiselmis virescens PC612 and the Chroomonas PC645 antenna complexes shown in Fig. 4.1 using a dual-MD multiscale protocol that combines classical MD and BOMD simulations based on DFT. QM/MM MD simulations provide a highly accurate description of the spectral density of electron-vibrational coupling in photosynthetic complexes, allowing for an accurate treatment of thermal effects.[174,175] We also compare this approach with an alternative protocol based on multiple QM/MM geometry optimizations along the trajectories.

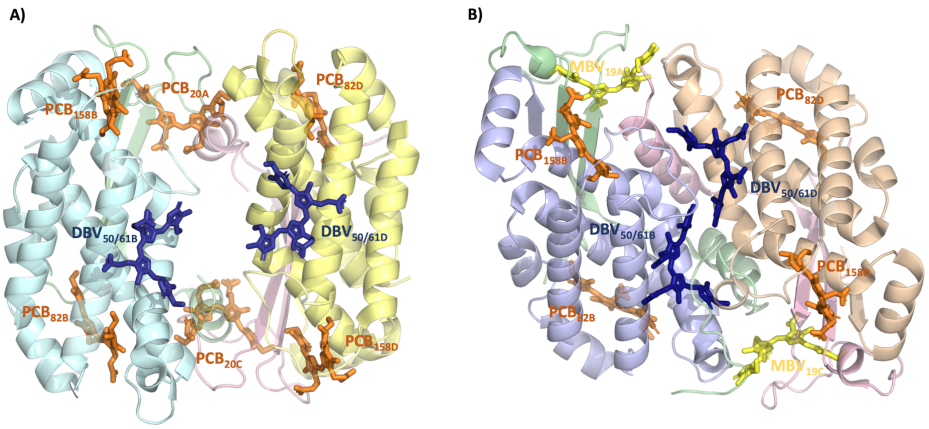


Figure 4.1. Structure and pigment composition of the phycocyanin cryptophyte antenna complexes a) PC612 and b) PC645.

Additionally, we adopt the full Cumulant Expansion (FCE) formalism to describe the absorption and circular dichroism spectra,[174,175] which includes non-Markovian and non-secular effects that are neglected in modified Redfield theory. Overall, this new protocol yielded absorption spectra with unprecedented accuracy for PC612 and PC645, providing an accurate excitonic Hamiltonian for future investigations into the implications of open and closed quaternary structures on phycocyanin light-harvesting properties.

4.2 Computational details

4.2.1 Simulated systems, MD simulations and geometry optimizations

We based our simulations on high-quality crystal structures obtained for Chroomonas PC645 (PDB 4LMS, resolution 1.35 Å) and Hemiselmis virescens PC612 (PDB 4LM6, resolution 1.8 Å) reported by Curmi and co-workers, with the PC numbering indicating their maximum absorption wavelength.[28] The preparation of the systems and the details of the classical MD simulations are described in

previous studies.[29,35] Both complexes have four polypeptide chains, but PC612 is organized as an $(\alpha\beta)_2$ homodimer, while PC645 displays a $(\alpha_L\beta)\cdot(\alpha_S\beta)$ organization, where α_L and α_S indicate long and short α subunits, respectively.[30] Additionally, PC612 features an open quaternary structure due to a ~73° rotation of the αβ protomers, whereas PC645 exhibits a closed configuration more commonly found in cryptophyte antennae. In terms of pigment composition, PC612 contains six PCBs and two DBVs, while in PC645, two of the PCBs are replaced by MBVs. These pigments are covalently linked to a Cys, except for DBVs, which involve two covalent links with Cys residues. Therefore, pigments are denoted as PCB82D or DBV50/61B, with A/C and B/D indicating α and β chains, respectively, and numbering referring to the Cys attachment residues in that chain. All bilins were modeled with a fully protonated tetrapyrrole backbone and anionic propionic side chains, while all amino acids were considered in their standard protonation state, except His21 C in PC645.

We started from the 200 ns classical MD simulations reported in Corbella et al. (2019),[29] performed using the ff14SB (Maier et al., 2015),[179] TIP3P,[180] and GAFF[99] force fields for the protein, water, and bilin chromophores, respectively. Classical NVT simulations at 300 K were extended to a total of 1 μs, using a time step of 2 fs, the SHAKE algorithm, periodic boundary conditions, the particle-mesh Ewald approach, and a non-bonded cutoff of 8 Å. From the last 500 ns, we extracted 34 frames at 14 ns intervals to be used as starting conditions for BOMD QM/MM simulations. Then, from each frame, we ran eight independent 1 ps BOMD simulations, where one of the bilin pigments was described at the B3LYP/6-31G level of theory, using a 0.5 fs time step. From the last 500 fs of BOMD trajectories, we extracted 25 frames at regular time intervals to be used as input geometries in QM/MMPol excited state calculations. This protocol is referred to as MD-BOMD. All runs were performed with the Amber20[181] - Gaussian 16 interface.[182,183]

On the other hand, QM/MM optimizations were independently performed for each chromophore geometry starting from the 34 frames extracted from the classical trajectories. We adopted the ONIOM subtractive scheme with electrostatic embedding as implemented in Gaussian.[183] In each optimization, a given pigment and selected amino acids interacting with it were fully relaxed and described at the B3LYP/6-31G(d) level of theory, while the rest of the system was kept frozen and described with the classical force field adopted in the MD. For PCBs and DBVs in both complexes, we included the Asp or Glu side chains coordinating the central rings, whereas for MBVs in PC645, we included selected amino acids (Asn22 A for MBV19A and His21 C and Glu25 C for MBV19C). QM/MM boundaries were defined at the residue-residue and Cys-bilin bonds using the link atom scheme.[184] The optimized geometries were used as inputs in subsequent QM/MMPol excited state calculations. This protocol is denoted as MD-OPT.

4.2.2 QM/MMPol calculations

Excited state calculations were performed using the QM/MMPol model[156] in its time-dependent density functional theory (TD-DFT) linear response formulation. In the MMPol model, the environment is described using a polarizable force field based on the induced point dipole model, which assigns point charges and isotropic polarizabilities to the environment atoms and accounts for full mutual polarization effects between the QM and MM regions. QM/MMPol calculations on structures extracted from BOMD trajectories and ONIOM optimizations were performed at the TD-CAM-B3LYP/6-31G(d) level of theory.[185] Previous work showed that different global and range-separated hybrid functionals with varying degrees of exact exchange (M06, M06-2X, and ωB97XD) led to similar results in phycobiliproteins.[29]

The MMPol region was described using the Amber pol12 AL parameters,[114,115] and QM/MM boundaries in the bilin-Cys bonds were treated using the link atom scheme.[184] Atomic charges for water and bilins, consistent with the Amber pol12 AL polarizabilities, were derived as described in Corbella et al. (2019).[29] Explicit polarization was limited to residues within a cutoff radius of 12 Å from the QM heavy atoms (MMPol region), while residues up to 35 Å were included using the additive force field (MM region). MMPol calculations were performed using a locally modified development version of the Gaussian package.[182]

4.2.3 Modelling of AD and CD pectra

Simulations of absorption (OD), circular dichroism (CD), and fluorescence (FLU) spectra were performed using the EXAT[173] and FCE codes.[175,186] EXAT simulations were based on modified Redfield theory, whereas FCE simulations included non-Markovian and non-secular effects as described in Ma and Cao (2015) and Cupellini et al. (2020).[174,175]

The excitonic Hamiltonian used to describe the multichromophoric system is as in the equation 3.55. Exciton states were obtained from the diagonalization of the excitonic Hamiltonian as it is indicated in the equation 3.56. We used the site energies averaged for each pigment over the 850 or 34 frames calculated using the dual quantum/classical MD (MD-BOMD set) or the ONIOM optimization protocols (MD-OPT set), respectively. Results were also compared to the energies estimated previously based on classical MD (MD set) or a single optimization performed on the crystal structure (Crystal-OPT set) in Corbella et al. (2019).[29] Electronic couplings, transition dipole moments, and pigment center coordinates were taken from previous QM/MMPol/ddCOSMO calculations based on the Crystal-OPT protocol, including dielectric screening effects exerted by the

surrounding environment. Beyond the DBV central pair in PC645, the pigment pairs in these complexes are not strongly coupled, and adopting the MD and Crystal-OPT sets of couplings and dipoles led to negligible differences in predicted spectra. Even the complete neglect of couplings leads to small changes, as shown in Figure A1.13 (see Appendix 1). For this reason, we did not recompute the couplings along the many MD-OPT optimizations. We also adopted the transition dipole moments from the Crystal-OPT set for consistency, but these values are very similar to those obtained by the MD-OPT and MD-BOMD protocols, as shown in Table A1.4 (see Appendix 1), so this choice leads to negligible changes in predicted spectra. On the other hand, BOMD simulations are performed for single pigments, so recalculation of electronic couplings on those geometries would lead to an unbalanced description of the pigment pair.

Electronic-vibrational coupling was accounted for using spectral densities for the intramolecular part calculated using the Vertical Gradient approach, while the continuous part due to slow environmental motions was added a posteriori using an overdamped Brownian oscillator with parameters adjusted to reproduce the Stokes shift, as described in Corbella et al. (2019).[29] Finally, static disorder was modeled by averaging realizations of the spectra over a random distribution of site energies characterized by a given standard deviation σ, which was adjusted to reproduce the broadening of the experimental emission lineshapes for each complex or estimated from the variation of site energies sampled from the MD-Opt set.

4.3 Structural flexibility of open and closed Phycocyanins

Recent studies have suggested that the sequence diversity of α subunits plays a significant role in determining the chromophore conformations observed in X-ray crystal structures and, consequently, the spectral properties of cryptophyte complexes.[30] To investigate the impact of thermal effects and quaternary structure

on the light-harvesting properties of PC612 and PC645 complexes, we conducted extensive classical and QM/MM MD simulations. These simulations allowed us to analyze and compare the structural flexibility of the open PC612 and closed PC645 complexes.

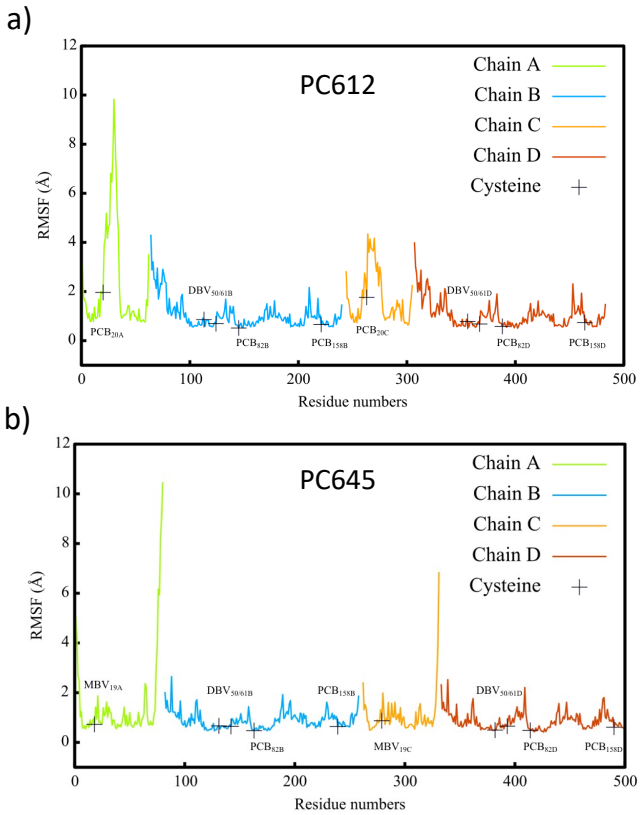


Figure 4.2. Root mean square fluctuation (RMSF) of atomic positions calculated along classical MD trajectories for a A/C and b B/D subunits of antenna complexes a) PC612 and b) PC645.

Figure 4.2 shows the root-mean-square fluctuations (RMSF) calculated along the classical MD trajectories for the α and β subunits of PC612 and PC645. The open

arrangement of the two αβ protomers in PC612 leads to higher flexibility in all subunits compared to PC645, which has a closed structure. Notably, in the β subunits B and D of PC612, there is increased flexibility in the N-terminal regions near PCB20A and PCB20C, where they interact with Tyr18 B and Tyr18 D through their propionic groups. In contrast, the C-terminal regions of the α subunits A and C in PC612 appear more rigid, despite the open quaternary arrangement. However, regions containing residues ~20–40 in chains A and C, starting at the anchoring points of PCB20A and PCB20C (Cys20 A and Cys20 C), show considerably more flexibility in PC612 compared to PC645.

In PC612, these regions form loops that mostly interact with PCB20A and PCB20C, partially shielding them from the solvent. This disposition is similar to that observed in PC645, which has the closed structure. However, in PC612, PCB20A and PCB20C are considerably more exposed to the solvent than MBV19A and MBV19C in PC645. Consequently, increased flexibility of PCBs in the open PC612 complex, likely due to higher solvent exposure, leads to more flexible protein loops. The RMSF values corresponding to the pigments in PC612 and PC645 (Table A1.1 in Appendix 1). confirm that increased fluctuations in the open protein PC612 result in more flexible pigments, particularly PCB20 bilins in α subunits, which display significantly larger fluctuations than the MBVs in PC645.

4.4 Spectral tuning of bilins by the protein scaffold

Next, we investigated how thermal fluctuations and protein flexibility influence the electronic excited states of the bilin pigments. Figure 3 presents the site energies (uncoupled transition energies) of the pigments computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations for PC612 and PC645. We compare the results obtained from classical MD geometries (MD set) or a single

optimization of the bilins in the protein scaffold based on the crystal structure (Crystal-OPT set)[29] with those obtained from multiple geometry optimizations of the pigments performed along classical MD trajectories (MD-OPT set) and short BOMD DFT-based simulations (MD-BOMD set). Figure A1.4-11 and Table A1.2-4 in Appendix 1 provide the averaged values and the complete distributions of site energies and transition dipole moments obtained from both sets.

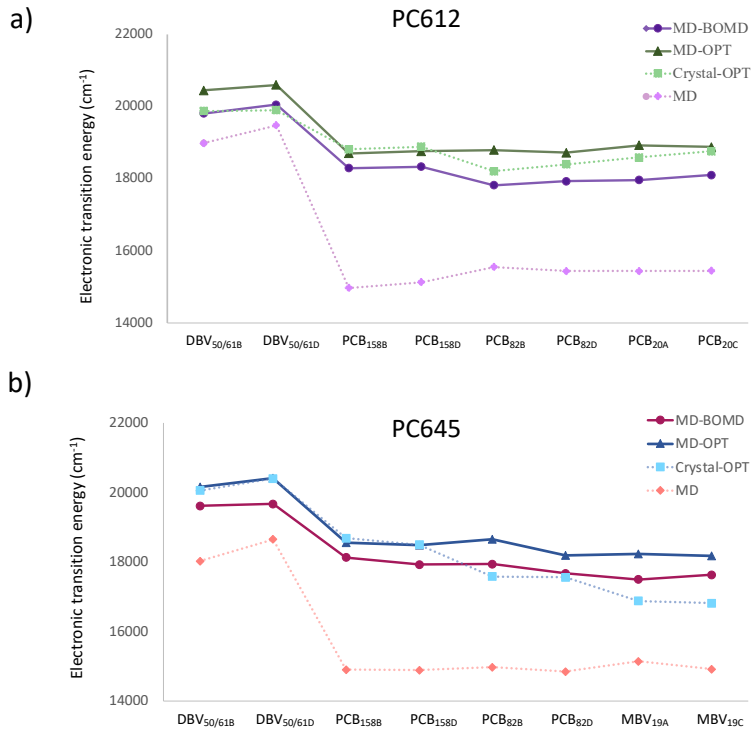


Figure 4.3. Electronic transition energies computed at the QM/MMPol TD-CAM-B3LYP/6-31G(d) level of theory for cryptophyte complexes using geometries obtained using the MD-BOMD and MD-OPT protocols compared to previous results based on classical MD or a single optimization of the bilins in the crystal structure:[187] a) PC612 and b) PC645.

In PC612, both the MD-BOMD and MD-OPT methodologies yielded similar site energies, closely matching those obtained using the Crystal-OPT strategy. This indicates that the energy ladder of PC612 is well explained by the crystal structure and the pigment-protein interactions, which constrain the bilin pigments to specific conformations. On the other hand, in PC645, the MD-BOMD and MD-OPT protocols significantly modify the picture provided by previous studies. While the energies of DBV and PCB158 pairs agree with previous values based on the crystal, the energetic position of PCB82 and MBV19 pairs is significantly modified, suggesting a crucial role of thermal effects. In terms of transition dipole moments, however, both MD-OPT and MD-BOMD sets provide results very similar to the Crystal-OPT ones in both complexes, as shown in Table A1.4 (see Appendix 1). Thus, despite the considerable distributions of dipoles obtained, which are comparable to those observed in site energies, as shown in Figure A1.11 (see Appendix 1) the averaged values turn out to be very close in all cases to those based on the crystal structure.

To explain the impact of thermal effects on the site energies of the PCB82 and MBV19 pairs in PC645, it is interesting to note that the comparison of the X-ray crystal structures of PC645 and a similar complex, PC630, showed that in the latter there is a hydrogen bond between ring D in PCB β82 chromophores and Gln6 in the αL and αS subunits, whereas in PC645 Gln6 is replaced by Leu5. The lack of this hydrogen bond is responsible for the red shift of the absorption spectra of PC645 compared to PC630.[30] Here, we show that beyond this analysis, thermal fluctuations of the protein environment are key to obtaining a balanced description of β82 PCB chromophores and to explaining why α19 MBVs in PC645 have site energies very close to the PCBs, despite their larger degree of conjugation. Indeed, bilin excitation energies are strongly controlled by the conformation of the tetrapyrrole backbone, and hydrogen bonds and steric clashes with the protein scaffold can twist the bonds connecting pyrrole rings, thus decreasing the conjugation of the π system, which increases the transition energies of the $\pi \rightarrow \pi^*$ states.

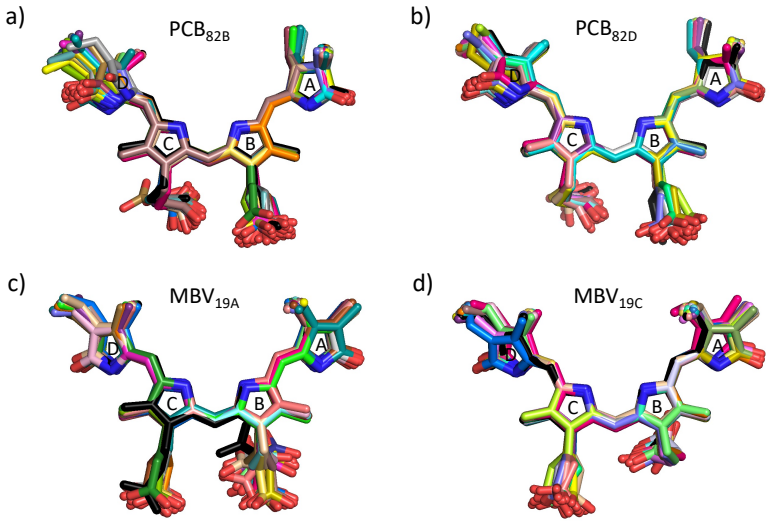


Figure 4.4 Alignment of PC645 pigment structures in the crystal structure (black) and optimized along the MD simulation (colors): a) PCB_{82B}, b) PCB_{82D}, c) MBV_{19A} and d) MBV_{19C}.

To understand the large impact of thermal effects on β82 PCB and α19 MBVs energy levels in PC645, Figure 4.4 shows an alignment of their structures as found in the crystal structure or optimized along the MD simulation following the MD-OPT protocol. Additionally, Tables 1 and 2 report the dihedral angles of the bonds connecting the bilin pyrrole rings (defined in Figure 4.5) obtained in the MD-BOMD calculations of PC612 and PC645 compared to the crystal references, as has been done in previous analyses of biliproteins either based on simulations or crystal structures.[30,178]

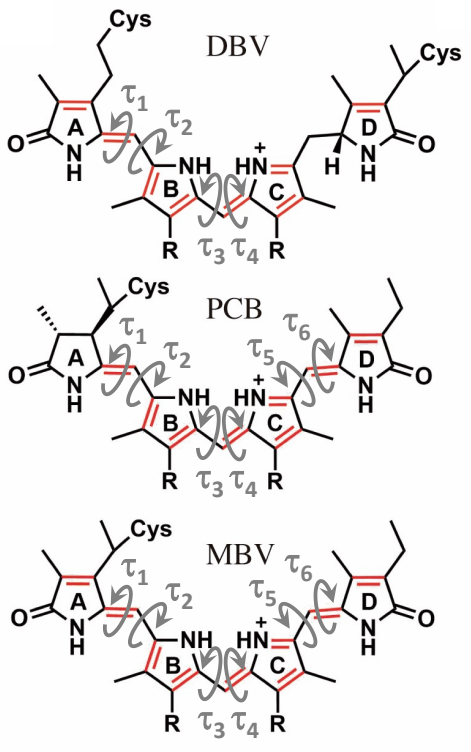


Figure 4.5. Structure of the linear tetrapyrrole arrangement of DBV, PCB and MBV bilin pigments indicating the key torsions that define the conformation of the conjugated region that characterizes the excited state.

Overall, Figure 4.4 indicates that there are no large conformational changes of the β82 PCB and α19 MBVs bilins in PC645 compared to the crystal reference, especially regarding the central B and C pyrrole rings. However, rings A and D and the propionic side chains display remarkable fluctuations compared to the crystal reference. Remarkable fluctuations are especially observed for ring D (shown on the left) in PCB82 bilins, which is confirmed by a standard deviation of ~16–19 degrees in the torsion τ5 reported in Table 4.2, which is almost twice the value observed in other bilins. As previously discussed, in PC645 the nitrogen atom of ring D in PCB β82 bilins interacts with Leu5, missing the two hydrogen bonds with Ser6 and Asp5 found in PC612, shown in Figure 4.6. In PC645, there

is indeed a hydrogen bond with Asp2, but its arrangement does not prevent considerable twists among the planes of C and D rings. Thus, in PC645, the lack of a hydrogen bond controlling the twist of ring D leads to larger fluctuations in PCB82 bilins that break the conjugation of the tetrapyrrole backbone, leading to a significant increase in the electronic transition energies of the $\pi \rightarrow \pi^*$ states, as shown in Figure 4.3. In MBVs, even stronger thermal effects are observed than in PCB82s. In this case, however, the fluctuations of rings A and D are comparable, and the strong impact of thermal effects seems to be linked to the fact that MBVs have a higher conjugation length that extends to ring A, so thermal fluctuations in the twist of that ring necessarily lead to a stronger impact than that found for DBVs or PCBs. Indeed, beyond the case of PCB82s in PC645 discussed above, Tables 4.1 and 4.2 show similar standard deviations for the torsions of MBVs and the other bilins in PC612 and PC645, with somewhat larger values for PCB20 and PCB158 pigments in PC612, which can be explained by the overall larger flexibility of the open structure of this complex.

Table 4.1. Averaged dihedral angles and their standard deviations computed from DFT-based BOMD simulations characterizing the twists among pyrrole units in the pigments of PC612. For comparison, the values corresponding to the crystal structure are provided in parentheses.

	t1	t2	t3	t4	t5	t6
$DBV_{50/61B}$	183 ± 9 (182)	32 ± 11 (20)	172 ± 10 (190)	182 ± 11 (168)	-	-
$DBV_{50/61D}$	178 ± 9 (178)	34 ± 11 (23)	181 ± 9 (186)	179 ± 9 (171)	-	-
PCB_{158B}	178 ± 6 (171)	-36 ± 14 (-20)	180 ± 7 (176)	181 ± 6 (178)	-32 ± 12 (-40)	178 ± 5 (183)
PCB_{158D}	177 ± 7 (171)	-39 ± 13 (-34)	184 ± 10 (179)	178 ± 8 (177)	-40 ± 11 (-43)	173 ± 9 (181)
PCB_{82B}	186 ± 7 (186)	-20 ± 11 (-24)	179 ± 9 (173)	187 ± 8 (189)	-33 ± 10 (-37)	170 ± 9 (175)
PCB_{82D}	186 ± 7 (189)	-18 ± 12 (-25)	177 ± 8 (174)	188 ± 9 (188)	-33 ± 11 (-38)	176 ± 8 (186)
PCB_{20A}	184 ± 6 (173)	11 ± 19 (-36)	184 ± 8 (190)	190 ± 10 (178)	-41 ± 11 (-33)	180 ± 9 (181)
PCB_{20C}	175 ± 8 (174)	-34 ± 15 (-37)	184 ± 10 (187)	179 ± 9 (179)	-40 ± 10 (-33)	176 ± 8 (186)

Table 4.2. Averaged dihedral angles and their standard deviations computed from DFT-based BOMD simulations characterizing the twists among pyrrole units in the pigments of PC645. For comparison, the values corresponding to the crystal structure are provided in parentheses.

	t_1	t_2	t_3	t_4	t_5	t_6
$DBV_{50/61B}$	190 ± 8	27 ± 11	190 ± 9	168 ± 8	-	-
	(195)	(20)	(187)	(174)		
$DBV_{50/61D}$	193 ± 8	33 ± 11	176 ± 9	187 ± 9	-	-
	(194)	(34)	(185)	(179)		
PCB_{158B}	176 ± 8	-37 ± 12 (-	182 ± 9	182 ± 8	-34 ± 11 (-	168 ± 9
	(174)	34)	(181)	(179)	27)	(168)
PCB_{158D}	171 ± 7	-36 ± 10 (-	187 ± 8	179 ± 8	-34 ± 11 (-	168 ± 9
	(171)	35)	(189)	(172)	25)	(168)
PCB_{82B}	183 ± 8	-24 ± 11 (-	179 ± 8	185 ± 9	27 ± 19 (-	189 ± 9
	(187)	29)	(181)	(188)	12)	(201)
PCB_{82D}	182 ± 8	-23 ± 11 (-	179 ± 8	186 ± 8	11 ± 16 (-	188 ± 9
	(191)	28)	(179)	(182)	6)	(203)
MBV_{19A}	182 ± 8	31 ± 11	181 ± 9	181 ± 8	-39 ± 10 (-	182 ± 3
	(184)	(29)	(185)	(183)	48)	(187)
MBV_{19C}	186 ± 8	29 ± 11	178 ± 9	184 ± 9	-43 ± 11 (-	175 ± 8
	(187)	(23)	(179)	(183)	45)	(176)

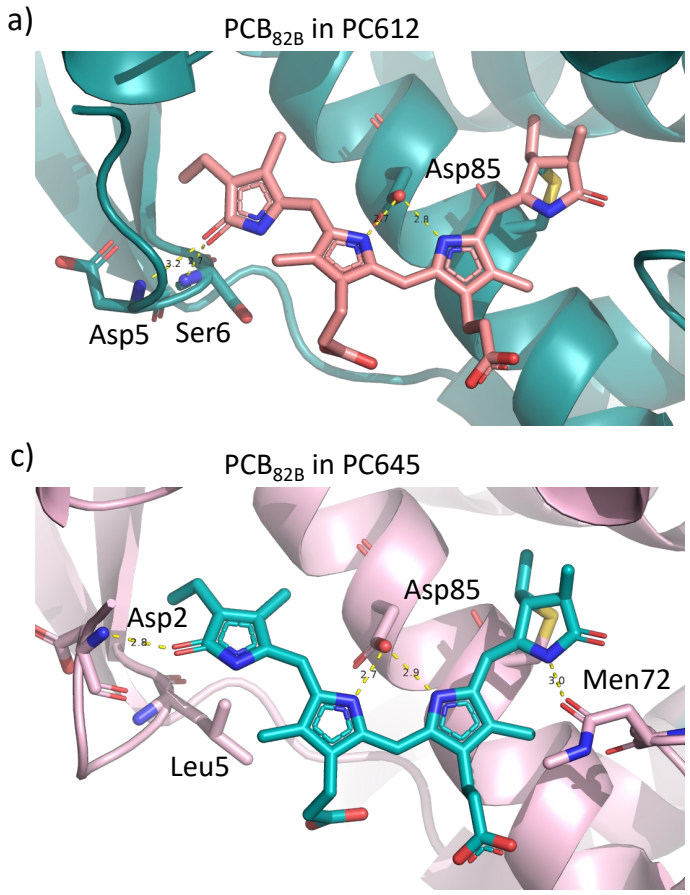


Figure 4.6. Hydrogen bonds established by PCB b82 bilins with the protein scaffold in 1) PC612 and 2) PC645. In PC612, the carbonyl group of ring D interacts with Asp5 and Ser6, stabilizing a considerable twist between rings C and D. In contrast, in PC645 the presence of Leu5 allows significantly larger fluctuations, which are not completely stabilized by the interaction with Asp2.

In addition to intramolecular fluctuations of the pigments, we analyzed how the quaternary structure of PC612 and PC645 and collective protein motions impact static disorder. Figure 4.7 shows an estimate of static disorder obtained as the standard deviation of site energies computed along the many MD-OPT optimized geometries. We also provide a similar analysis based on the energies averaged over

different BOMD QM/MM simulations. In the latter, short BOMD runs necessarily translate into overestimated disorder values, but the comparison between both protocols shows similar trends that suggest they both capture similar physics in the complexes. Indeed, both analyses show a similar disorder for DBV50/61 and PCB158 pairs in the two proteins, but opposite trends for PCB82 and the PCB/MBV α19/20 pairs. The increased flexibility and solvent exposure of PCB20 in the open structure of PC612 compared to MBV19 in PC645 seems to be the reason for the increased disorder. In contrast, in PC645, despite the closed quaternary structure leading to an overall more rigid protein, PCB82 pigments consistently display larger disorder compared to PC612. As discussed previously, PCB82 bilins in PC645 lack a key hydrogen bond with αL and αS subunits that lead to larger fluctuations in the twist of the bonds linking rings C and D, which explains such increased disorder. Our best estimates for disorder, shown in Figure 4.7, are larger than the value of 100 cm^{-1} we derived previously from a fit between experiment and simulations of absorption and emission spectra.[29] In the next section, we discuss the reliability of the site energies and the disorder values in describing the optical spectra of the complexes.

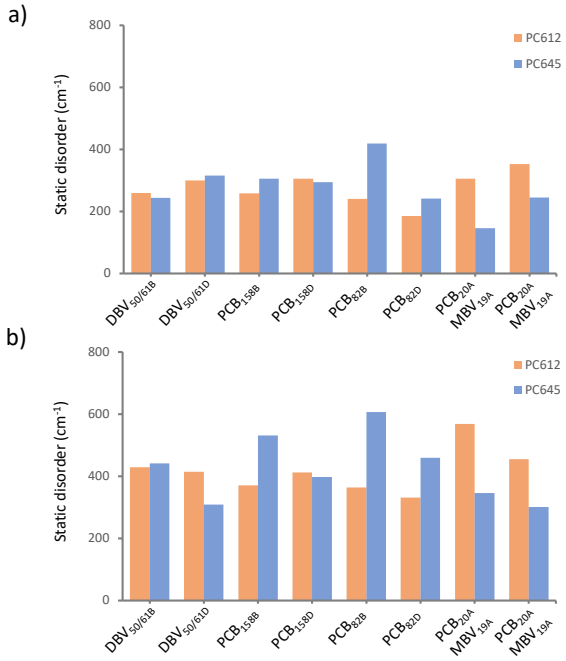


Figure 4.7. Static disorder (cm^{-1}) estimated from a) MD-OPT and b) MD-BOMD averaged transition energies computed along the classical MD trajectories of PC612 and PC645.

4.5 Absorption and circular dichroism spectra

To validate the excitonic Hamiltonian derived for PC612 and PC645 based on the MD-BOMD and MD-OPT site energies, we simulated OD and CD spectra of the complexes and compared them to experimental data. In Figure 4.8, we present our most accurate estimates, based on the dual classical/quantum MD strategy and the MD-BOMD protocol. Unlike previous studies, this approach accurately accounts for thermal effects by avoiding the geometry mismatch problem associated with using MD trajectories based on classical force fields.[29] Additionally, we adopted the full Cumulant Expansion (FCE) formalism,[174,175] which includes non-Markovian and non-secular effects neglected in modified Redfield theory. To

assess the impact of these effects, we also report spectra based on modified Redfield theory obtained using the EXAT code.[173]

The performance of the MD-BOMD set of site energies, combined with the FCE formalism, in describing the OD spectrum is remarkable. For both complexes, we achieve excellent agreement with experimental data, accurately reproducing the position of the two main peaks, their relative intensities, and the overall broadening in the spectra. This is a significant improvement over previous simulations reported on these systems,[29] where the absence of thermal effects led to a strong underestimation of MBV energies in PC645, too large differences among PCB158 and PCB82 bands, and overestimated DBV energies. For PC612, previous reports were more accurate, but the peaks were too close, resulting in an overall broad band.

In terms of CD spectra, the MD-BOMD simulations also correctly positioned the three experimentally observed bands, which are obscured in the OD spectra dominated by two bands. Interestingly, the MD-OPT protocol also produced accurate spectra very similar to those obtained using the MD-BOMD benchmark protocol, as shown in Figure A1.12 (see Appendix 1).Because the multiple geometry optimizations in this protocol do not fully account for thermal effects associated with intramolecular vibrations of the pigments, we conclude that accounting for the collection of protein conformations is crucial, which is not achieved in analyses limited to the crystal structure.

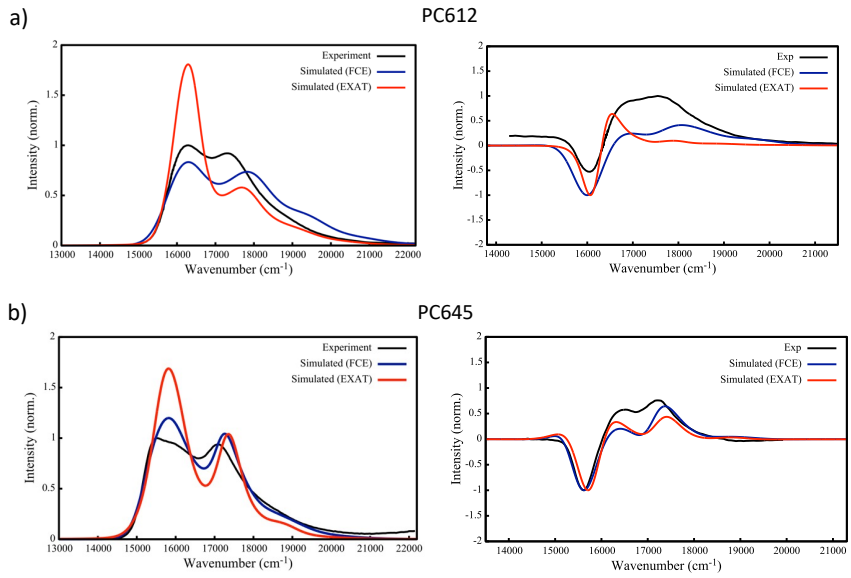


Figure 4.8. Experimental[188–190] and simulated absorption (left) and circular dichroism (right) spectra of a) PC612 and b) PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-BOMD geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 1600 cm^{-1} (EXAT PC612), 1000 cm^{-1} (FCE PC612), 1700 cm^{-1} (EXAT PC645) and 1480 cm^{-1} (FCE PC645). Static disorder was modeled by applying a common $\sigma = 100$ cm^{-1}.

As extensively discussed in the previous section, thermal effects are crucial for accurately describing the energies of PCB82 and MBV bilins in PC645. Additionally, accounting for non-Markovian and non-secular effects through the FCE formalism is important for correctly describing the relative intensities of the two main absorption peaks in OD spectra. Figure 8 shows that EXAT simulations overestimate the intensity of the low-energy band, while the FCE formalism provides a more accurate representation. This improvement is also observed in the relative intensities of CD bands, especially in PC612. For PC612, the FCE and

EXAT methods produce similar OD lineshapes in the blue edge of the spectrum, but for PC645, the FCE formalism offers a more accurate lineshape. This can be attributed to a better description of the coupling between vibrations and the exciton states delocalized on the central DBV50/61B—DBV50/61D dimer in PC645, which has a strong electronic coupling of 238 cm^{-1}.[29] In the open quaternary structure of PC612, this coupling is reduced to 16 cm^{-1}, leading to mostly localized states. This distinction is evident in Figure A1.13 (see Appendix 1), where we compare the OD spectra to simulations in which all couplings are zeroed. While neglecting excitonic couplings does not significantly alter the PC612 spectra, which are characterized by localized states, the impact on PC645 is substantial, particularly in the high-energy band characterized by the strongly coupled DBV50/61B—DBV50/61D dimer.

The theoretical estimates of static disorder shown in Figure 4.7 are larger than the 100 cm^{-1} value used in the simulations reported in Figure 4.8, previously fitted by comparison to absorption and emission experimental spectra.[29] In Figure A1.14-16 (see Appendix 1), we present OD, CD, and FLU spectra obtained using our theoretical estimates of disorder based on MD-OPT simulations. In all cases, the OD spectra display minor changes. For CD and FLU spectra, PC612 shows reasonable broadening. However, the CD and FLU spectra of PC645 are slightly too broad, indicating that our methodology still has limitations in estimating this important parameter, likely due to the simulations being based on MDs spanning a limited μs timescale.

CHAPTER 5. COMPUTATIONAL SCREENING OF PUTATIVE I2 RECEPTORS

5.1 Background

The growing prevalence of neurodegenerative diseases, particularly Alzheimer's Disease (AD) and Parkinson's Disease (PD), has increased efforts to develop effective treatments. As these diseases continue to rise with the aging global population, the need for innovative therapeutic strategies becomes even more urgent. Among the potential therapeutic targets, Imidazoline receptors (IRs) have shown considerable promise, especially the type 2 receptor (I2-IR). Recent studies suggest that I2 ligands may help reduce both cognitive and non-cognitive symptoms in AD, showing their neuroprotective effects. However, among the three types of Irs-type 1 (I1-IR), type 2 (I2-IR), and type 3 (I3-IR), I2-IRs remain the least understood. Despite the growing interest in their potential, a lack of structural understanding continues to limit the development of targeted drugs that can fully take advantage of their therapeutic benefits.[40,43,47,191]

I2 receptors are nonadrenergic binding sites initially identified for their high affinity for [3H]-idazoxan, distributed across tissues such as the brain, kidneys, liver, colon, vascular smooth muscle, astrocytes, glial cells, platelets, and pancreatic cells. Primarily located on the outer mitochondrial membrane, they are involved in energy regulation and show complex associations with multiple proteins. Four distinct proteins (30, 45, 66, and 85 kDa) in rat and rabbit brain tissues bind I2 ligands like idazoxan and 2-BFI. While some binding sites are associated with monoamine oxidase B (MAO-B), significant binding persists in MAO knockout mice, suggesting other proteins are involved. [50,59,60,192] Brain creatine kinase, crucial for energy metabolism, also binds I2 ligands like 2-BFI, hinting at roles in brain energy homeostasis. Additionally, the I2 receptor ligand BU99006 upregulates Nischarin expression in the mouse hippocampus, indicating a compensatory mechanism during I2 receptor inhibition.[193] However, despite significant experimental efforts, the identities, binding sites, and functions of other

associated proteins remain unclear. This highlights the regulatory complexity of I2 receptors and underscores the critical need for their identification and characterization.

In collaboration with the groups of Prof. Carmen Escolano, Prof. Mercè Pallàs and Dr. Christian Griñán Ferré, we aim to characterize the binding sites of the canonical I2 ligand 2-BFI to their putative targets, enabling the derivation of meaningful structure-activity relationships to guide the identification and optimization of novel I2 ligands. In this thesis, we have devised a computational protocol to screen putative targets identified in previous proteomics analysis conducted in a senesce-accelerated mouse model to identify suitable binding pockets able to accommodate 2-BFI. This multidisciplinary approach seeks to lay the groundwork for the future development of innovative treatments for neurodegenerative diseases targeting I2 receptors.

5.2 Computational details

The three-dimensional structures of seventeen proteins were built from the AlphaFold protein structure database.[73] To match the proteomics experiments we used the *Mus musculus* proteins. When required, the position of co-factors absent in the AlphaFold predictions, such as the Flavin-Adenine-dinucleotide (FAD) from Monoamine oxidase B (MAOB), were superimposed from crystallographic structures of homologous proteins. Specifically, MAOB was built from the AlphaFold structure (AF-Q8BW75-F1) with coordinates for FAD and crystallographic water molecules obtained from PDB 2XFN;[194] the Acidic Residue Methyltransferase 1 (AF-A6H630-F1) with Mg^{2+} and water molecules from PDB 6UQM.[195] The Creatine kinase B-type (AF-Q04447-F1) was modeled using with dummy Mg^{2+}, ADP,[196] and water molecules from PDB 3DRB for the closed form, and PDB 3DRE[197] is used for the open form. To obtain the open form of the *M. musculus* Creatine kinase-B, we mutated the structure of 3DRE to match the mouse

sequence using the PyMol mutagenesis wizard tool. Finally, the Hydroxymethylglutaryl-CoA lyase (AF-P38060-F1) incorporated Mg^{2+} and water molecules from PDB 2CW6.[198] The remaining proteins were taken from the AlphaFold protein structure database without further modification. Specifically, the Alpha-2a adrenergic receptor (AF-Q01338-F1), Nischarin (AF-Q80TM9-F1), the RUN and FYVE domain-containing protein 2 (AF-Q8WXA3-F1-v4), Peroxiredoxin-6 (AF-O08709-F1), 40S ribosomal protein S9 (AF-Q6ZWN5-F1), Splicing factor U2AF 65 kDa (AF-P26369-F1), S-formylglutathione hydrolase (AF-Q9R0P3-F1), Heterogeneous nuclear ribonucleoprotein U (AF-Q8VEK3-F1), WD repeat-containing protein 47 (AF-Q8CGF6-F1), Rho GTPase-activating protein 44 (AF-Q5SSM3-F1), La-related protein 4B (AF-Q6A0A2-F1), Ribonucleoprotein G-like (AF-Q91VM5-F1) and Receptor expression-enhancing protein 1 (AF-Q8BGH4-F1). Protonation states were evaluated with PROPKA to ensure proper ionization at physiological pH, and the disordered regions for which the AF prediction was deemed of low confidence were removed and appropriate capping groups added.

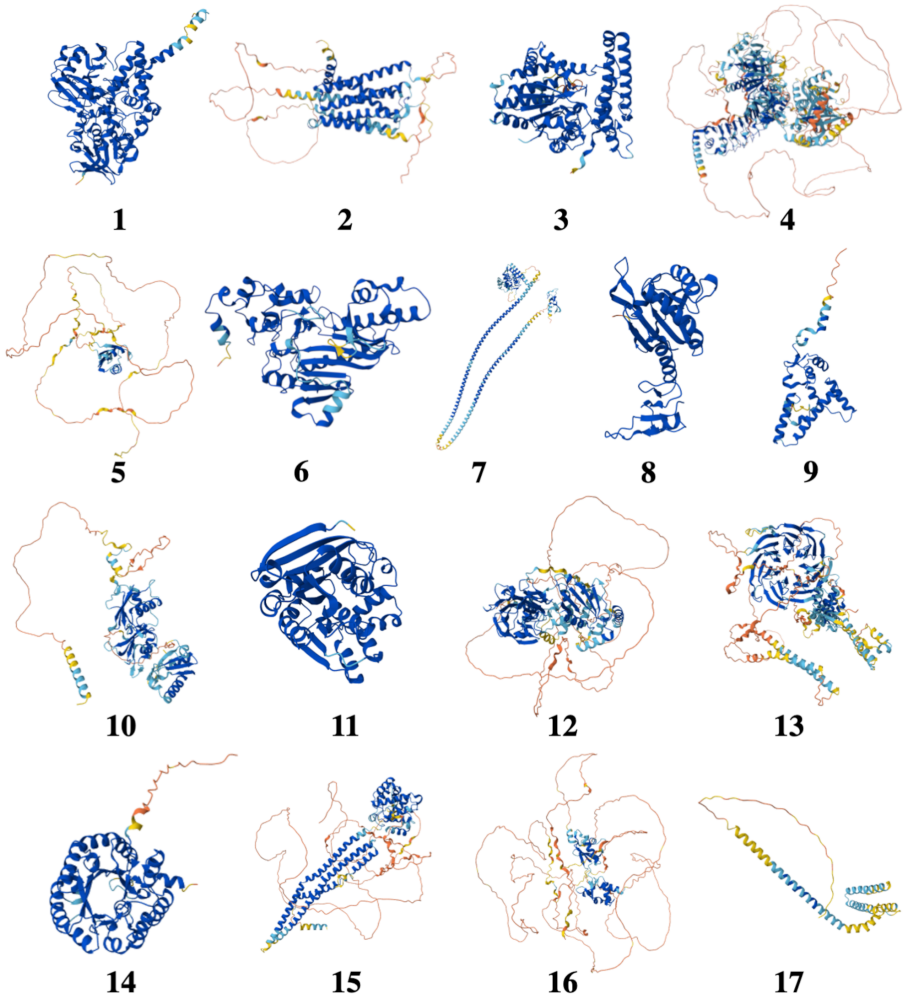


Figure 5.1 AlphaFold structures of the proteins computationally screened for binding pockets of 2-BFI. Further details are provided in the main text.

To identify 2-BFI I2 ligand binding site, we developed a systematic protocol that combined different common methodologies in CADD (Figure 5.2). In brief, the protocol starts with MDmix simulations in organic solvent/water mixtures to identify high-affinity interaction hotspots, with retained solvent probes indicating potential binding sites. Based on the structure of 2-BFI co-crystallized with

MAOB, we identified the physicochemical properties and relative spatial orientation of the hotspots that match the 2-BFI binding site (Figure 5.3). In the next step, we filtered out the hotspots that did not match the pattern obtained for 2-BFI and used the remaining hotspots as pharmacophoric constraints for docking calculations using rDock.[138] Finally, we evaluated ligand stability on each identified binding site using MD simulations. We validated this protocol by recovering the crystallographic binding site of 2-BFI in MAOB and subsequently used it to explore potential binding sites for I2 ligands in the remaining sixteen proteins. Further details into each stage of the protocol are described in the subsequent sections.

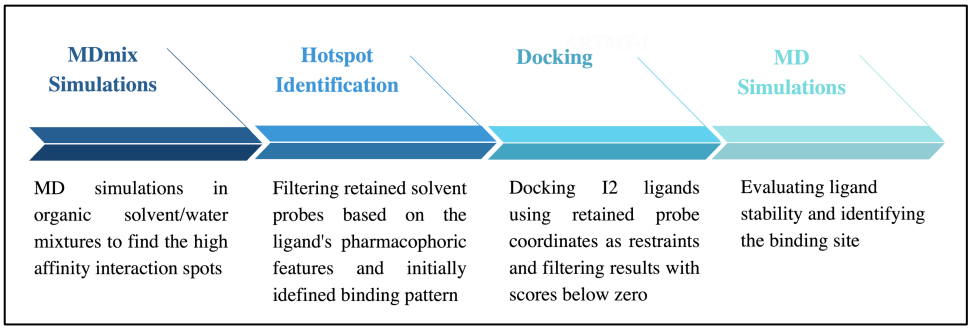


Figure 5.2 Workflow for identifying 2-BFI binding sites, including MDmix simulations, hotspot filtering, ligand docking, and MD simulations to confirm stability and interactions.

5.2.1 MDmix simulations using mixtures of water and organic solvents

Three different mixtures of water and organic solvents were used to identify binding hotspots for 2-BFI: a mixture of 20% v/v ethanol-water was used to identify buried H-bond donor/acceptor hotspots, while a mixture of 20% v/v isopropanol-water was used to identify potential H-bond interaction points in more solvent exposed regions. Finally, a mixture of 1% v/v pyridine in water was employed to mimic the potential interactions of the benzofuranyl aromatic moiety (5.3). Each protein was solvated in a truncated octahedral box with each solvent

mixture, and the systems were minimized with Cartesian restraints on heavy atoms using a force constant of 0.01 kcal/mol·Å^2. Equilibration was performed stepwise, gradually heating from 100 K to 300 K over four 200 ps phases in the NVT ensemble. Once at 300 K, the system underwent an additional 1 ns equilibration under the NPT ensemble at 1 bar, maintained by the Berendsen barostat with a coupling constant of 2 ps. The production phase involved three independent 100 ns simulations for each protein-solvent combination at 300 K in the NVT ensemble, with heavy atoms restrained using a 0.01 kcal/mol·Å^2 force constant. After the simulations, all replicas were superimposed onto a reference structure (based on the protein's backbone atoms in crystallographic coordinates). A 0.5 Å grid was constructed around the protein for each solvent probe, allowing observed solvent densities to be compared with expected bulk densities. Binding free energy (ΔG_{bind}) values for each grid element were calculated using the inverse Boltzmann relationship. Regions with the most negative ΔG_{bind} values, representing highly favorable interactions, were identified as binding hotspots. These hotspots reveal regions where solvent molecules preferentially interact with the protein surface.

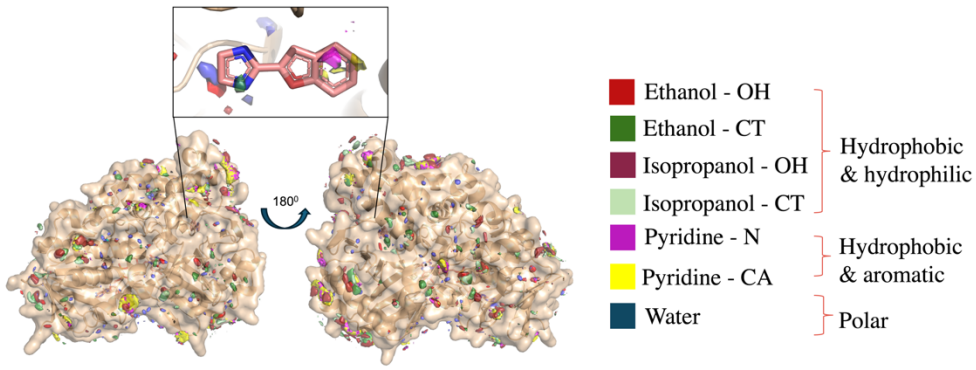


Figure 5.3 The MAO-B solvent densities from MDmix simulations are color-coded: red/dark red for OH, green for ethyl, light green for isopropyl, purple for nitrogen, yellow for the pyridine ring, and blue for water. The probes align with the pharmacophoric features of the superimposed 2-BFI ligand at its original binding site in the MAO-B X-ray structure.

5.2.2 Finding the pharmacophoric hotspots

MDmix simulations were used to identify solvent probes occupying the 2-BFI binding site in MAOB according to the crystallographic structure with PDB ID 2XFN (**Figure 5.4**). We hypothesized that the spatial arragement of these solvent hotspots could be used as pharmacophoric features to highligh regions capable of accommodating 2-BFI. After superimposing the hotspot location with the structure of 2-BFI, we identified an aromatic hotspot (from the pyridine probe) and a H-bond hotspot (from the ethanol and isopropanol probes) in the same plane and with a separation of 7.6 Å. A third hotspot for a water molecule was identified in the same plane, at a distance of 2.7 Å from the H-bond hotspot and 6.6 Å from the aromatic hotspot. This spatial arrangement of hotspots was used to filter out all the putative binding sites for the solvent probes that did not display a similar arrangement (with a tolerance of 1 Å in the distances), thus reducing the number of possible docking sites.

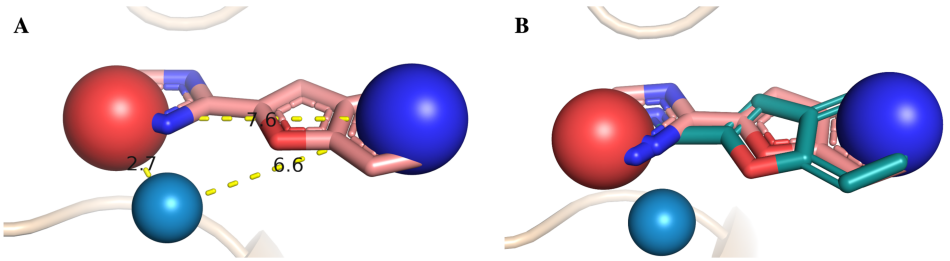


Figure 5.4 (A) The hotspot is located in the entrance cavity of MAO-B, with 2-BFI (pink) superimposed from the X-ray structure. Solvent probes are color-coded: navy blue for pyridine (aromatic interactions), red for ethanol/isopropanol OH (hydrogen bond donor/acceptor), and light blue for water. These coordinates were used to establish aromatic and hydrogen bond restraints during docking. (B) The docked 2-BFI (greenish blue) represents the best-scored structure, showing close alignment with the original X-ray structure of 2-BFI (pink) in the entrance cavity.

5.2.3 Molecular docking

Docking calculations were performed using the rDock program based on hotspots identified and filtered as described above. The docking cavity was defined using the two-sphere method, in which the binding pocket center is first set at the center of the three hotspots with a radius of 15.0 Å to confine the docking search space. A larger probe of 6.0 Å excludes shallow, solvent exposed areas, while a smaller probe of 1.5 Å serves to define buried cavities. The method grids the target region, excludes overlaps with the protein structure, and identifies cavities of enough volume to accommodate drug molecules. Ligands prepared in SDF format using the ChemAxon software suite[199] were used as input for docking, with receptor flexibility set to 3.0 Å. Docking involved 50 runs of the Genetic Algorithm per ligand and per cavity. The poses with favorable interaction score (negative according to rDock's scoring functions) were retained for further analysis, while those not meeting this criteria were discarded.

5.2.4 MD simulations

MD simulations were conducted to assess the structural stability of the identified binding modes. All simulations used the pmemd.cuda module from the Amber package, with system preparation handled by the tleap tool from AmberTools20. [181] The ff14SB[179] force field was applied to the protein and the gaff2 forcefield to the ligand, using the TIP3P water model. The system was solvated in a truncated octahedron water box extending 12 Å beyond the protein and neutralized with Na^+ and Cl^- ions. A pKa of 8.3 was calculated for 2-BFI, hence the protonated form was used in all calculations. After optimizing the structure of 2-BFI at the B3LYP/6-31+G(d) level of theory, partial charges were obtained with the RESP method at the HF/6-31G(d) level of theory. All runs were performed with the

Amber20[182] and Gaussian 16. The off-center dummy model was employed for magnesium cations.[200]

Prior to the production runs, minimization was performed in two stages: first, the solvent and ions were minimized with a harmonic restraint of 10 kcal/mol Å^2 on the protein-ligand complex using 500 steps of steepest descent, followed by 2000 steps of conjugate gradient minimization without restraints. Equilibration was carried out in five heating steps from 0 K to 300 K in the NVT ensemble, each lasting 50 ps, followed by 250 ps in the NPT ensemble at 1 bar. For production, MD simulations were run at 300 K in the NVT ensemble for 50 ns per replica using the H mass repartitioning method with a 4 fs timestep. Long-range electrostatics were computed with the Particle Mesh Ewald method, and van der Waals interactions were truncated at 8 Å. The SHAKE algorithm constrained bonds involving hydrogens, and the Langevin thermostat maintained a collision frequency of 1 ps^{-1}. Production runs involved three independent 50 ns MD simulations for each docked 2-BFI molecule with negative docking scores, totaling 150 ns per docked position. The root mean square deviation (RMSD) to the initial pose was used to monitor the structural stability of each binding pose.

5.2.5 Protocol validation on the Monoamine Oxidase-B I2 ligand binding site

We initially sought to validate the developed protocol by recovering the crystallographic binding pose of 2-BFI in MAOB. After the MDmix simulations and hotspots filtering, we identified 66 suitable binding sites for 2-BFI based on the spatial positioning of solvent hotspots. Docking calculations yielded 15 binding poses with scores below zero, which were chosen for further analysis (Figure 5.5). MD simulations were then conducted for these poses, with three replicas run for 50 ns each, 2.25 μs in total. The replicas were aligned to the initial protein backbone structure, and RMSD values of the ligand were calculated for the last 10 ns of the simulation using the initial position as a reference, along with

standard error of the mean (SEM). Two binding positions remained below a 5 Å threshold (Figure 5.6). The first binding site (11), D1, corresponds to the entrance cavity where 2-BFI was originally located in the crystal structure, with an RMSD value of 1.66 ± 0.04 Å. The second binding site (8), D2, represents a previously unknown location, with an RMSD value of 4.6 ± 0.04 Å. Notably, one out of three replicas from D1 and two out of three replicas from D2 exhibited movement toward the aromatic cage formed by F103, W119, and F168. To confirm this transition, starting from both binding sites, additional set of simulations were conducted, which demonstrated a consistent pattern, as shown in Figure 5.7. While D1 and D2 cavities were selected for the initial ligand placement, the aromatic cage dynamically rearranged itself during the simulation to accommodate the 2-BFI ligand, creating a binding configuration that could not be captured through docking with the initial protein structure.

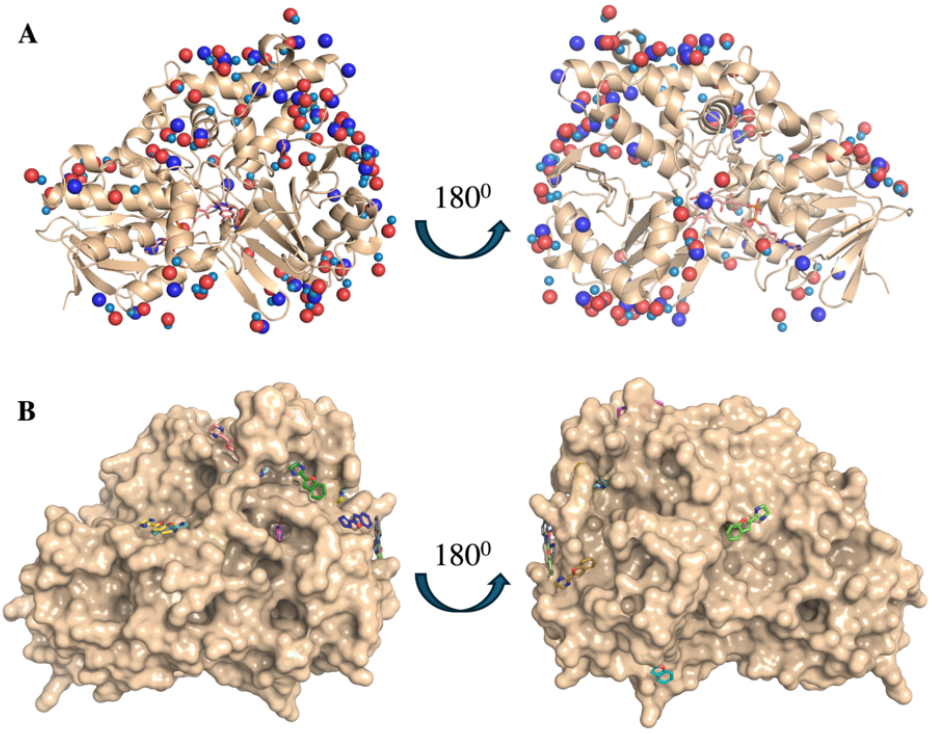


Figure 5.5 (A) Hotspot analysis from MDmix simulations shows pyridine (N-navy blue), ethanol/isopropanol (OH-red), and water (light blue) probes. (B) Docking results illustrate 2-BFI molecules with best scores positioned using these hotspots as pharmacophoric restraints, showing only poses with scores below zero.

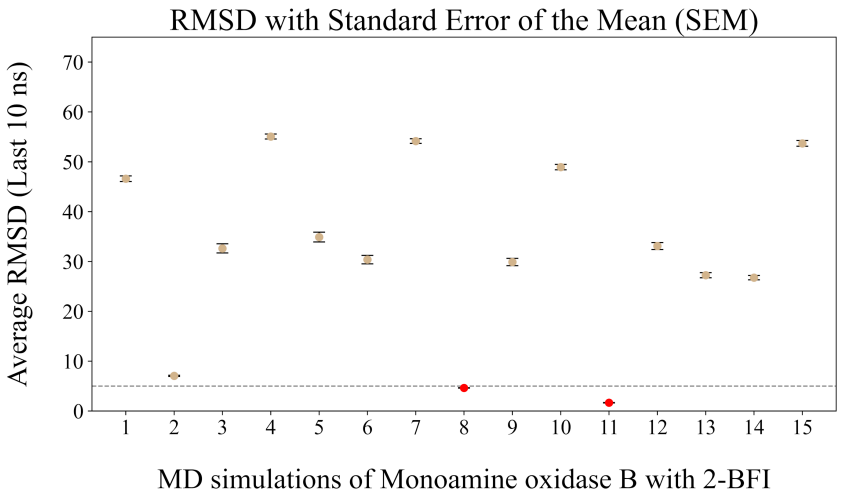


Figure 5.6 Average RMSD values of the 2-BFI ligand relative to its initial position, calculated after aligning the protein backbone over the final 10 ns of the MD simulations. The graphs represent data from three replicas and include the standard error of the mean, with red data points indicating positions below the 5 Å threshold.

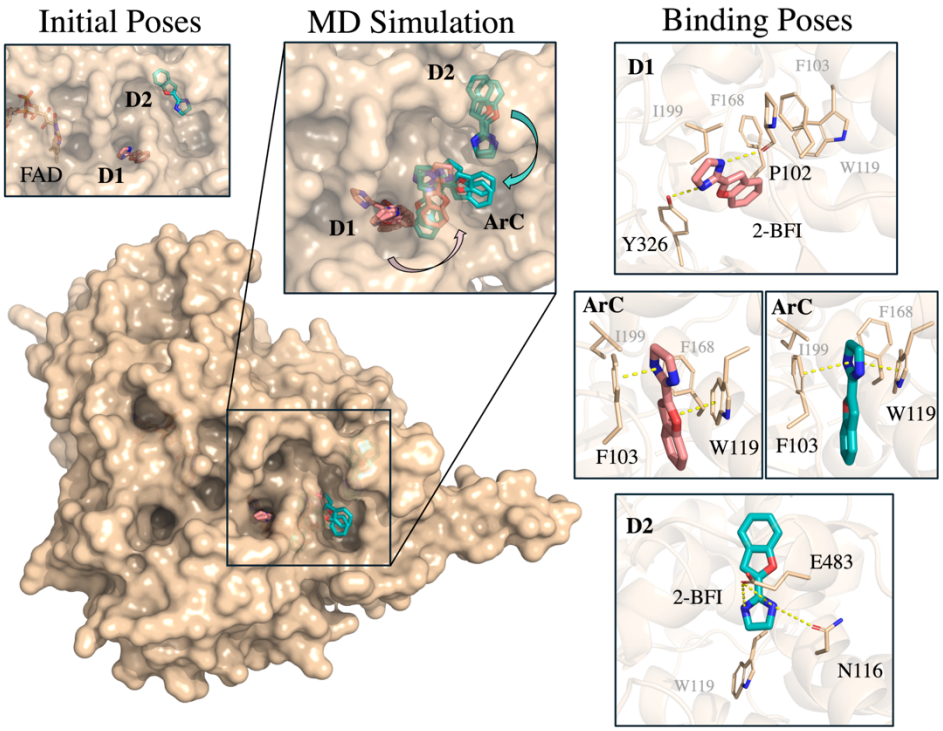


Figure 5.7 The 2-BFI binding site on MAO-B, highlighting two distinct binding modes. Binding Site 1 (D1) is represented by pink molecules, Binding Site 2 (D2) by blue molecules, and binding poses of the ligands in the aromatic cage (ArC). Binding poses illustrate the final positions of the ligands in the trajectory.

In the D1 cavity, 2-BFI establishes hydrogen bonds with Y326 and the backbone oxygen atom of P102 via the nitrogens in its imidazoline ring, with averaged bond distances of 3.05 ± 0.34 Å and 2.88 ± 0.24 Å, respectively. In the D2 cavity, 2-BFI forms a salt bridge with E483 and a hydrogen bond with N116, both with bond distances of 4.1 ± 1.19 Å and 3.8 ± 0.8 Å, respectively. Within the aromatic cavity, the ligand alternates between a cation-π interaction with F103 and W119 through its imidazoline ring, and a π-π interaction via its aromatic ring. In the aromatic cage, 2-BFI is 4.05 ± 0.58 Å and 3.7 ± 0.27 Å from F103 and W119 when moving from D1, and 3.85 ±0.3 Å and 3.8 ± 0.25 Å when moving from D2.

In summary, we were able to validate the protocol by successfully recovering the crystallographic binding pose of 2-BFI in MAO-B and, simultaneously, identified new putative interaction sites. We hypothesize that these sites may be able to allosterically modulate the enzymatic activity of MAO-B, which is further investigated in Chapter 6.

5.3 Virtual screening of putative I2-IR receptors

In addition to MAO-B, the analysis of proteomics results previously conducted by Dr. Griñan identified sixteen possible I2-IR receptors. The results of the application of our protocol to these potential proteins of interest are summarized in Table 5.1.

We identified highly-stable binding sites in six out of the sixteen proteins, hinting for high-affinity binding sites for 2-BFI. The Alpha 2a adrenergic receptor and Nischarin both showed two possible 2-BFI binding sites each, while we identified one possible binding site for 2-BFI in Creatine Kinase-B, S-formylglutathione hydrolase, Heterogeneous nuclear ribonucleoprotein U and Hydroxymethylglutaryl-CoA lyase. In the remaining proteins, the protocol did not identify possible high-affinity sites for 2-BFI, and thus we hypothesized that their identification in the proteomics assays maybe due to downstream effects of the signalling cascade after treatment with 2-BFI or other indirect effects of the ligand. The identified hotspots and docking interactions of the 2-BFI molecule with proteins that did not show strong binding, including Acidic Residue Methyltransferase 1, Ribonucleoprotein G-like, RUN and FYVE domain-containing protein 2, Peroxiredoxin-6, 0S ribosomal protein S9, Splicing factor U2AF 65 kDa subunit, WD repeat-containing protein 47, Rho GTPase-activating protein 44, and La-related protein 4B, are detailed in Appendix 2.

Table 5.1: Summary of I2-IR receptor screening.

	Protein Name	Protein residues	Docking sites	Binding poses and number of replicas	Identified binding sites
1	Monoamino oxidase B protein	449	66	15 × 3	**2**
2	Alpha 2a adrenergic receptor	302	38	13 × 3	**1**
3	Acidic Residue Methyltransferase 1	439	41	17 × 3	0
4	Nischarin	1272	223	96 × 3	**2**
5	Ribonucleoprotein G-like	90	44	10 × 3	0
6	Creatine Kinase-B (open and closed)	377	61	37 × 3	**1**
7	RUN and FYVE domain-containing protein 2	233	41	14 × 3	0
8	Peroxiredoxin-6	224	59	20 × 3	0
9	0S ribosomal protein S9	160	32	19 × 3	0
10	Splicing factor U2AF 65 kDa subunit	350	56	26 × 3	0
11	S-formylglutathione hydrolase	282	9	2 × 3	**1**
12	Heterogeneous nuclear ribonucleoprote-in U	429	90	25 × 3	**1**
13	WD repeat-containing protein 47	643	36	23 × 3	0
14	Hydroxymethylglutaryl-CoA lyase, mitocondria	303	63	29 × 3	**1**
15	Rho GTPase-activating protein 44	220	62	20 × 3	0
16	La-related protein 4B	173	39	17 × 3	0
17	Receptor expression-enhancing protein 1	107	21	11 × 3	0

5.3.1 Identification of 2-BFI binding sites in Nischarin (Uniprot: Q80TM9)

Nischarin is a mice multifunctional protein involved in cellular signaling, receptor regulation, and cytoskeletal organization. It plays critical roles in neuronal migration, synaptic plasticity, neuroprotection, cell migration, apoptosis, tumor suppression, and glucose metabolism, while supporting neuronal function, especially under stress.[193,201–205] Initially identified as a binding protein for integrin α5β1 and in humans as an I1-imidazoline receptor expressed in neurons and astrocytes,[203] Nischarin is linked to conditions such as cancer, opioid dependence, and neurodegenerative diseases. The human homolog, imidazoline receptor antisera-selected protein (IRAS), shares 80% homology with rodent Nischarin.[202,206]

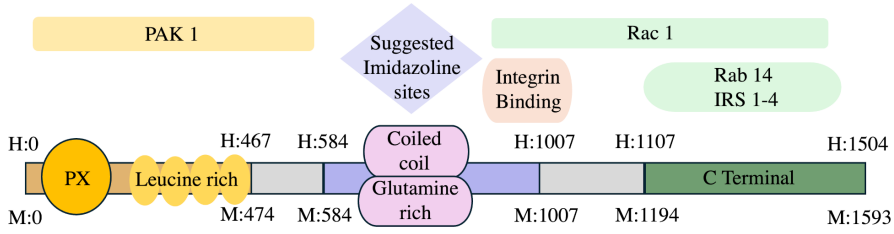


Figure 5.8 Important domains and interactions in human and mouse Nischarin.

Nischarin is composed of three structured regions that contain different functional domains: The N-terminal region (1-474) contains the PX domain (11–121) and a Leucine rich sequence and it is essential for endosomal targeting and binding to PAK1. The second structured region (584-1007) contains a Glutamin rich region that organizes as a coiled-coil domain (634–695) and also enables protein-protein interactions. The C-terminal domain (1120–1504) mediates intracellular transport and interacts with proteins like Rab14, Rac1 GTPases, and insulin receptor substrates (IRS 1-4) to activate ERK and regulate metabolic and survival pathways (Figure 5.8).[202,207–209] Reduced expression of IRAS lowers I1R density and impairs

ERK activation, highlighting its importance in cellular signaling.[201] The 966–1020 region has been identified to be the primary binding site for I1-IR ligands.[210]Also, docking studies with Autodock Vina Software suggested key binding sites for the I1 ligand agmatine in the C-terminal domain of Nischarin, including R1260, A1353, D1384, and E1385, which form significant hydrogen bonds and electrostatic interactions.[204,205]

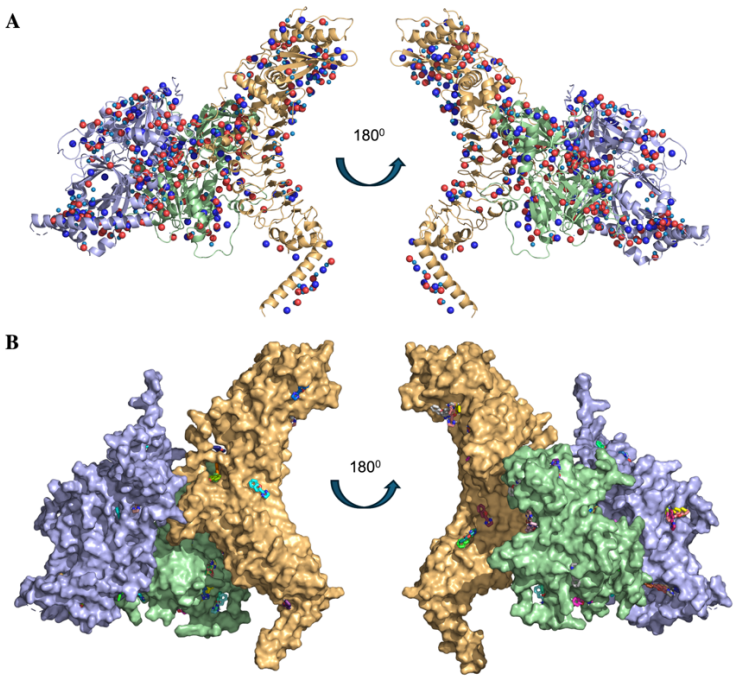


Figure 5.9 (A) Analysis of MDmix simulations for Nischarin domains highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

Due to the size of the protein, we independently applied the same protocol to each domain. The MDmix analysis identified 89 high-affinity interaction spots in the N-terminal region, 68 hotspots in the secondary region (coiled-coil), and 67 hotspots in the C-terminal region. Subsequent docking analysis revealed 45 putative binding sites in the N-terminal (PX) domain (residues 1–474), 27 binding

sites in the secondary structured domain (residues 584–1007), and 24 binding sites in the C-terminal region (residues 1194–1593) with docking scores below zero (Figure 5.9).

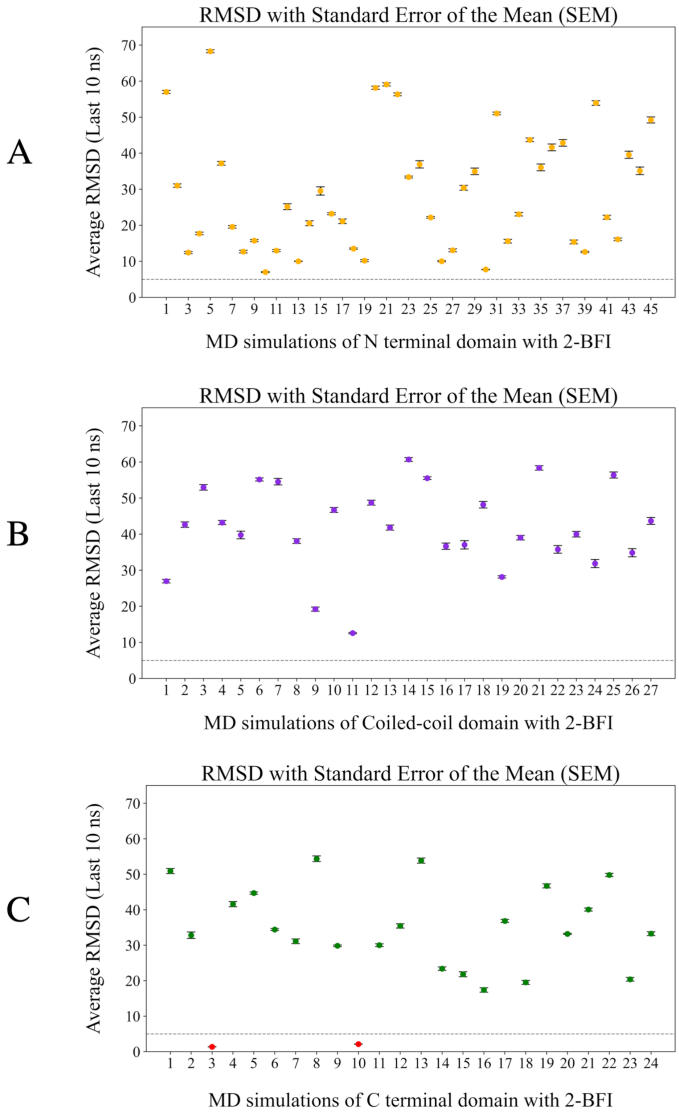


Figure 5.10 Average RMSD values of 2-BFI relative to its initial position during the final 10 ns of MD simulations. Graphs show three replicas with SEM, highlighting red data points below the 5 Å threshold for (A) N-terminal PX domain, (B) Coiled-coil domain, and (C) C-terminal domain of Nischarin.

Further filtering was performed using MD simulations, with a total simulation time of 6.75 μs for the N-terminal domain, 4 μs for the coiled-coil region, and 3.6 μs for the C-terminal domain. Only two binding sites in the C-terminal region exhibited stable binding modes. These stable modes correspond to binding number 3 (RMSD 2.14 $\pm$ 0.04 Å) and binding number 10 (1.37 $\pm$ 0.02 Å) on the C-terminal domain (Figure 5.10).

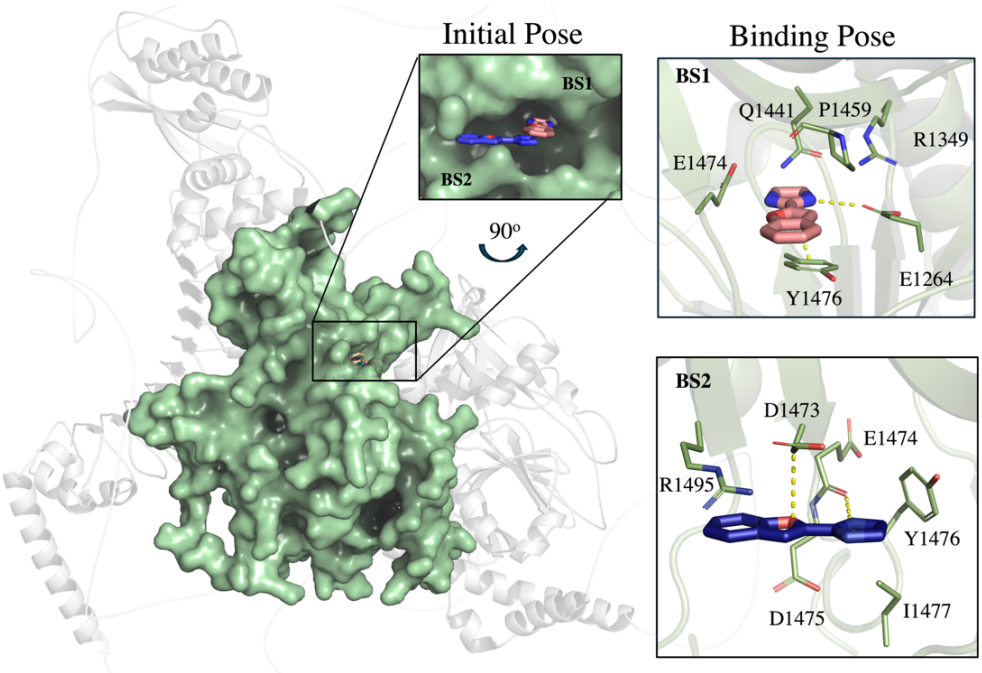


Figure 5.11 The binding sites of 2-BFI on the C-terminal domain of Nischarin are shown. The 2-BFI molecules, depicted in blue (molecule 3) at binding site 1 and pink (molecule 10) at binding site 2, represent their respective binding modes.

We further analyzed the two binding poses. The binding (10) in Figure 5.10, referred to as Binding Site 1 (BS1), is defined by residues Q1441, R1349, E1264, P1459, Y1476 and E1474. E1264 forms an intermintent salt bridge with the nitrogen atom of the imidazoline ring of 2-BFI, with an average distance of 3.94

±1.40 Å. Additionally, Y1476 engages in a cation–π interaction with 2-BFI, with an average distance of 4.1 Å ± 0.35 Å. The molecule is also near P1459 and E1474. The binding (3) in Figure 5.10, referred to as Binding Site 2 (BS2), shows that the primary interaction in BS2 involves E1474, which forms a hydrogen bond with the nitrogen atom of the imidazoline ring of 2-BFI through its backbone carbonyl oxygen, with an average distance of 2.86 ± 0.14 Å. The cavity is surrounded by residues D1473, D1475, I1477, Y1476 and R1495.

Previously, imidazoline I1 binding sites were suggested to be located in the secondary domain of Nischarin (represented in purple in Figure 5.8 and Figure 5.9) and, in IRAS, a site equivalent to BS1 was proposed as the binding site of the I1 ligand agmatine. The BS2 site was identified exclusively in our study and is positioned at the entrance of the cavity of BS1, hence the boundaries of the pockets share a few residues: in Figure 5.12, we compare the aminoacid composition of the pockets in the Nischarin and IRAS sequences. The high sequence identity of residues delineating these binding sites between Nischarin and IRAS suggests that these cavities may serve as binding sites for both I1 and I2 receptors. This is further supported by studies from B. Keller and G. Sevilla, demonstrating that the I2-IR ligand BU99006 upregulates Nischarin expression in the mouse hippocampus, indicating a compensatory response following I2 receptor modification.[193]

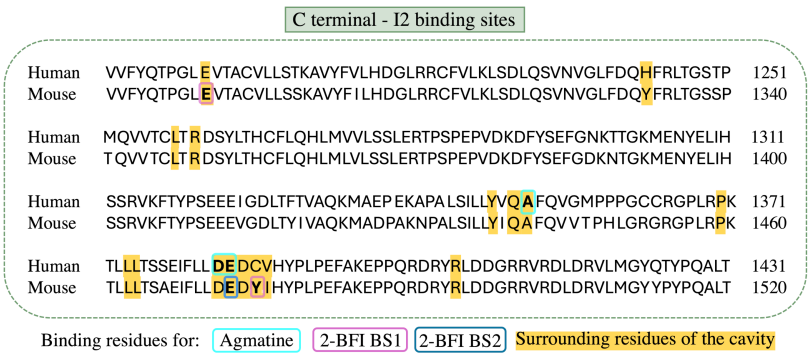


Figure 5.12 Human and Mouse Nischarin sequences indicating the amino acids participating in the C-terminal I2 interaction site.

As previously discussed, Nischarin is involved in various biological processes and interacts with numerous proteins critical for neurological health. Here, we hypothesize that binding of 2-BFI to Nischarin may disrupt protein-protein interactions in the I2 region, potentially downregulating one of its functions and triggering upregulated Nischarin expression as a compensatory response. However, it is possible that given the multiple roles the protein, the excess Nischarin may enhance other functions unrelated to the I2 binding site, potentially resulting in neurological benefits.

5.3.2 Identification of 2-BFI binding sites in S-formylglutathione hydrolase (Uniprot: Q9R0P3)

S-formylglutathione hydrolase (FGH) plays a critical role in detoxifying formaldehyde and protecting cells from oxidative stress caused by reactive oxygen species (ROS) like H_2O_2 and O_2^-.[211,212] It catalyzes the conversion of S-formylglutathione, a conjugate formed when formaldehyde reacts with glutathione (GSH), into GSH and formate, thereby neutralizing formaldehyde and regenerating GSH to maintain cellular redox balance and combat ROS.[211] FGH is widely expressed across various organs, with the highest levels observed in the liver and kidneys.[213] It has critical role in antioxidant defense, oxidative stress management, aging, and cellular health.[212,214,215]

The crystal structure of human Esterase D (ESD, alternate name for S-formylglutathione hydrolase) closely resembles the mouse FGH and adopts a typical α/β-hydrolase fold, with a central twisted β-sheet surrounded by α-helices.[216] The catalytic triad, S149, H260, and D226, forms the active site: S149 initiates substrate attack, forming a covalent intermediate, H260 activates S149 and stabilizes the transition state, and D226 stabilizes H260 via hydrogen bonding.

The reaction proceeds through a tetrahedral intermediate stabilized by the backbone amides of M150 and L54, with a water molecule, activated by H260, completing the hydrolysis.[216]

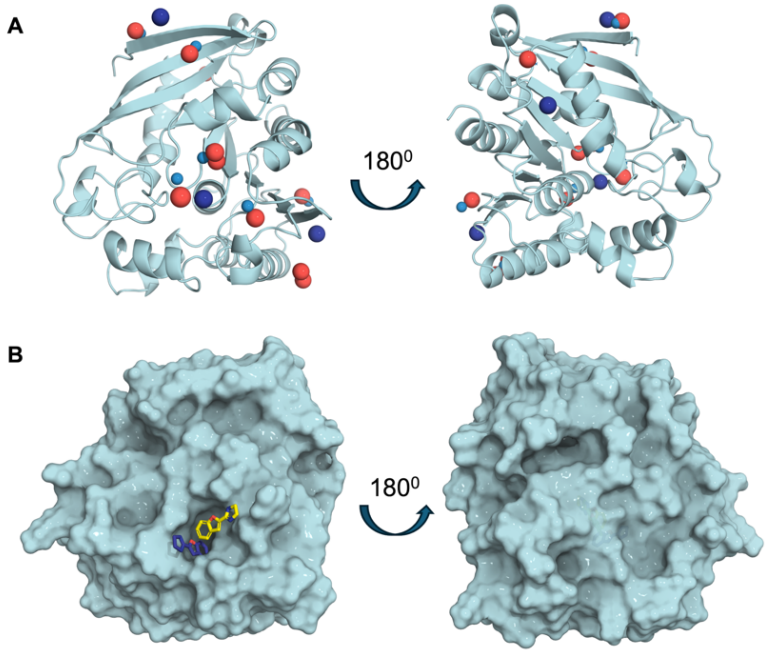


Figure 5.13 (A) Analysis of MDmix simulations for S-formylglutathione hydrolase highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

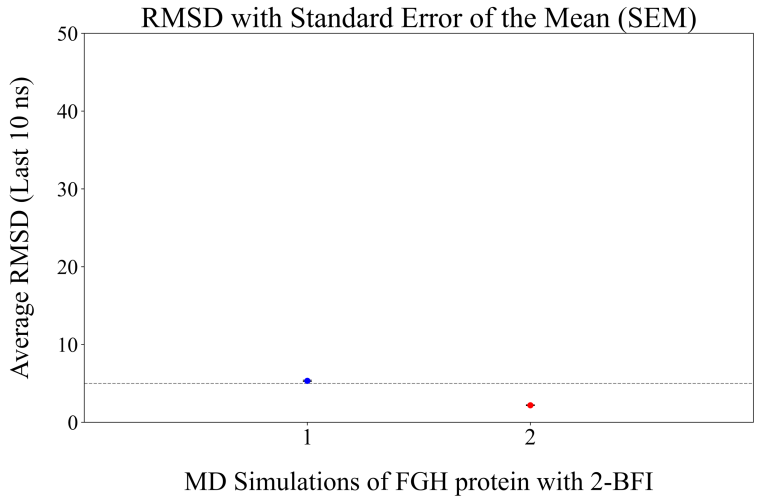


Figure 5.14 The average RMSD values of the 2-BFI ligand relative to its initial position, aligned to the S-formylglutathione hydrolase backbone, were calculated over the final 10 ns of MD simulations, including the standard error of the mean.

Hotspot analysis of S-formylglutathione hydrolase revealed very few potential binding spots and subsequent docking narrowed it down to only two possible binding modes (Figure 5.13). The MD simulations, spanning a total of 300 ns, revealed that molecules in BS2 remained stable throughout the simulation, while molecules in BS1, in two replicas, shifted toward BS2 despite remaining within the protein (Figure 5.14). The average RMSD values are 5.33 $\pm$ 0.08 Å for BS1 and 2.2 $\pm$ 0.04 Å for BS2.

The surrounding residues in the active site are W183, F228, D226, H260, F172, Y262, S149, L54, and M150. 2-BFI interacts with key receptor binding site residues (Figure 5.15): It establishes a cation—π interaction with W97 (4.3 $\pm$ 0.77 Å) and stays in a close proximity with catalytic residue H260 (5.5 $\pm$ 0.8 Å). Furthermore, it seems to interfere with the arrangement of the catalytic residues H260 and S149 increasing the distance between their interacting atoms by 4Å $\pm$ 0.98 Å with respect to the 2.9 Å distance observed in the crystallographic structure

of the human variant (PDB: 3FCX)[216]. The positioning of 2-BFI in the active site of S-formylglutathione hydrolase suggests it may act as a competitive inhibitor of the protein, a hypothesis that will be investigated by experimental assays that are beyond the scope of the current thesis. Furthermore, the potential neuroprotective effects (if any) of this inhibition would need to be further established.

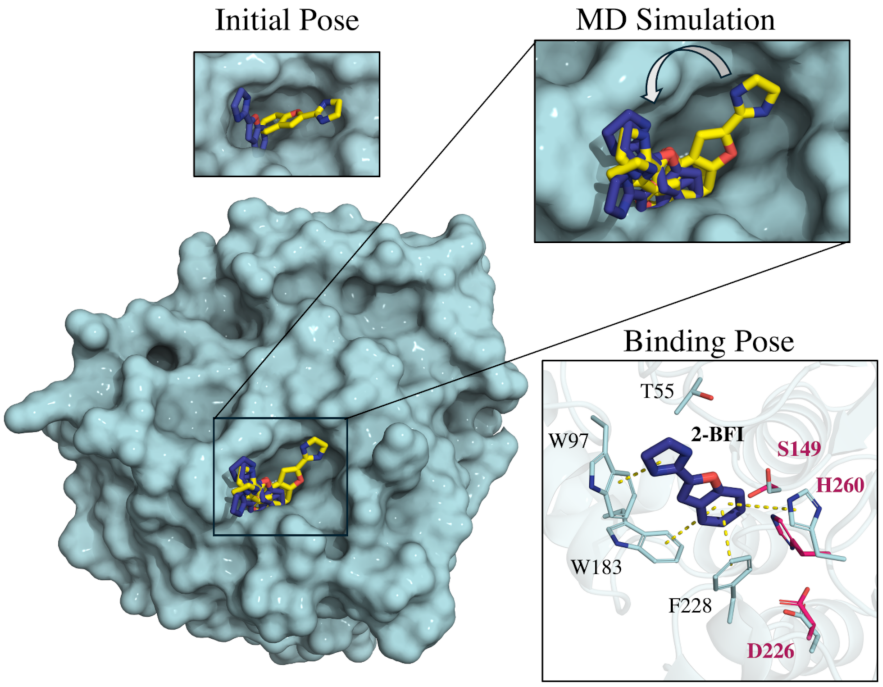


Figure 5.15 The 2-BFI binding site on S-formylglutathione hydrolase is illustrated, showing two binding modes. The yellow molecules represent Binding Site 1 (BS1), as indicated in the RMSD graph, while the dark blue molecules represent Binding Site 2 (BS2).

5.3.3 Identification of 2-BFI binding sites in alpha-2 adrenergic receptor (Uniprot: Q01338)

The alpha-2 adrenergic receptor (α2-AR) belongs to the Class A G-protein-coupled receptors (GPCRs) family. This class of receptors is a well-stablished pharmacological target, since they regulate key processes such as neurotransmitter release, cardiovascular function, regulation of blood pressure, and pain modulation. Structurally, the α2-AR consist of seven transmembrane helices (7TMs) forming an orthosteric binding pocket for norepinephrine (NE).[217] Similarly to many other GPCRs, the binding of NE activates the receptor by inhibiting adenylyl cyclase and reducing intracellular cyclic AMP (cAMP) via the activation of G-protein. As many other GPCRs, α2-ARs are validated therapeutic targets for conditions such as hypertension, pain, and neurodegeneration.[217,218]

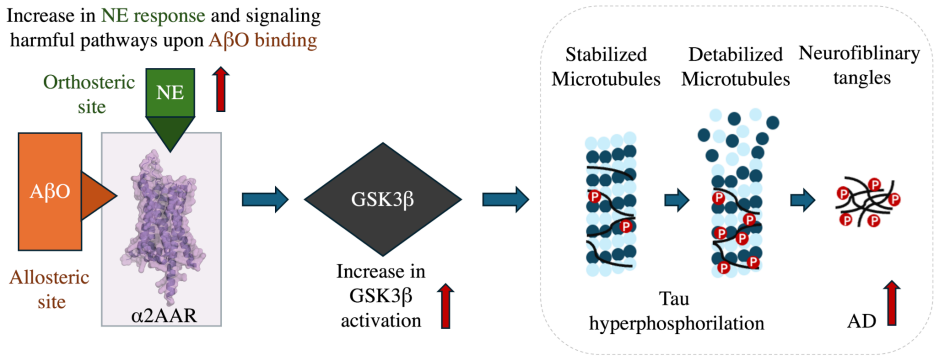


Figure 5.16 AβO allosteric binding, activating glycogen synthase kinase 3β (GSK3β) pathways, leading to tau hyperphosphorylation and neurodegeneration in Alzheimer's disease.

The role of allosteric modulation of α2-AR is well stablished. In the CNS, it has been recently described how Amyloid-β oligomers (AβO) bind allosterically at the third extracellular loop, amplifying norepinephrine signaling and activating pathways like glycogen synthase kinase 3β (GSK3β), leading to tau

hyperphosphorylation, one of the pathological hallmarks of Alzheimer's disease. Furthermore, in preclinical Alzheimer's mouse models, the α2-adrenergic receptor (α2-AR) antagonist and I2 ligand idazoxan has shown promise by reversing GSK3β activation, reducing tau hyperphosphorylation, and lowering Aβ plaque levels. [217] Additionally, it has been shown to increase norepinephrine in the hippocampus. [219] However, it is not yet proven whether idazoxan acts as an orthosteric or allosteric antagonist.

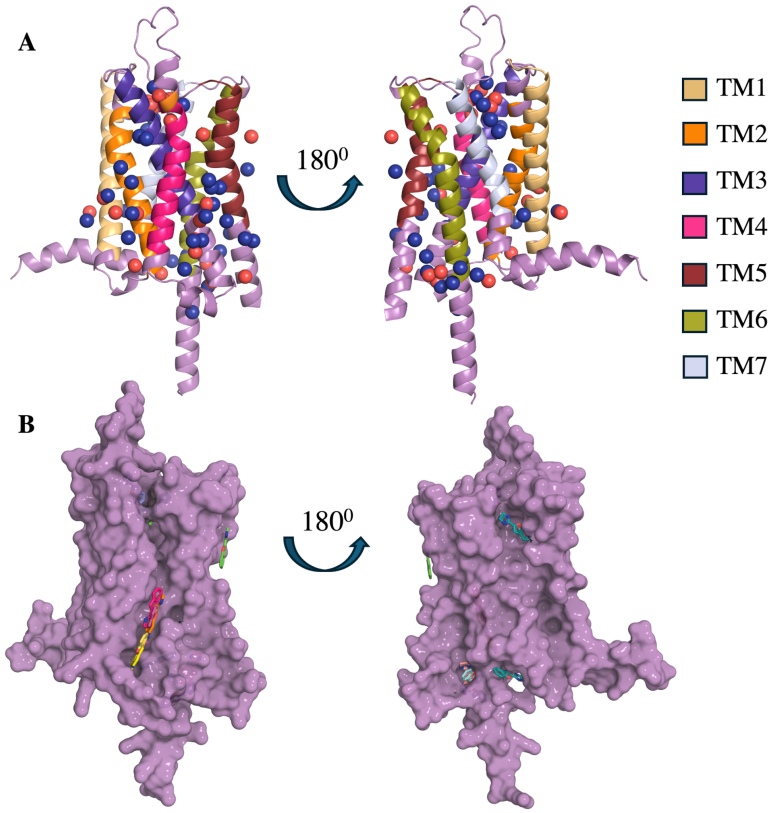


Figure 5.17 (A) Analysis of MDmix simulations for alpha-2 adrenergic receptor highlighting probe hotspots for pyridine (navy blue) and ethanol/isopropanol (red). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

The analysis of α2-AR revealed 13 putative binding sites over 38 identified solvent binding hotspots, predominantly located around the transmembrane domains and the orthosteric site (Figure 5.17). Subsequent MD simulations (1.95 μs in total) identified one stable binding site: BS1, with an RMSD value of 1.7 $\pm$ 0.03 Å (Figure 5.18).

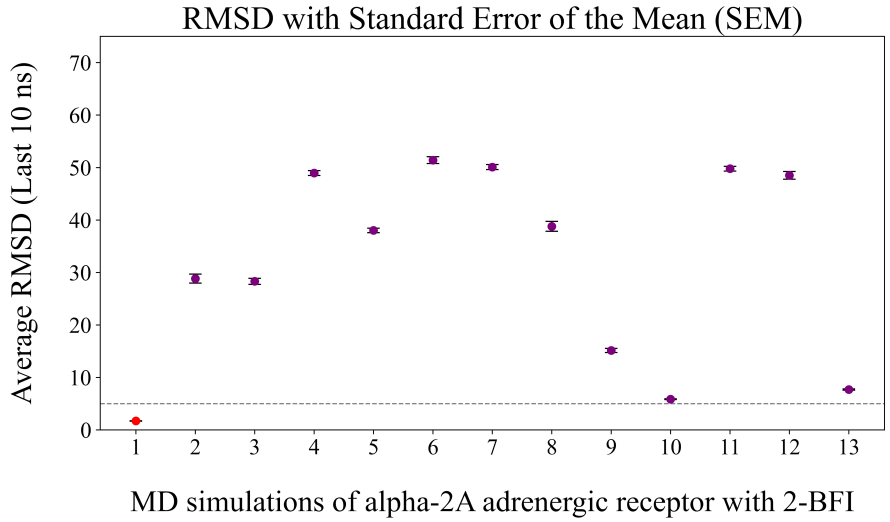


Figure 5.18 The average RMSD values of 2-BFI, relative to its initial position, were calculated over the final 10 ns of MD simulations, with graphs showing data for three replicas and the standard error of the mean.

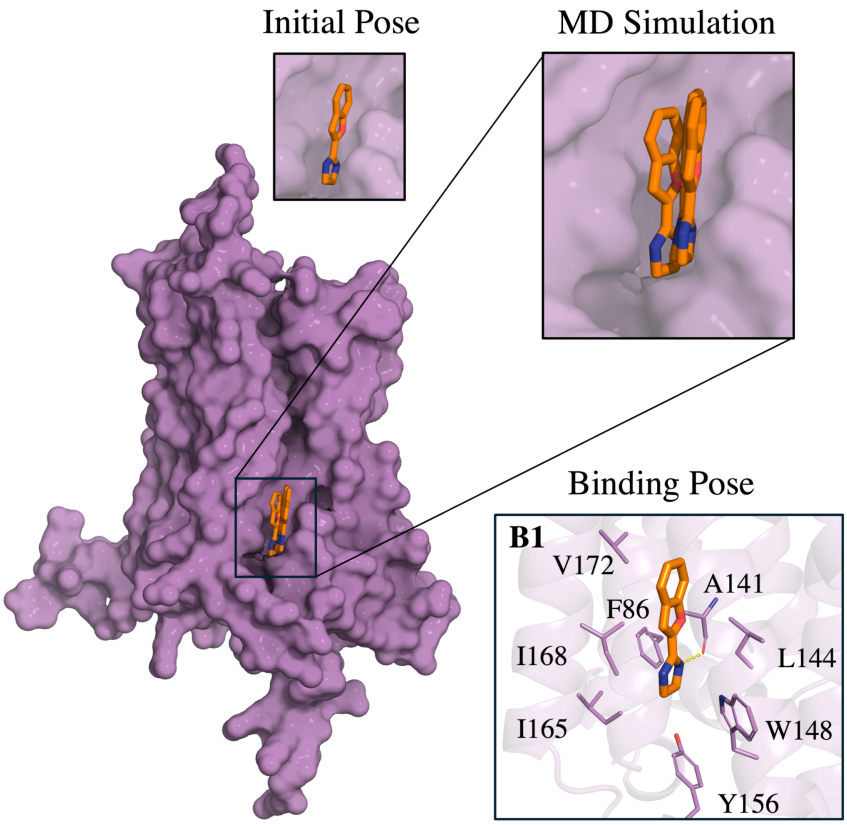


Figure 5.19 The 2-BFI binding site on alpha-2 adrenergic receptor is illustrated. Orange molecules represent binding site B1, an allosteric site, in the transmembrane TM345 region.

The cavity of BS1 is delineated by residues V172, F86, A141, I168, I165, Y156, W148, and L144. 2-BFI forms a hydrogen bond with the backbone carbonyl of A141 via the imidazoline ring of 2-BFI with a distance of 3.15 ± 0.29 Å (Figure 5.19). Our study suggests the potential presence of an imidazoline-binding site in this transmembrane region (TM345_mid), however our simulations did not include an atomistic description of the membrane and further simulations should be performed to validate this potential binding site. Nevertheless, recent

computational studies have identified novel allosteric sites in Class A GPCRs in the A-EH-TM345_mid site, in the same region as BS1. Furthermore TM3, TM4, and TM5 serve as a critical interface for structural communication, facilitating signaling between the extracellular ligand-binding region and the intracellular G-protein interaction region, thereby stabilizing or destabilizing the receptor's activation state and modulating its activity.[220] Additionally, the role of the I2 ligand idazoxan as an α2-AR antagonist and its role in inhibiting allosteric pathways mediated by AβO,[217] hints that BS1 binding site could be involved in the distortion of the allosteric signaling triggered by AβO binding.

5.3.4 Identification of 2-BFI binding sites in Heterogeneous nuclear ribonucleoprotein U (Uniprot: Q8VEK3)

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a family of RNA-binding proteins essential for RNA metabolism, including transcription, splicing, stabilization, transport, and translation regulation.[221] Among them, hnRNP U, also known as scaffold attachment factor A (SAF-A), is a multifunctional nuclear protein involved in chromatin organization, transcription regulation, RNA processing, and nuclear architecture. Structurally, mouse hnRNP U comprises six key structured regions: the SAP domain (8–42) for chromatin binding,[222,223] the SPRY domain (244–440),[224] which function is not completely understood, but is suggested to mediate protein-protein interactions, the ATPase region (464–648) which binds ATP[225] and is essential for SAF-A oligomerization, the actin-binding region (587–602) that mediates cytoskeletal interactions, the coiled-coil region (626–653) that stabilizes complexes with other proteins and the RNA-binding RGG box (690–715) that is responsible to stablish the interactions with RNA molecules. Moreover, hnRNP U features an Intrinsically Disordered Domain (IDD) that is considered to enhance its flexibility in interacting with RNA and chromatin.

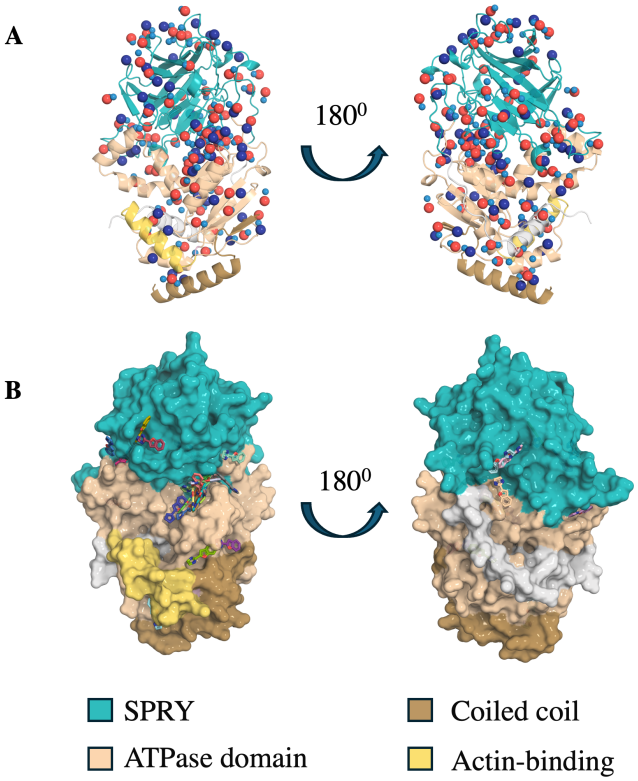


Figure 5.20 (A) Analysis of MDmix simulations for Heterogeneous nuclear ribonucleoprotein U highlighting probe hotspots for pyridine (navy blue) and ethanol/isopropanol (red) throughout ATPase, SPRY, Coiled coil and Actin-binding regions. (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

Due to the low reliability of the AF2 prediction for the disordered regions of the protein, we focused our analysis in the six structured domains of hnRNP U and identified 90 solvent binding hotspots, which were further explored through docking simulations. These docking studies revealed 25 specific binding poses with docking scores below zero (Figure 5.20), that were further assessed using MD simulations (total simulation time 3.75 μs). These simulations identified only one

binding site structurally stable (BS17) with an RMSD value of 3.99 ± 0.27 Å (Figure 5.21).

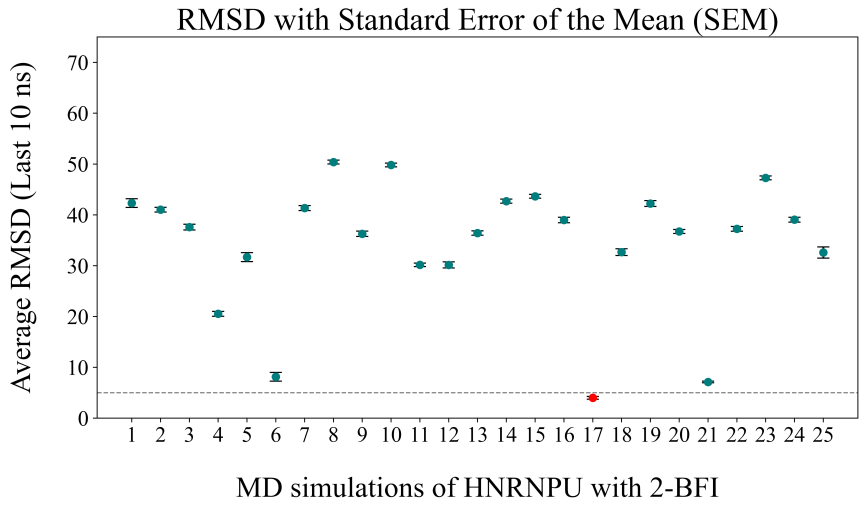


Figure 5.21 The average RMSD values of 2-BFI, relative to its initial position, were calculated over the final 10 ns of MD simulations, with graphs showing the average value for for three replicas and the standard error of the mean.

The BS17 is located at the interface between the SPRY and the ATPase domains, delineated by T343, N270, Y269 and P464. The charged nitrogen atom in the imidazoline ring of 2-BFI forms hydrogen bond with the side-chain amide group of N270 (3.5 ± 0.73 Å) and it is in close proximity with the carbonyl oxygen of T343 (4.90 ± 1.20 Å), stabilizing its position within the cavity (Figure 5.22). Dysregulation of hnRNP U is linked to AD, where it regulates the translation of amyloid precursor protein (APP) mRNA, thereby influencing amyloid-beta production. hnRNP U is also essential for chromatin stability, RNA metabolism, and synaptic function, making it critical for neuronal health. Its dysfunction is associated with neurodegenerative processes and chronic neuroinflammation.[225,226] Although the binding site of 2-BFI is located away from the RNA and DNA interaction sites of hnRNP U, we speculate that binding at BS17 could

allosterically modulate the protein function and maybe decrease the transcription of APP. However, the interaction between hnRNP U and I2 ligands has not been previously reported, and therefore, these results would require of further experimental validation.

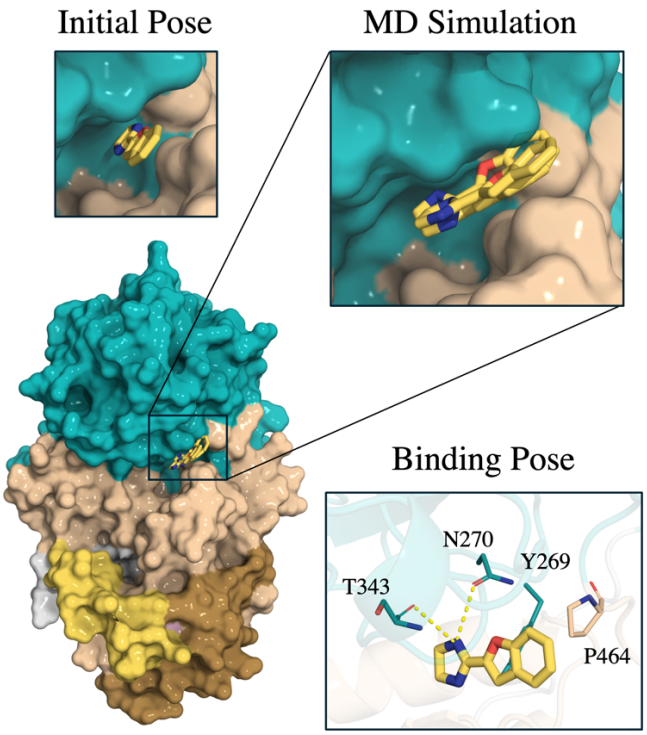


Figure 5.22 The binding mode of 2-BFI on heterogeneous nuclear ribonucleoprotein U is illustrated, depicting one stable binding pose identified in this study.

5.3.5 Identification of 2-BFI binding sites in Hydroxymethylglutaryl-CoA lyase (Uniprot: P38060)

Hydroxymethylglutaryl-CoA lyase (HMG-CoA lyase or HMGCL) is a key mitochondrial enzyme in ketogenesis, catalyzing the Mg^{2+}-dependent cleavage of HMG-CoA into acetoacetate and acetyl-CoA, precursors of ketone bodies like beta-hydroxybutyrate (BHB).[227,228] This process provides alternative energy sources for the brain during glucose scarcity, such as fasting, ketogenic diets (KD), or prenatal development. The critical role of HMGCL is evidenced by knockout studies in mice, causing embryonic lethality, and mutations in humans leading to fatal metabolic diseases.[229] In neurological diseases like AD and PD, where glucose metabolism is impaired, ketones generated via HMGCL support neuronal energy needs and exhibit neuroprotective effects.[230,231] Studies show ketones enhance mitochondrial function, reduce oxidative stress by increasing NAD+/NADH ratios, and modulate inflammation by decreasing microglial activation)(**Figure 5.23**).[232] BHB has been shown to reduce amyloid-beta toxicity and enhance brain function in Alzheimer's disease (AD) models. Clinically, ketogenic interventions, such as medium-chain fatty acid supplementation that increases ketosis, have been shown to improve cognitive function and delay AD progression by alleviating oxidative stress and inflammation.[233,234] These findings highlight HMGCL as a crucial enzyme and promising therapeutic target for neurodegenerative diseases.

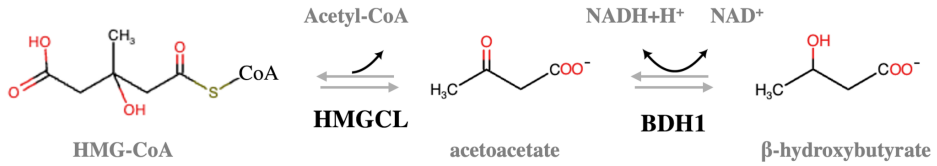


Figure 5.23 Hydrolysis of HMG-CoA into acetoacetate and acetyl-CoA catalyzed by HMG-CoA lyase and conversion of acetoacetate into beta-hydroxybutyrate by BDH1.

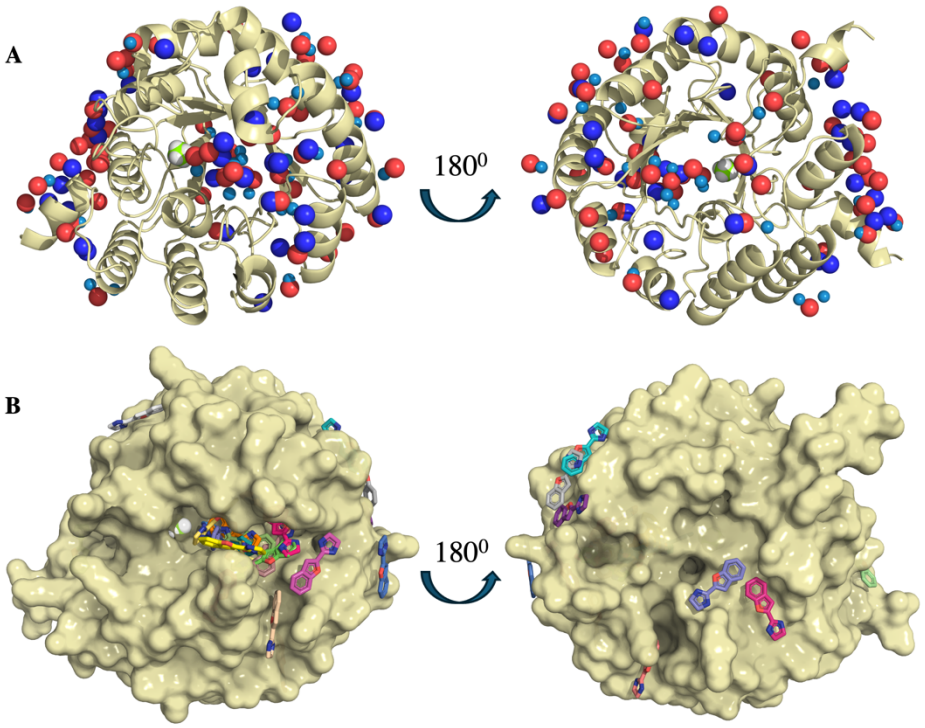


Figure 5.24 (A) Analysis of MDmix simulations for hydroxymethylglutaryl-CoA lyase (HMG-CoA lyase) highlighting probe hotspots for pyridine (navy blue) and ethanol/isopropanol (red). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

MDmix simulations revealed 63 high-affinity interaction spots, of which docking calculations identified 29 putative binding sites with favorable docking scores. Further analysis through MD simulations (4.35 μs in total) highlighted one stable binding site among the identified candidates (Figure 5.24).

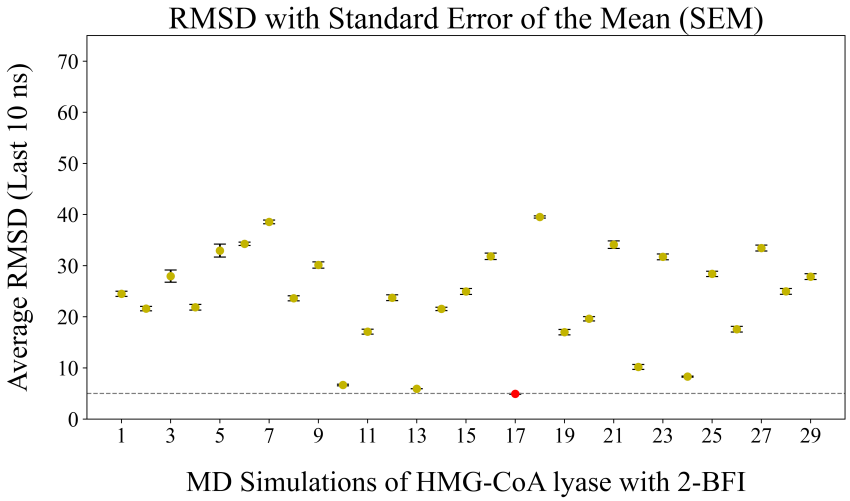


Figure 5.25 The average RMSD values of 2-BFI, relative to its initial position, were calculated over the final 10 ns of MD simulations, with graphs showing data for three replicas and the standard error of the mean.

BS17 in Figure 5.25 displayed an RMSD value of 4.5 ± 0.13 Å. The relatively high RMSD is attributed to the flexibility of the loop covering the entrance to the orthosteric binding site. Despite the fact that the active site offers multiple high-affinity interaction spots, the 2-BFI molecules located inside failed to remain bound within the cavity in a stable manner, and instead 2-BFI binds at the entrance of the active site.

The cavity that accommodates 2-BFI is surrounded by residues N138, A117, N140, K137, W81, V82, and V77. Among these, 2-BFI forms a cation–π interaction with W81 at a distance of 4.13 ± 0.89 Å. The hydrogen bond with N140 through its imidazoline ring, below the distance of 3.5 Å, was observed in only 10% of the trajectory due to the flexibility of the loop where N140 is located, and as a result, it does not provide a considerable hydrogen bonding interaction. (Figure 5.26).

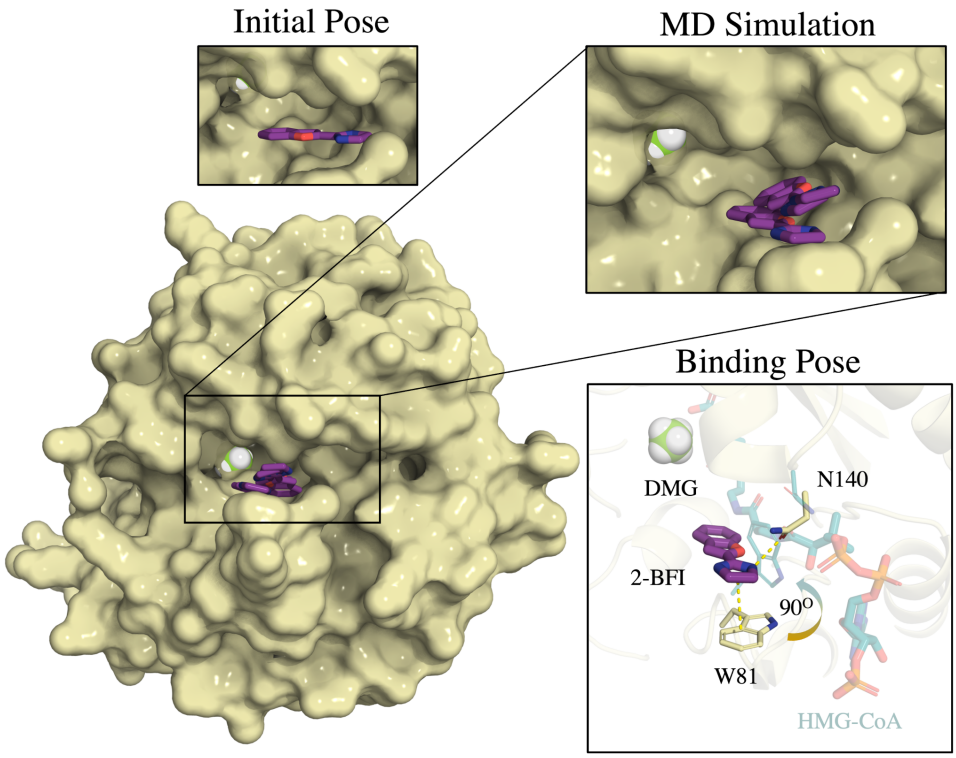


Figure 5.26 The binding mode of 2-BFI (purple) with dummy magnesium ion (green) on mouse HMGCL (yellow) is illustrated by superimposing the crystal structure of the human HMG R41M mutant bound with HMG-CoA (blue), highlighting one stable binding pose identified in this study and its spatial relationship to the natural substrate, HMG-CoA.

To compare the position of 2-BFI with that of the substrate and propose a potential mechanism of action, we superimposed the crystal structure of the human HMGCL catalytically deficient R41M mutant bound to the substrate HMG-CoA (PDB: 3MP5)[235]. This structure reveals a conserved $(\alpha/\beta)_8$ TIM barrel fold and a flexible loop (L263–N275) that repositions upon substrate binding to facilitate catalysis. The catalytic site is defined by key residues, including D42, R41, H233, H235, and C266, which coordinate the Mg^{2+} ion and play crucial roles in HMG-CoA catalysis.[235]

Although 2-BFI is positioned far from the reaction site, stabilization of the substrate HMG-CoA is facilitated through interactions with S169, T205, and E72 in the active site, as well as N138, N140, W81, and S78 at the entrance, which ensure precise alignment during the reaction. Interestingly, 2-BFI interacts with N140 and W81, which rotate by 90 degrees relative to their position in the HMG-CoA-bound structure (Figure 5.26). This proximity highlights a potential effect of 2-BFI in the binding of HMG-CoA within the enzyme. However, whether 2-BFI binding exerts a stabilizing or destabilizing effect on HMGCL cannot be derived from the simulations. Further experiments are needed to clarify the potential role of 2-BFI in the metabolism of HMG-CoA.

5.3.6 Identification of 2-BFI binding sites in creatine kinase – B (Uniprot: Q04447)

Creatine kinase–B (CK-B) is expressed in neurons, oligodendrocytes, and astrocytes. It is a homodimeric enzyme essential for maintaining energy homeostasis in the CNS.[236] It catalyzes the reversible phosphorylation of creatine from ATP to produce phosphocreatine and ADP. This process supports energy-demanding functions, such as ion transport via Na^+/K^+ ATPase and ATP-gated K^+ channels, with mitochondrial CK-B generating phosphocreatine as an energy reserve and cytoplasmic CK-B regenerating ATP during periods of high energy demand.[197,237,238] CK-B contains an N-terminal helical domain (residues 1–100) and a C-terminal α/β domain (residues 125–381), connected by a hinge region linker.[197] Structural studies reveal two conformational states: a ligand-free open form and a ligand-bound closed form. Due to its critical role in brain energy metabolism, CK-B is implicated in neurodegenerative diseases like Alzheimer's, where its dysfunction disrupts energy regulation and contributes to neuronal damage.[197]

Experimental evidence from the literature indicates that CK-B contains an I2-R motif. In an experimental autoimmune encephalomyelitis (EAE) model of multiple

sclerosis (MS), spinal cord injury in mice reduced brain CK-B activity, leading to disrupted energy production, impaired calcium homeostasis, and increased tissue damage. Treatment with 2-BFI restored CK-B activity, stabilizing ATP production, regulating calcium levels, and preventing calpain overactivation. Proteomic studies reported by Wang et al. further support CK-B's involvement in EAE pathology and suggest that 2-BFI stabilizes this enzyme, consistent with proteomics results from the Pallàs group. In these studies, 2-BFI improved neurological outcomes and reduced tissue damage, demonstrating its potential to modulate neurological diseases. Additionally, previous molecular docking studies using the crystal structure of chicken brain CK-B and the irreversible I2 ligand BU990006 identified key residues T71, V72, V75, L201, L202, C283, and S285, as critical for their interaction.[60]

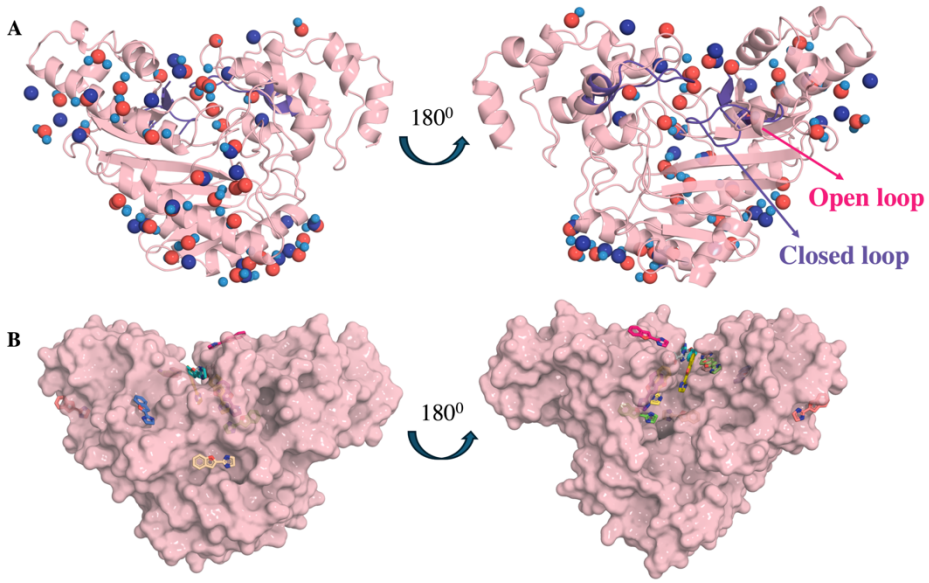


Figure 5.27 Hotspot analysis from MDmix simulations for Mus musculus creatine kinase B-open (mCK-B_open), with the closed loop superimposed, is shown in A. The analysis highlights pyridine (navy blue) and ethanol/isopropanol (red) probes. These identified hotspots were applied as pharmacophoric restraints in docking studies, with molecules scoring below zero depicted in **B**.

Considering the conformational plasticity of CK-B reported in the literature, we explored the existence of binding hotspots in both an open conformation (hereafter CK-B_open) and closed conformation (CK-B_closed) of CK-B. MDmix simulations identified 25 and 36 high-affinity interaction spots for CK-B_open and CK-B_closed, respectively, with subsequent docking reducing the putative binding sites to 12 and 25, respectively (Figure 5.27). MD simulations (aggregated simulation time of 5.6 μs) revealed that, despite the higher number of solvent interaction hotspots and favorable docking sites in CK-B_closed none of them was stable during the course of the simulations. In contrast, in the open conformation one binding site (BS10) was identfied with and average RMSD of 3.62 ± 0.01 Å. (Figure 5.28). The relatively high RMSD value can be attributed to the ligand's initial position being slightly outside the binding cavity. Throughout the simulation, several conformational changes allowed the ligand to adjust and accommodate itself in closer proximity to the residues surrounding the binding site (Figure 5.29).

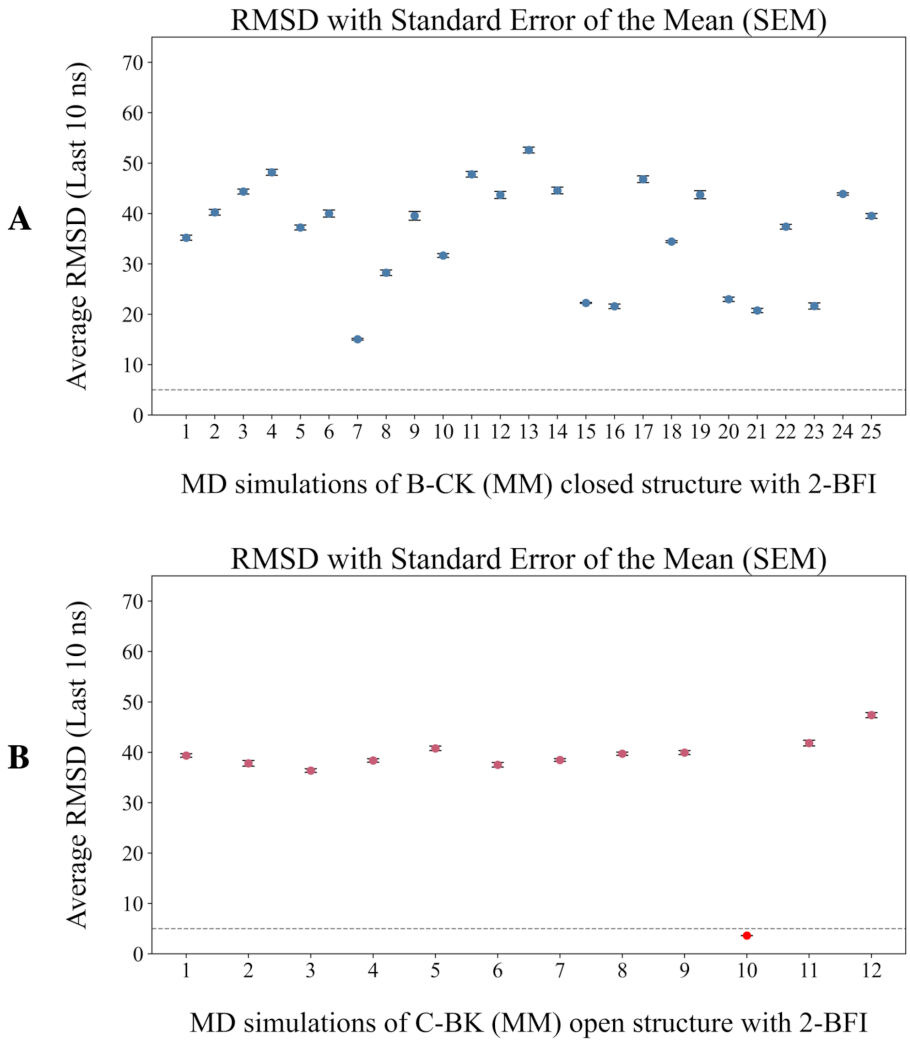


Figure 5.28 Average RMSD values of 2-BFI relative to its initial position, calculated after aligning the protein backbone during the final 10 ns of MD simulations. Graphs represent three replicas with standard error of the mean, highlighting red points below the 5 Å threshold. Results are shown for (**A**) mouse CK-B_closed including ADP and MG, and (**B**) mouse CK-B_open.

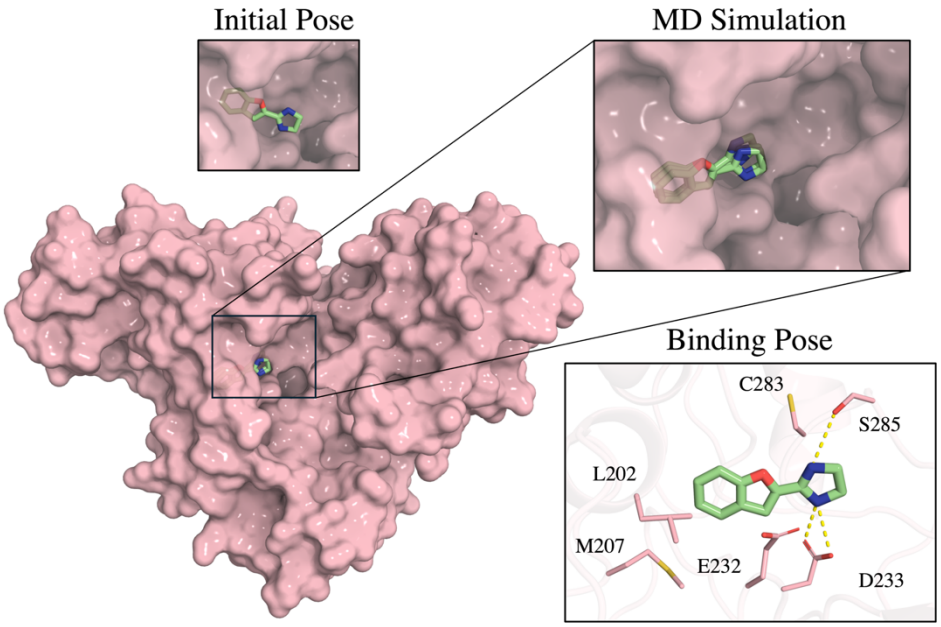


Figure 5.29 The binding mode of 2-BFI on mouse CK-B_open is illustrated, depicting one stable binding pose identified in this study.

In the final position of BS10, 2-BFI is accommodated in the cavity delineated by residues E232, S205, V75, T282, C283, S285, P144, L202, and L201. 2-BFI forms hydrogen bonds with S285 (3.5 ± 0.4 Å) and with D233 (2.9 ± 0.26 Å and 3.0 ± 0.19 Å to each carboxylic oxygen) through the nitrogens in its imidazoline ring. To compare the position of 2-BFI relative to the active site substrates, we superimposed the final pose of BS10 with the crystal structure of closed human creatine kinase B (PDB: 3B6R).[197] Key residues in this human structure include H296 and H191 for nucleotide stabilization; R130, R132, R236, R292, and R320 for phosphate binding and transition-state stabilization; C283 and S285 for anchoring and positioning creatine; E232 and V72 for orienting the guanidine and carboxylate groups of creatine; and I69 and V325 for forming the creatine-binding pocket. Additional contributions include D335 for water coordination and Mg^{2+} for substrate alignment.[197]

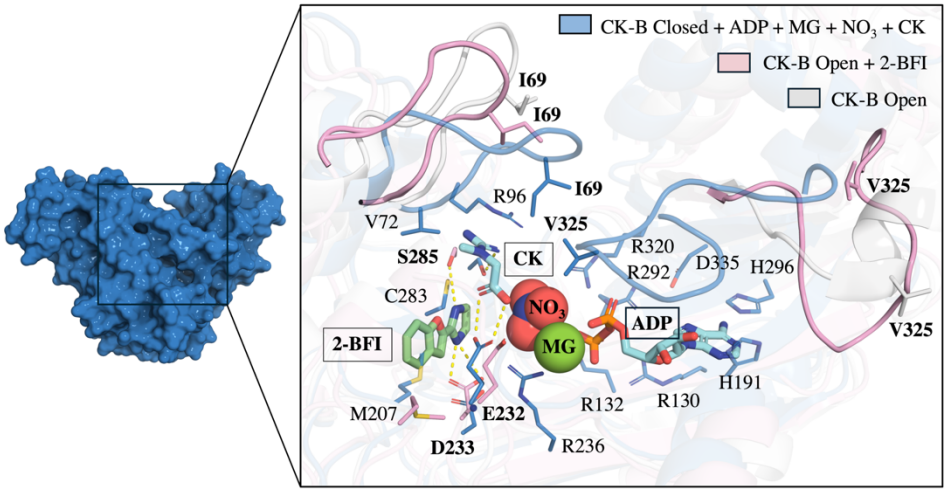


Figure 5.30 Mouse CK-B_open (pink), bound to 2-BFI (green) represents the last frame of the trajectory and it is superimposed on the unbound mouse crystal CK-B_open structure (grey) and crystal structure of human CK-B_closed (blue), which contains magnesium, ADP, creatine, nitrate (NO_3^-), and key active-site residues (PDB: 3B6R).

As shown in the Figure 5.30, 2-BFI is not positioned directly in the catalytic center but is in close proximity to the reacting creatine and its interacting residues. The only residue shared between creatine and 2-BFI is S285, which simultaneously interacts with the nitrogen atom of 2-BFI's imidazoline ring and the guanidine group of creatine. Given that 2-BFI has been identified as an enhancer of CK-B activity, we hypothesize that it induces conformational rearrangements in the active site, potentially stabilizing the reaction's transition state. Further simulations on a structure already accommodating 2-BFI would be required to validate its role in stabilizing the substrates, but this is beyond the scope of this thesis.

CHAPTER 6.
THE ALLOSTERIC EFFECT OF 2-BFI ON MAO-B

6.1 Background

Monoamine oxidase B (MAO-B) is a mitochondrial enzyme predominantly located in glial cells and serotonergic neurons. It catalyzes the oxidative deamination of dopamine and other biogenic amines using its co-factor FAD (flavin adenine dinucleotide), which is covalently attached to C397. This reaction, occurring in a specialized hydrophobic active site, involves key residues Y398 and Y435 forming an "aromatic cage" that stabilizes dopamine's amino group and positions it for optimal interaction with the FAD cofactor. [239,240] The reaction generates neurotoxic products such as 3,4-dihydroxyphenylacetaldehyde (DOPAL), ammonia, and hydrogen peroxide (H_2O_2).[239,240] DOPAL is particularly harmful to dopaminergic neurons as it can promote protein aggregation and cellular stress, while H_2O_2 contributes to oxidative stress, leading to neuronal damage and neurodegeneration. Elevated MAO-B activity, which increases with age, is associated with Alzheimer's and Parkinson's diseases, promoting reactive oxygen species (ROS) production, microglial activation, and amyloid beta plaque formation. Its central role in dopamine metabolism and its contribution to neurotoxic by-products make MAO-B an important target for therapeutic intervention. Recent studies have also highlighted MAO-B's interaction with Imidazoline I2 ligands, further emphasizing its importance in neurodegenerative disease mechanisms.[40,194]

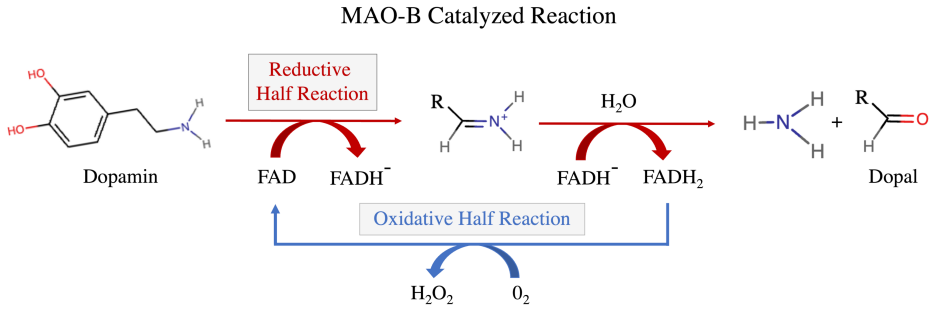


Figure 6.1: Complete metabolic reaction of the neurotransmitter dopamine, showing its enzymatic oxidation and the production of hydrogen peroxide as a byproduct.

Particularly 2-BFI shows promise for selective modulation and therapeutic potential.[194,241] It binds MAO-B through a non-competitive mechanism, indirectly affecting its activity without interacting with the enzyme's catalytic site.[192] This suggests that I2 binding sites (I2BS) might function as regulatory sites for MAO-B, influencing key neurological functions. Monoamine oxidase B (MAO-B) is inhibited by 2-(2-Benzofuranyl)-2-imidazoline (2-BFI) with a (K_i) of approximately 8.3 μM. While 2-BFI shows moderate inhibition, its binding affinity improves significantly when MAO-B is pre-treated with tranylcypromine. This covalent modification induces a conformational shift, involving residues I199 and Q206, creating a hydrophobic environment. Consequently, tranylcypromine-inhibited MAO-B binds 2-BFI with a (K_d) of 9 ± 2 nM, reflecting an increased binding energy (ΔG) of 3.9 kcal/mol and stronger affinity.[194]

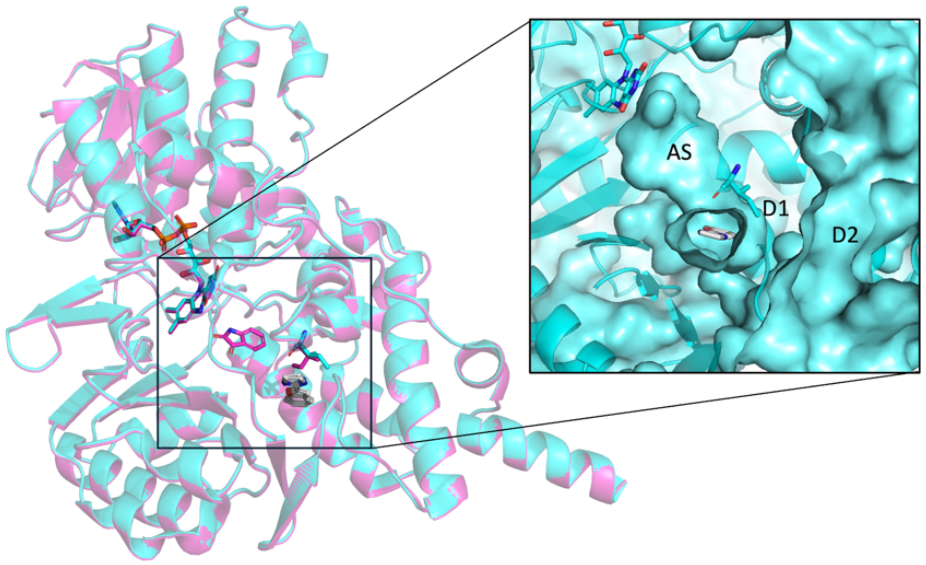


Figure 6.2: Superimposed structures of MAO-B: PDB 2XFN [194](blue) showing an open Ile199 conformation and PDB 2XFP [194]with the inhibitor isatin, inducing a closed Ile199 conformation. Both structures include 2-BFI (grey) in the entrance cavity. Further MD simulations were performed with 2-BFI, dopamine, and DOPAL placed in the active site (AS), entrance cavity (D1), and the newly identified surface cavity (D2).

In this chapter two crystal structures of MAO-B have been used. In PDB 2XFN,[194] 2-BFI is located in the entrance cavity adjacent to the active site, with Ile199 in an open conformation, connecting the entrance cavity to the active site. In PDB 2XFP,[194] 2-BFI is also found in the entrance cavity, but Il99 adopts a closed conformation due to the presence of the inhibitor isatin in the active site, whose presence was not considered during our study. As discussed in Chapter 5, we observed the dynamic binding of 2-BFI between the entrance cavity (D1) and a newly identified binding site on the protein surface (D2). This chapter aims at characterizing these interaction sites, as well as their relationship with the active site (AS), and their potential role in modulating the entrance and exit pathways of the natural substrate dopamine and its metabolite DOPAL. A second objective of

this chapter is to investigate the allosteric modulation of 2-BFI on the reaction barrier of the rate limiting step of dopamine's enzymatic degradation. Using the ONIOM QM/MM methodology, we aim to explore the dopamine reaction, focusing on the previously identified hydride transfer mechanism involved in dopamine conversion.

6.1.1 Catalysis of dopamine deamination by MAO-B

To lay the groundwork, we have examined the reaction mechanism, which has been widely debated with alternative proposals over the years. Early models, such as the radical mechanism,[242] suggested dopamine oxidation via one-electron transfers to form radical intermediates, but this was dismissed due to the lack of experimental evidence, such as paramagnetic resonance signals. Similarly, the polar nucleophilic mechanism,[243] proposing proton transfer from dopamine's α-carbon to FAD's N5 atom, was deemed unlikely due to high activation barriers. The hydride transfer mechanism is currently considered the most likely pathway. Although there is ongoing debate about whether the mechanism involves a direct hydride transfer or a two-step process, Erdem et al. supported a direct hydride transfer mechanism in their QM/QM gas-phase study,[244] while Mavri et al., using a QM/MM EVB approach, proposed a two-step mechanism.[245] The direct mechanism involves a single, concerted hydride transfer from the substrate's α-carbon to the N5 atom of FAD, forming an iminium cation and reduced flavin. In contrast, the two-step mechanism involves an intermediate where the hydride first forms a covalent adduct with flavin N5, followed by proton abstraction at flavin N1, assisted by water molecules, yielding the imine and reduced flavin.[245,246]

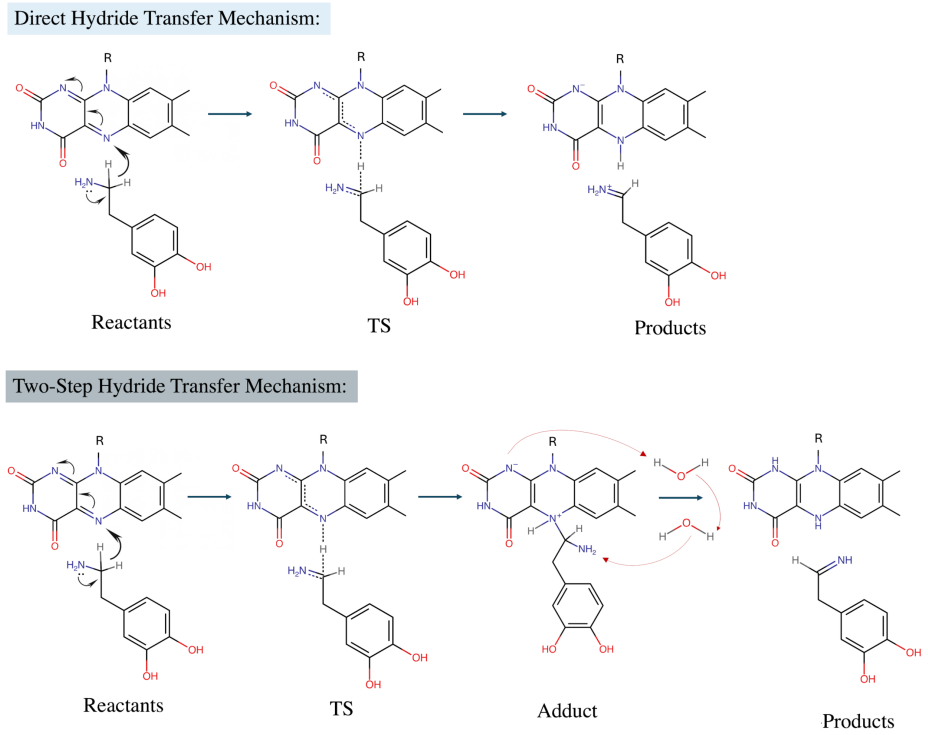


Figure 6.3: Illustration of the direct hydride transfer and two-step hydride transfer reaction mechanisms of dopamine with FAD, depicting the reactant, transition state, intermediate adduct, and product states.

Mavri et al. emphasized the importance of proper substrate orientation within the active site, where the amino group of dopamine forms a hydrogen bond with the O4 atom of FAD.[245]This spatial arrangement ensures alignment of the pro-R hydrogen of dopamine with FAD, enabling efficient hydride transfer. Additionally, they indicated that the deprotonated form of dopamine is essential for catalysis, despite the substrate binding in its protonated form. Water molecules facilitate intraenzymatic deprotonation, and a free energy cost of 1.9 kcal/mol for transfer to the bulk solvent was estimated for dopamine. In their simulations, adding this energy to the reaction barrier of 14.2 kcal/mol led to a total activation free energy of 16.1 kcal/mol, closely matching the experimental value of 16.5 kcal/mol, supporting the hydride transfer mechanism.[247]

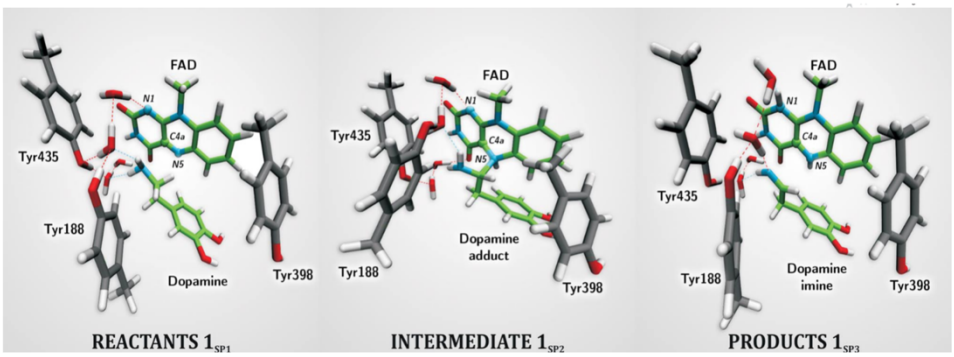


Figure 6.4: The key stages of the hydride transfer mechanism, as determined by Mavri et al. in their previous QM/MM study, are illustrated.[245]

In a recent study, Stare et al. utilized DFT calculations (employing the M06-2X functional with the 6-31+G(d,p) basis set) to investigate the energy barriers of the two-step hydride transfer mechanism involving the triple-ring isoalloxazine moiety of FAD and PEA, a representative substrate of MAO. The calculated energy barriers varied across different conditions: ~35.8 kcal/mol in the gas phase, ~31.7 kcal/mol in implicit solvent, and 24–26 kcal/mol in explicit solvent with several water molecules. These results highlight the importance of environmental stabilization in facilitating the reaction. By systematically removing water molecules and re-optimizing intermediate structures, they showed that at least four water molecules are necessary to stabilize the intermediate adduct.[248]

Based on insights into the reaction mechanisms, our goal is to generate a detailed reaction profile for dopamine both alone and with 2-BFI to understand the impact of 2-BFI on the catalytic process, considering direct electrostatic effects exherted by the 2-BFI ligand on the reaction barrier but also allosteric changes in the binding site structure and substrate binding mode. In collaboration with Prof. Maria João Ramos from the University of Porto, we used the ONIOM methodology to obtain the reactant, transition state, and product structures.

6.2 Computational details

6.2.1 MD simulations

MD simulations were conducted to investigate the dynamic behavior of the human MAO-B protein-ligand systems using crystal structures of MAO-B obtained from the Protein Data Bank (PDB IDs 2XFP and 2XFN),[194] retaining only the crystal waters and the FAD cofactor covalently bound to the protein. 2-BFI was placed in the AS, D1, and D2 positions based on the binding site analysis presented in Chapter 5. Dopamine and DOPAL (3,4-dihydroxyphenylacetaldehyde) were manually positioned into the active site by aligning dopamine with the inhibitor's binding mode observed in the 2XFP structure. Additionally, dopamine was manually aligned to the D1 and D2 cavities based on the binding position of 2-BFI.

System preparation was performed using the tleap tool from AmberTools20,[181] applying the ff14SB[179] force field for the protein and the general AMBER force field (GAFF2)[99] for the ligands. Specific parameters for 2-BFI and dopamine were generated accordingly (as indicated in chapter 5). The combined system, including the protein, FAD cofactor, ligands, and structural water molecules, was solvated in a truncated octahedral box of TIP3P[180] water molecules, ensuring a minimum distance of 12 Å between the solute and the box edge. The system was neutralized by adding sodium (Na^+) and chloride (Cl^-) ions. Hydrogen mass repartitioning[123] was applied, allowing for an increased integration time step of 4 fs. Energy minimization was conducted in two stages: initially, the solvent and ions were minimized with a harmonic restraint of 10 kcal/mol·Å^2 on the protein-ligand complex using the steepest descent method, followed by conjugate gradient minimization without restraints. Equilibration involved five heating steps, gradually increasing the temperature from 0 K to 300 K under the NVT ensemble, each step lasting 50 ps, followed by 250 ps under the NPT ensemble at 1 bar pressure. The SHAKE algorithm was employed to constrain bonds involving

hydrogen atoms, and a Langevin thermostat with a collision frequency of 1 ps^{-1} was used to control the temperature. Production MD simulations were carried out at 300 K in the NVT ensemble for 1 μs for each replica with a 4 fs time step. Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) method, with an 8 Å cutoff for non-bonded interactions. The total simulation time extended to a total of 10 μs with 10 replica, with a restraint of 10 kcal/mol·$Å^2$ applied to residues 482–501, which are located within the membrane. All simulations were executed using the pmemd.cuda module from the AMBER20 software package.[181]

Positioning dopamine, DOPAL, and 2-BFI individually in three distinct locations, the active site (AS), D1 (entrance cavity), and D2, resulted in 9 different systems. The same systems were recreated using the 2XFP MAO-B structure, yielding a total of 18 systems. Each system was simulated for 1 μs across 10 replicas, resulting in a cumulative simulation time of 180 μs. Additionally, a separate set of simulations was performed using neutral dopamine and 2-BFI, corresponding to an additional 60 μs of simulation time. These simulations started from the three binding locations.

To further investigate the effect of ligand co-occupancy on their behavior and affinities, double-ligand systems were prepared, as summarized in Table 6.1. In these systems, two different ligands were placed in two distinct positions (AS, D1, or D2). A total of 11 double-ligand systems were simulated using the 2XFN MAO-B structure, incorporating protonated and deprotonated dopamine, 2-BFI, and neutral DOPAL. These simulations, with a combined simulation time of 110 μs, were conducted to analyze how the presence of one ligand influences the dynamic behavior and binding affinity of the other.

6.2.2 ONIOM QM/MM calculations

QM/MM studies were based on the human monoamine oxidase B (MAO-B) crystal structure (PDB ID: 2XFN), which contains the 2-BFI ligand positioned in the entrance cavity (D1). Two main systems were constructed: the first included only dopamine in the active site, and the second included dopamine in the active site and 2-BFI in the D1 cavity.

To identify the most probable catalytic conformation of dopamine for the hydride transfer reaction, frames from a 10-μs MD simulation of the dopamine-only system were analyzed. Frames were selected where the distance between the alpha carbon of dopamine and the nitrogen atom of the FAD isoalloxazine ring was within 5 Å. These structures were clustered into three groups using the k-means algorithm. Then, a representative structure from the most abundant cluster was selected for QM/MM study. Additionally, a structure similar to the input described by Mavri et al. was chosen from our simulations for subsequent ONIOM QM/MM calculations.

The ONIOM method was applied to describe the system at two levels of theory: a high-level QM layer for the reactive region and a low-level MM layer for the surrounding protein and solvent environment. The QM layer, which described the electronic structure of the active site, included neutral dopamine, the isoalloxazine ring of FAD, and key residues Y188, Y398 and Y435. The B3LYP DFT functional was employed with the 6-31G(d) basis set, providing an appropriate balance between accuracy and computational cost. All calculations were carried out with the Gaussian16 program package.[249] To improve the treatment of non-covalent interactions such as dispersion forces, Grimme's empirical dispersion correction (GD3BJ) was applied.[250] The MM layer consisted of the remaining protein residues

and solvent molecules, modeled using the AMBER ff14SB force field for the protein and the general AMBER force field (GAFF2) for the ligands.

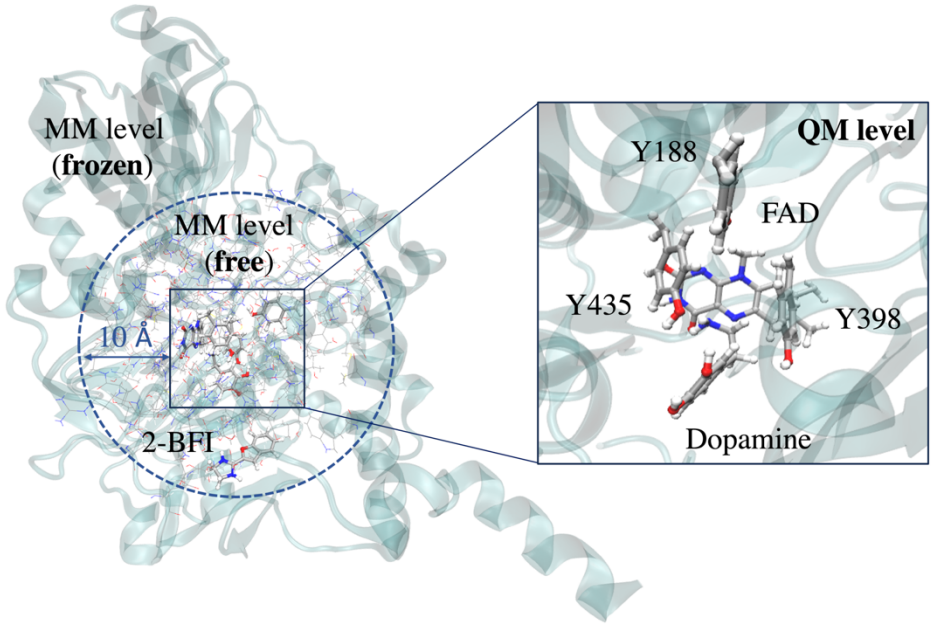


Figure 6.5: The system prepared for QM/MM simulations includes: the QM layer (FAD, Y435, Y398, Y188, and dopamine), the free MM layer (2-BFI, MM residues, and water molecules within 10 Å of the QM layer), and the remaining frozen MM layer.

To optimize computational efficiency while maintaining accuracy, the MM region within 10 Å of the QM layer was kept flexible, allowing atoms in this region to relax freely during optimizations. The remainder of the MM system was frozen to reduce computational cost. For the second system, the 2-BFI ligand, located in the D1 cavity, was included in the flexible MM region to evaluate its influence on the catalytic environment.

The geometry of the system was optimized in two stages to ensure accurate convergence. Initially, mechanical embedding was used, where the electrostatic impact of the MM layer on the QM calculation was neglected to allow for rapid optimization. In the second stage, electrostatic embedding explicitly incorporated

interactions between the QM and MM layers, allowing partial charges from the MM layer to influence QM calculations. This approach provided a more realistic description of the system's electrostatics by allowing the account for the differential stabilization that environment electrostatics introduce in the reactant, TS and product states, and hence in the reaction profile.

To investigate the hydride transfer from dopamine to the FAD isoalloxazine ring, a reaction coordinate scan was performed on the optimized system. The bond distance between the hydrogen atom of dopamine's α-carbon and the nitrogen atom of FAD was systematically reduced in 0.05 Å steps, generating intermediate structures along the reaction pathway. The structure with the highest energy in the scan was selected for TS optimization. The TS was identified as a geometry with a single imaginary frequency in the vibrational spectrum, corresponding to hydrogen movement along the reaction coordinate, confirming it as a saddle point on the potential energy surface. To further characterize the reaction, an IRC calculation was performed starting from the optimized TS. The IRC traced the minimum energy pathway in both forward and reverse directions, linking the TS to the reactant and product states. This provided a continuous trajectory of the reaction, revealing structural transitions during the hydride transfer. The reactant and product structures determined from the IRC were subsequently re-optimized with electrostatic embedding. Subsequently, single-point energy calculations were performed on the optimized reactant, product, and TS structures using both the B3LYP and M06-2X functionals with the 6-311++G(2d,2p) basis set to ensure result robustness. Thermal corrections from vibrational frequency calculations were applied to determine the reaction energy profile, including the activation energy and the overall reaction energy.

6.3 Investigating ligand interactions through MD simulations

In Chapter 5, we identified a second binding site (D2) alongside the entrance cavity (D1), previously recognized as a 2-BFI binding site in crystallographic data. Observed molecular movements between these cavities, particularly toward the aromatic cavity, suggested potential entrance and exit pathways for I2 ligands and dopamine. To investigate whether I2 ligands could interfere with the normal pathways of the substrate or metabolites, we conducted 10-μs MD simulations (10×1 μs replicas) for nine systems, placing 2-BFI, dopamine, and DOPAL in the active site (AS), entrance cavity (D1), and the newly identified D2 cavity. By aligning these simulations, we visualized molecular movements and preferences, showing their distances of center of mass to the N5 atom of FAD and the α-carbon of Phe103 using kernel density graphs. Yellow dense regions in the plots highlight where molecules spent the most time during the simulations.

The results revealed that 2-BFI tends to migrate toward the aromatic cavity, consistent with previous observations, while this behavior was not seen in dopamine and DOPAL simulations. Instead, dopamine and DOPAL showed transitions between cavities. Notably, 2-BFI did not enter the active site from the D1 cavity. Its increased density in the active site likely results from restricted movement after being placed there, rather than active migration through the D1 cavity (Figure 6.6). On the other hand, dopamine exhibits two distinct regions of high binding affinity within the active site, indicating its potential for two distinct binding modes. Both dopamine and DOPAL show comparable binding affinities in the D1 cavity, with their affinities surpassing those observed for 2-BFI. Moreover, DOPAL demonstrates a more uniform distribution across the three cavities (D1, D2, and AS).

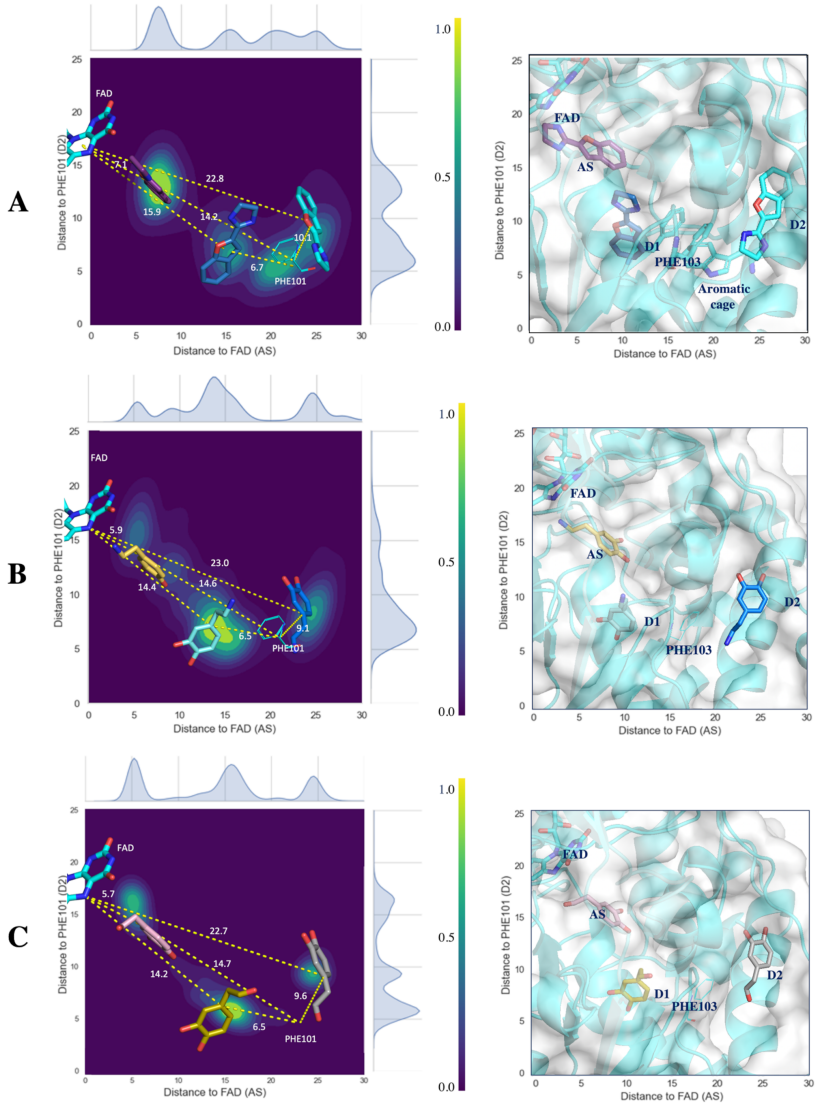
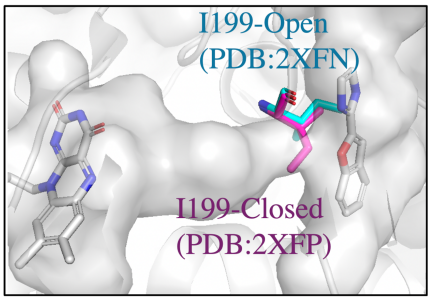


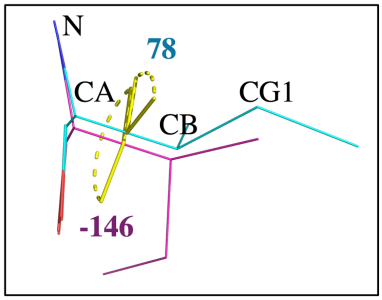
Figure 6.6: Kernel density graphs for (A) 2-BFI, (B) dopamine, and (C) DOPAL showing their distances of center of mass to FAD (N5) and Phe103 (C_α). The graphs highlight the ligands' preferred positions between the D2 and AS cavities, based on combined simulations starting from different positions (30 μs per ligand).

To investigate whether the conformation of I199 affects ligand movement, we analyzed the dihedral angle (N-CA-CB-CG1) of I199 across trajectories for two distinct systems: one starting with an open conformation (PDB: 2XFN) characterized by a dihedral angle of 78°, and the other with a closed conformation (PDB: 2XFP) with a dihedral angle of -148°, both accommodating the 2-BFI ligand in AS, D1, and D2 positions. The percentage of frames with open and closed ILE199 dihedral angles within a ±30° range is illustrated in Figure 6.7.

A



B



C

Trajectory with the initial location of 2-BFI in two crystal str. of MAO-B	(%) Frames with closed ILE dihedral -148° (+/- 30)	(%) Frames with open ILE dihedral 78 ° (+/- 30)
Open-BFI-AS	48.4	40.8
Open-BFI-D1	63.1	23.1
Open-BFI-D2	74.3	19.2
Closed-BFI-AS	70	7.6
Closed-BFI-D1	57	37.3
Closed-BFI-D2	71.8	24.9

Figure 6.7: (A) Starting conformations of MAO-B: open (PDB: 2XFN) and closed (PDB: 2XFP). (B) Dihedral angles of I199 (N-CA-CB-CG1). (C) Percentage of frames with open (78° ± 30°) and closed (-148° ± 30°) ILE199 dihedrals during simulations of MAO-B with 2-BFI in AS, D1, and D2 positions,

Regardless of the starting conformation, the closed state consistently emerged as the dominant form during simulations. Trajectories initiated with the open

conformation showed a marked shift toward the closed state, particularly when 2-BFI was present in D1 (63.1%) or D2 (74.3%) positions. Systems starting with the closed conformation predominantly retained this state, especially when 2-BFI started in AS (70%) or D2 (71.8%), with minimal transitions to the open conformation. These findings suggest that the closed conformation is energetically favored for 2-BFI binding, underscoring its stability and the critical role of I199's flexibility in ligand stabilization and MAO-B function.

The occupancies of the molecules are calculated as the percentage of frames in which ligands are found within each site (Figure 6.8). Each ligand system was initially placed in one of these pockets and subjected to MD simulations for a total of 10 μs across 10 replicas using the 2XFN crystal structure of MAO-B (with open I199 conformation). Although 2-BFI and dopamine are expected to be positively charged under physiological conditions, we also performed simulations with their neutral forms to account for the possibility of deprotonation occurring inside the protein. The colored boxes represent the percentage of frames during which ligands remained in their initial pocket throughout the trajectories. Protein backbones were aligned and ligand positions were then determined using their center of masses distance from their initial positions, AS (0–7 Å), D1 (0–6 Å), and D2 (0–7 Å). The connections between boxes in figure 6.8 indicate transitions of ligands between pockets. Additionally, the occupancy of the aromatic cavity (ArC) was calculated for molecules that transitioned from D1 or D2 to assess their affinity for this intermediate site.

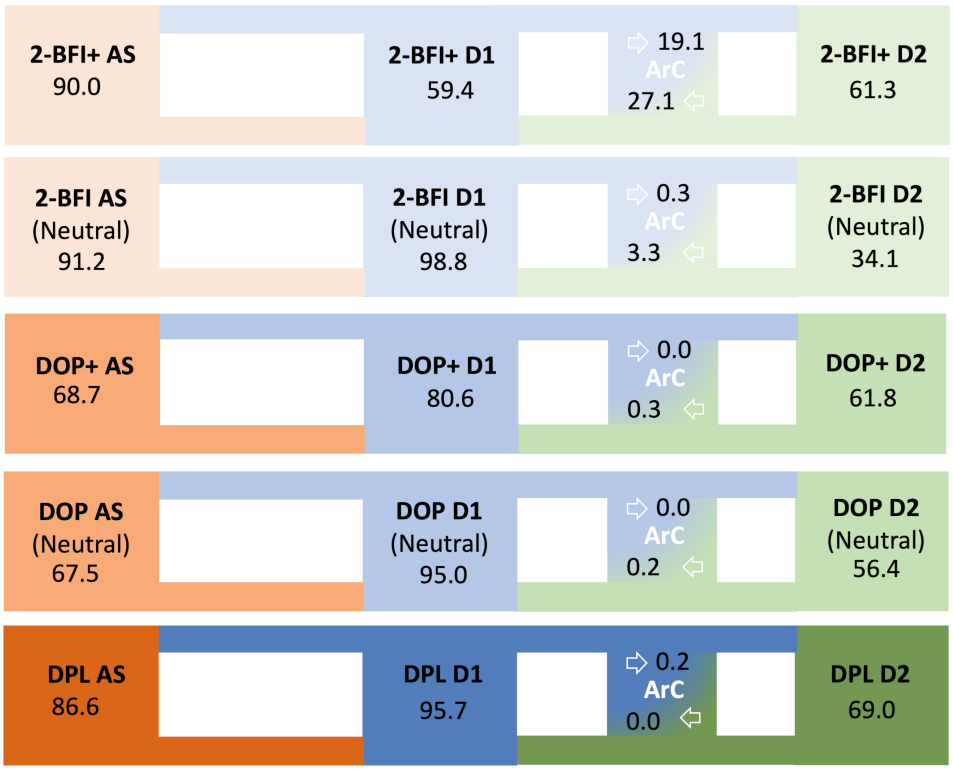


Figure 6.8: Ligand occupancies in the Active Site (AS), Entrance Cavity (D1), and Secondary Site (D2) during 10-μs simulations with 10 replicas for each system. Colored boxes show the percentage of frames ligands remained in their initial pocket. Arrows indicate transitions from D1 and D2 to aromatic cavity (ArC). The data is based on the systems with 2XFN crytal structure of MAO-B displaying an open conformation.

In the AS cavity, charged 2-BFI shows strong retention, with occupancies of 90.0% (99.9% in 2XFP, the closed I199 conformation), while the neutral form is similarly stable at 91.2%. Charged dopamine (DOP) exhibits differing occupancies between the two crystal structures, with 68.7% occupancy in 2XFN and 92.8% in 2XFP, while the neutral form remains slightly lower at 67.5%. Neutral DPL displays higher stability in the AS cavity, with 86.6% occupancy (73.9% in 2XFP).

In the D1 cavity, neutral 2-BFI demonstrates the highest stability, with an occupancy of 98.8%, while charged 2-BFI shows lower retention at 59.4% (25.0%

in 2XFP), due to increased mobility toward the aromatic cavity. Neutral DOP exhibits strong occupancy (95.0%), whereas the charged form is moderately stable at 80.6% (70.1% in 2XFP). DPL retains the highest stability in this cavity, with occupancies of 95.7% in 2XFN (98.9% in 2XFP).

In the D2 cavity, charged 2-BFI shows moderate occupancy at 61.3% in 2XFN (68.4% in 2XFP), while the neutral form exhibits weaker retention at 34.1%. Charged DOP displays moderate stability at 61.8%, which increases significantly to 93.2% in 2XFP. Neutral DOP shows slightly lower occupancy at 56.4%, while neutral DPL demonstrates the highest stability in D2 at 69.0% in 2XFN, though its retention drops substantially in 2XFP (28.5%).

The aromatic cavity (ArC) acts as a dynamic site, particularly for charged ligands. Charged 2-BFI frequently transitions to ArC, with 27.1% transitions in 2XFN and 28.8% in 2XFP, due to strong π-cation interactions that trap it within this cavity. Neutral 2-BFI, on the other hand, shows minimal interaction with ArC, at just 3.3%. Charged DOP rarely transitions to ArC, favoring stability in D1 and D2, while neutral DOP shows similar behavior. DPL also rarely interacts with ArC in either conformation, reflecting its strong retention in D1 and D2.

Overall, the retention and dynamics of ligands in various cavities demonstrate a highly context-dependent behavior influenced by ligand charge, cavity characteristics, and protein structural conformation. Charged and neutral 2-BFI exhibit exceptional stability in the AS cavity, largely due to being trapped within it. The charged form further demonstrates dynamic transitions to the aromatic cavity (ArC) driven by strong π-cation interactions with F103, which previously suggested to be the gate, highlighting ArC as a unique site for 2-BFI trapping, which might be potentially blocking the path of the natural substrate, or causing conformational changes of the residues. In contrast, dopamine and DPL show a clear preference for retention in the D1 and D2 cavities, with neutral forms of dopamine generally displaying greater stability than its charged counterparts in D1. Notably, charged dopamine exhibits significantly enhanced stability in the

closed I199 conformation of 2XFP compared to 2XFN, emphasizing the critical role of structural dynamics in influencing ligand behavior. The differential stability of dopamine across these conformations underscores the importance of structural flexibility in modulating binding interactions. Meanwhile, DPL's consistent stability in D1 and D2 suggests strong binding affinity, potentially aiding in the export of the product from the protein via D2.

Double ligand systems

To better understand how the presence of one ligand affects the affinity of another, we constructed 11 double-ligand systems and performed a total of 10-μs simulations with 10 replicas per system, resulting in a cumulative simulation time of 110 μs. Table 6.1 presents the occupancy percentages of ligands in each binding pocket across these simulations. Ligands were filtered based on the distance values of their center of mass to the initial structure to determine their positions: molecules in the active site (AS) were filtered between 0-7 Å, in D1 between 0-6 Å, and in D2 between 0-7 Å. Colored entries (orange for AS, blue for D1, green for D2) represent the primary filtering for the specified ligand in each system. Grey-shaded entries indicate secondary filtering, applied to measure the occupancy of the second ligand within the trajectory of the first. This approach highlights frames where both ligands are simultaneously stable in their respective pockets. For example, in System 1, the blue box shows the percentage of frames where BFI remains in D1, while the grey-shaded DOP value indicates the stability of DOP during this time. Similarly, the green box shows DOP's occupancy in D2, and the grey-shaded BFI value reflects BFI's presence in D1 during DOP's stability in D2.

Table 6.1: Ligand occupancy percentages in AS (0–7 Å), D1 (0–6 Å), and D2 (0–7 Å) for double-ligand systems, showing how one ligand affects the other's stability.

Double Ligand systems	AS	D1	D2
System 1		**BFI+**: 50.88%	DOP+: 97.43%
		BFI+: 60.57%	**DOP+**: 82.57%
System 2		**DOP+**: 91.09%	BFI+: 82.13%
		DOP+: 97.95%	**BFI+**: 79.21%
System 3	**BFI+**: 90.72%	DOP+: 53.42%	
	BFI+: 84.15%	**DOP+**: 57.49%	
System 4	**BFI+**: 99.96%		DOP+: 73.35%
	BFI+: 99.71%		**DOP+**: 74.37%
System 5	**DOP+:** 84.92%	BFI+**:** 62.29%	
	DOP+: 94.17%	**BFI+:** 55.09%	
System 6	**DOP+:** 90.75%		BFI+**:** 73.93%
	DOP+**:** 92.32%		**BFI+:** 69.53%
System 7	**DOPneut:**99.46%	BFI+**:** 80.40%	
	DOPneut**:** 99.50%	**BFI+:** 80.40%	
System 8	**DOPneut:**59.60%		BFI+**:** 23.64%
	DOPneut**:** 100.0%		**BFI+:** 34.88%
System 9	**DOPneut:**98.46%	BFIneut**:**99.21%	
	DOPneut**:** 96.07%	**BFIneut:**99.14%	
System 10	**DOPneut:**68.42%		BFIneut**:** 32.66%
	DOPneut**:** 66.55%		**BFIneut:**33.63%
System 11	**DPL:** 92.82%	BFI+: 76.7 %	
	DPL: 97.34%	**BFI+:** 73.09%	

In System 1 (DOP+ in D2 and 2-BFI+ in D1), DOP+ shows increased occupancy in D2 compared to its single-ligand system, while 2-BFI+ maintains similar occupancy in D1, indicating stabilizing interactions between the two ligands. In System 2 (DOP+ in D1 and 2-BFI+ in D2), both DOP+ and 2-BFI+ exhibit increased occupancies compared to their single-ligand systems, suggesting mutual stabilization. In System 3 (DOP+ in D1 and 2-BFI+ in AS), the presence of 2-BFI+ in AS slightly reduces DOP+'s stability in D1. System 4 (DOP+ in D2 and 2-BFI+ in AS) demonstrates enhanced stability for both ligands, with DOP+ in D2 and 2-BFI+ in AS showing higher occupancies than in their single systems. In System 5

(DOP+ in AS and 2-BFI+ in D1), DOP+ shows increased occupancy in AS, while 2-BFI+ in D1 retains similar stability as in the single-ligand system. Similarly, in System 6 (DOP+ in AS and 2-BFI+ in D2), DOP+ occupancy increases in AS, while 2-BFI+ in D2 remains stable. System 7 (DOP neutral in AS and 2-BFI+ in D1) exhibits enhanced stability for both ligands, with DOP neutral and 2-BFI+ showing higher occupancies, representing a likely protonation scenario during the catalytic reaction. In System 8 (DOP neutral in AS and 2-BFI+ in D2), DOP neutral maintains its occupancy, but 2-BFI+ shows reduced stability in D2, likely due to transitions toward D1. In System 9 (DOP neutral in AS and 2-BFI neutral in D1), DOP neutral displays increased stability in AS, while 2-BFI neutral retains its stability in D1. System 10 (DOP neutral in AS and 2-BFI neutral in D2) shows similar occupancy for DOP neutral but decreased stability for 2-BFI neutral, likely due to transitions toward D1 as in System 8. Finally, in System 11 (DOPAL in AS and 2-BFI+ in D1), DOPAL retains its stability in AS, while 2-BFI+ shows higher occupancy in D1, suggesting stabilizing interactions between the two ligands.

Overall, the presence of a second ligand significantly influences the stability of ligands across binding pockets. Charged ligands, such as 2-BFI+ and DOP+, often show enhanced stability in their respective pockets when paired with another ligand, especially in systems where both ligands occupy complementary cavities (e.g., Systems 2 and 4). Neutral ligands, such as DOP neutral and 2-BFI neutral, exhibit strong retention in the D1 cavity, with minimal destabilization from the second ligand. As previously suggested, 2-BFI exhibited higher binding affinity in the presence of another inhibitor. The simulations further support that ligand affinity increases when paired with another ligand.

In the next stage, we applied the QM/MM methodology to the selected systems with dopamine, both with and without 2-BFI, to investigate the secondary ligand's effect on the reaction profile. We began by focusing on System 7 (neutral dopamine in the AS and protonated 2-BFI in D1). The hydride transfer mechanism

is expected to occur with neutral dopamine, and 2-BFI is assumed to be protonated in solution. This choice reflects the most likely state of these molecules together, especially considering that 2-BFI was also observed in the D1 cavity in the crystal structure. In future stages of this work, we plan to extend the study to other systems where charged and neutral 2-BFI are placed in D1, D2, and the aromatic cavity.

6.4 Determination of reaction barrier for dopamine hydride transfer with FAD using ONIOM QM/MM methods

This section aims to analyze the effect of 2-BFI on the activation barrier of dopamine's hydride transfer reaction with FAD. Initially, we investigated how the presence of a secondary ligand influences dopamine's position in the active site. After aligning and calculating the RMSD of the protein, dopamine was filtered based on its α-carbon distance to FAD (N5), with frames within the 0–4 Å range selected. The analysis (Figure 6.9) shows how the proximity of dopamine's alpha carbon to FAD varies across different systems. For neutral dopamine (DOPneu) in the solo system, 4.08% of frames show the alpha carbon within 4 Å of FAD, indicating catalytic suitability to the active site. In contrast, charged dopamine (DOP+) in the solo system shows only 1.29%.

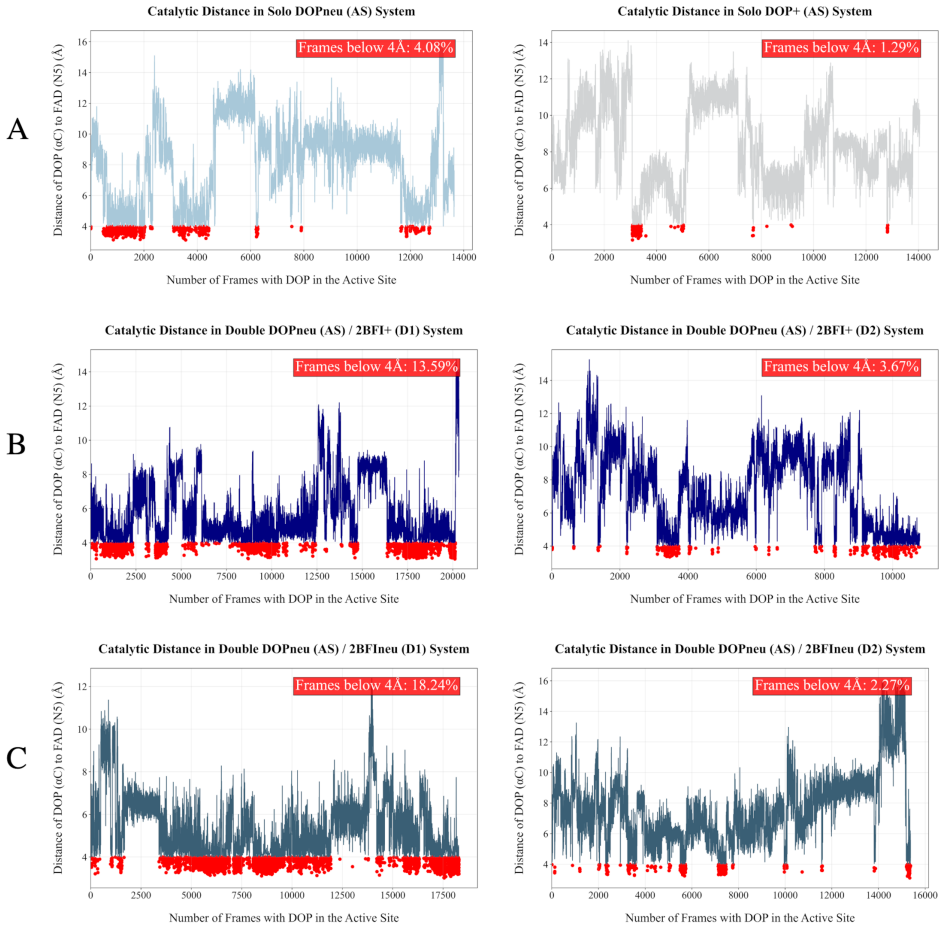


Figure 6.9: The catalytic distance of dopamine's α-carbon to FAD (N5) across various systems. (A) depicts neutral and charged dopamine in the active site (AS). (B) shows neutral dopamine with charged 2-BFI in the D1 and D2 cavity, and (C) presents neutral dopamine with neutral 2-BFI in the D1 and D2 cavity. The x-axis represents frames with dopamine in the active site, while the y-axis shows the distance to FAD. Red bars indicate frames where the distance is below 4 Å, with the percentage of these frames shown at the top of each plot.

When paired with 2-BFI (DOPneu (AS) + 2-BFI+ (D1)), neutral dopamine shows a substantial increase to 13.59%, suggesting enhanced stability and proximity to FAD. However, in the D2 cavity (DOPneu (AS) + 2-BFI+ (D2)), this percentage

drops to 3.67%, indicating that 2-BFI in D2 is less favorable for dopamine binding. Neutral 2-BFI (DOPneu (AS) + 2-BFIneu) in D1 increases dopamine stability further, with 18.24% of frames showing proximity to FAD. In D2 (DOPneu (AS) + 2-BFIneu (D2)), this percentage decreases to 2.27%. These results suggest that both the charge state of dopamine and the cavity location of 2-BFI influence dopamine's structural arrangement in AS, with the closest interaction with FAD occurring when neutral 2-BFI occupies the D1 cavity. The increased proximity of dopamine to FAD in the presence of 2-BFI raises the question of whether this enhanced stability results from dopamine adopting a specific conformation different from that observed in the single-ligand neutral dopamine systems, which further investigated by clustering of the dopamine conformations. Indeed, a closer arrangement does not necessarily translate into a more favorable catalytic configuration.

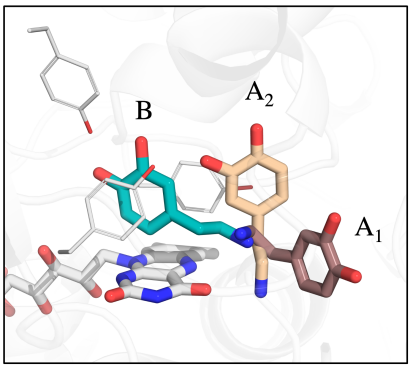


Figure 6.10: Representative dopamine conformations (the centroids) of K-means clustering, based on ligand RMSD within the filtered trajectory (0–5 Å from the α-carbon to FAD N5), were aligned to the protein backbone using only the neutral dopamine trajectory. Conformation A includes two subtypes: A1 (horizontal) and A2 (perpendicular) aromatic ring alignments relative to FAD. The initial input structure from Mavri et al. falls between clusters A1 and A2, while conformation B represents a novel arrangement, where the aromatic ring forms π-π interactions with Y398 and Y435.

To identify dopamine conformations for QM/MM studies, initially frames were filtered based on the α-carbon's distance to FAD (N5) in the trajectory of single

neutral dopamine system, selecting those within 0–5 Å for sufficient sampling.

Subsequently, K-means clustering with various cluster numbers was applied to the trajectory, revealing that dopamine conformations primarily fell into three categories (Figure6.10): parallel to FAD and interacting with site residues (centroid B), or facing FAD in distinct orientations (centroids A1 and A2). To capture these differences and the tendencies of the molecule in various systems, we clustered the conformations into three groups: (A1) parallel to FAD, (A2) vertical to FAD, and (B) parallel to FAD with aryl groups forming a π-π stacking interaction with Y98 and Y435. Notably, the initial input structure from Mavri et al. falls between clusters A1 and A2, while the novel conformation (B) has not been previously reported.

To investigate the effect of 2-BFI on dopamine's conformational preferences, we analyzed both single-ligand and double-ligand systems. Single-ligand systems included dopamine (neutral or protonated) with and without 2-BFI (neutral or protonated).

Table 6.2: The distribution (%) of dopamine conformations within 5 Å of the catalytic site under different dopamine protonation states, and the presence and location of a second ligand (2-BFI).

DOP p.state in AS	**Second ligand system**	**Second ligand location**	**Frames (%) within cat. distance <5** Å	**Conf. A1 (%)**	**Conf. A2 (%)**	**Conf. B (%)**
DOPneu	-	-	13.8	29.1	22.2	48.7
DOPneu	2-BFIneu	D1	50.7	4.1	35	60.9
DOPneu	2-BFI+	D1	47.9	31.8	58.1	10.1
DOPneu	2-BFI+	D2	12.5	7.7	52.2	40.1
DOP+	-	-	3.8	2.4	43.5	54.1
DOP+	2-BFI+	D1	23.5	100	-	-
DOP+	2-BFI+	D2	3.3	74.7	25.3	-

Table 6.2 summarizes the distribution of dopamine conformations near the catalytic site (<5 Å) under various conditions. Without a second ligand, neutral dopamine (DOPneu) shows limited proximity (13.8%). K-means clustering applied to this filtered trajectory reveals Conformation B as most frequent (48.7%), followed by A1 (29.1%) and A2 (22.2%). Neutral 2-BFI in the D1 cavity increases dopamine's proximity to 50.7%, with B remaining dominant 60.9% of the filtered trajectory, A2 rising to 35%, and A1 dropping to 4.1%. Protonated 2-BFI (BFI+) in D1 maintains high proximity (47.9%), shifting preferences to A1 (31.8%) and A2 (58.1%) as B significantly decreases (10.1%). Protonated 2-BFI in D2 reduces dopamine proximity to 12.5%, favoring A2 (52.2%) and B (40.1%), with minimal A1 (7.7%). Protonated dopamine (DOP+) without a second ligand shows very low proximity (3.8%), with B dominating (54.1%), followed by A2 (43.5%) and negligible A1 (2.4%). Adding 2-BFI+ in D1 increases proximity to 23.5%, where A1 becomes the sole conformation (100%). In contrast, 2-BFI+ in D2 reduces proximity to 3.3%, favoring A1 (74.7%) and A2 (25.3%), with no B observed.

This analysis highlights the impact of 2-BFI, particularly in its protonated form, on dopamine's proximity and conformational dynamics near the catalytic site. Without 2-BFI or with neutral 2-BFI, Conformation B dominates, while protonated 2-BFI (BFI+) shifts preferences towards A1 and A2. The presence of 2-BFI in the D1 cavity promotes closer catalytic proximity compared to D2, and the solvent-exposed nature of these cavities suggests 2-BFI is likely protonated. Protonated 2-BFI may stabilize dopamine near the catalytic site, increasing binding frequency and residence time. However, this stabilization could distort active site geometry or impose steric/electrostatic constraints, raising the reaction barrier and slowing catalytic progression. This may cause dopamine to adopt non-productive conformations, creating a kinetic bottleneck where the enzyme is trapped in an inefficient state. The effects of 2-BFI depend on its protonation state and location, with D1 offering greater catalytic relevance than D2.

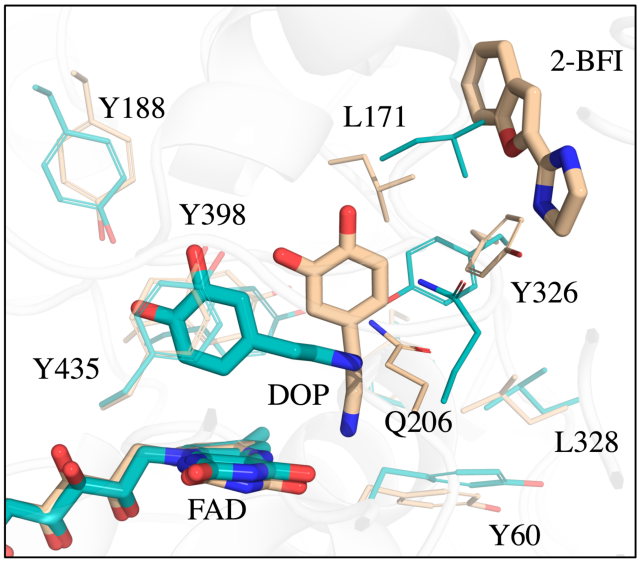


Figure 6.11: Superimposed representative dopamine conformations showing the most abundant Conformation B from the single-ligand neutral dopamine system (blue) and Conformation A2 from the double-ligand system with positively charged 2-BFI in the D1 cavity (beige), as determined by K-means clustering.

The superimposed structures indicate that while several surrounding residues, such as Y188, Y398, Y435, Y60, and L328, maintain a consistent orientation, others like Y326, L171, and Q206 exhibit distinct conformations depending on dopamine's position. Notably, in Conformation B, Y326 shifts by 90 degrees, coming into close proximity with Y398, with an O-O distance of approximately 2.7 Å.

At this stage, we applied the ONIOM QM/MM methodology to representative frames from the most abundant cluster B of the single dopamine system. Additionally, we selected a conformation from cluster A1 that closely resembled the input structure from the Mavri et al. study (Figure 6.4), which suggested that the R configuration of the hydrogen atom on the alpha carbon was more likely to be transferred to FAD. Ultimately, three representative structures were chosen: the

A-R conformer, representing the configuration described in Mavri et al., and two conformations from cluster B, representing B-R and B-S hydrogen configurations, selected from frames with the closest proximities to FAD. The same selection process was applied to the double-ligand systems with 2-BFI, ensuring alignment with the most similar structures from the single-ligand systems.

The optimization of these six systems was successfully completed using mechanical embedding, followed by electrostatic embedding. Freezing the MM layer beyond 10 Å from the QM layer significantly improved the efficiency and speed of the optimization process. From the optimized structures, scan calculations were performed by systematically bringing the hydrogen atom on dopamine's α-carbon closer to the FAD(N5) atom in increments of 0.05 Å per step.

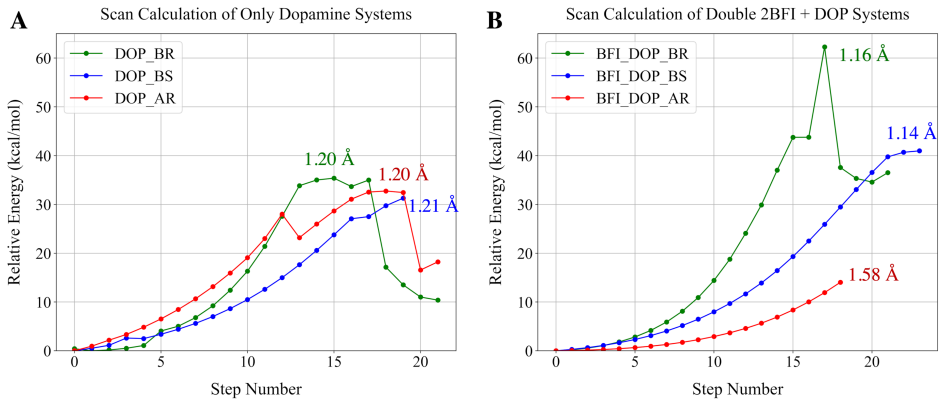


Figure 6.11: Scan calculations depicting the relative energy profiles for the approach of the α-carbon hydrogen atom of dopamine toward FAD(N5) in different systems computed at the B3LYP/6-31G(d) level of theory with GD3BJ empirical dispersion corrections. Panel A shows results for dopamine-only systems, while Panel B includes systems with a 2-BFI ligand alongside dopamine. The distance values indicate H-FAD(N5) distance when the energy is the highest.

The scan calculations identified the highest energy structures for single dopamine systems, which were subsequently used for transition state optimization. While the

transition states DOP-AR-TS and DOP-BR-TS were successfully optimized, the optimization of BS-TS was unsuccessful as the scan calculations crashed at later steps, preventing further optimization. For the double-ligand systems, the incomplete scan calculations were due to system crashes when dopamine reached a certain distance. Significantly higher relative energies were observed for the 2-BFI$^+$-DOP-BR and 2-BFI$^+$-DOP-BS systems. The structures with the highest energies were subjected to transition state optimization; however, these calculations did not converge. Furthermore, the scan calculation with the 2-BFI$^+$-DOP-AR system has crashed and did not permit optimization when the hydrogen atom approached closer than 1.58 Å to FAD(N5). Consequently, the transition state optimization for dopamine was not feasible in the presence of 2-BFI$^+$.

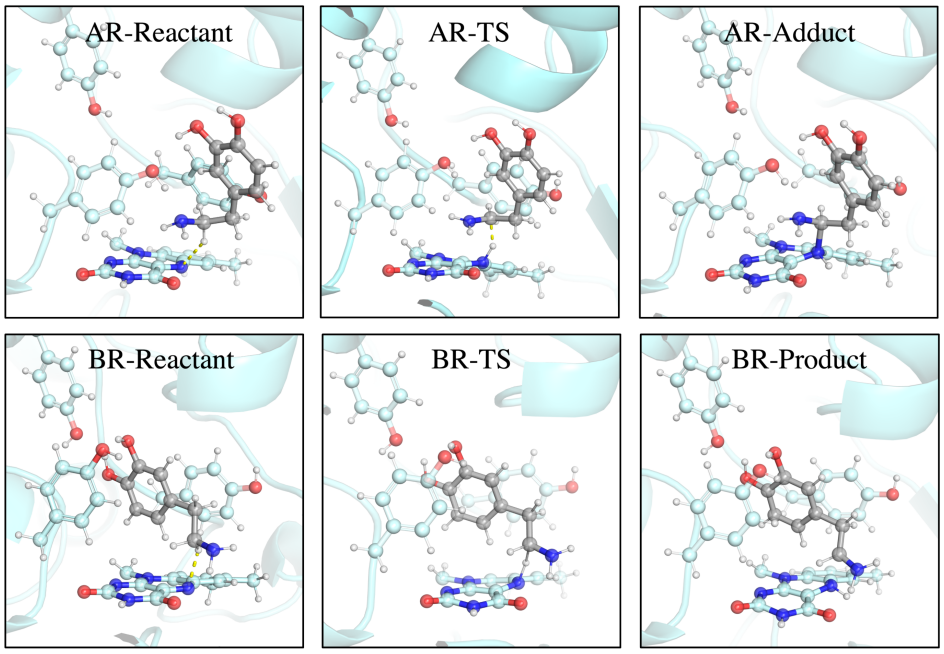


Figure 6.12: Optimized reactant, transition state, and product structures for single-ligand dopamine systems. AR follows a two-step hydride transfer, while BR undergoes direct hydride transfer, bypassing adduct formation.

The transition state optimization for single-ligand dopamine systems was successfully completed and confirmed with a negative vibrational frequency. IRC calculations were subsequently performed to trace the reaction pathway and identify the reactant and intermediate product structures. Figure 6.12 illustrates optimized reactant, transition state, and product for the single dopamine systems. Interestingly, the product from the BR system follows a direct hydride transfer mechanism, bypassing the adduct formation and hydrolysis with water required in the AR system, which follows a two-step hydride transfer. Consequently, the direct hydride transfer in the B conformation results in the release of a positively charged imine, which can later be converted to an aldehyde via nucleophilic attack by waters.

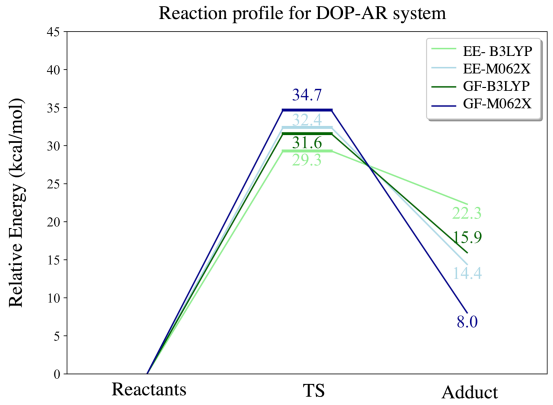


Figure 6.13: Activation and reaction energies of the DOP-AR structure using EE-B3LYP (light green) and EE-M062X (light blue) functionals for electronic energies, and GF-B3LYP (dark green) and GF-M062X (dark blue) for Gibbs free energies. Calculations include Empirical Dispersion GD3BJ at the 6-311++G(2d,2p) level.

DOP-AR reaction profile is illustrated in Figure 6.13. Thermal corrections were applied to the QM system to avoid inaccuracies related to protein vibrations. The graph shows electronic energies (light green and blue) and Gibbs free energies (dark green and blue) for the reactants, TS and adduct. Electronic activation

energies from B3LYP and M06-2X are 29.3 kcal/mol and 32.4 kcal/mol, respectively, while Gibbs free energy barriers are 31.6 kcal/mol (B3LYP) and 34.7 kcal/mol (M06-2X). All energies include empirical dispersion GD3BJ at the 6-311++G(2d,2p) level. M06-2X predicts a higher barrier but lower reaction energy compared to B3LYP. The activation energy for the DOP-BR system, based on the optimized structures with the B3LYP functional and 6-31G(d) basis set, is currently calculated to be 33 kcal/mol using the electrostatic energy, with the reaction energy estimated to be 18 kcal/mol. The single-point and Gibbs free energies remain to be calculated.

To summarize, our single dopamine simulations yielded results structurally consistent with the findings of Mavri et al., successfully optimizing the DOP-AR conformation and obtaining its TS and the adduct. While experimental studies suggest a reaction barrier of approximately 16.5 kcal/mol for the hydride transfer reaction of dopamine,[247] our calculations resulted in a higher barrier (around 30 kcal/mol). This discrepancy can be attributed to the absence of explicit water molecules and fewer residues in the QM layer, a limitation imposed by computational resource constraints. Additionally, the chosen free MM layer within a 10 Å cutoff may have restricted the flexibility needed for better TS and product optimization, which would also require a larger computational effort.

Interestingly, the second conformation, B, which has not been previously described in hydride transfer studies, suggests a comparable energy barrier for its optimized transition state. This pathway bypassed the formation of the adduct, which was previously suggested to require at least four water molecules for stabilization and proceeded directly to the imine form of dopamine in a single step.

In the presence of 2-BFI, our results indicate a significant shift in dopamine's conformational preferences. While dopamine alone predominantly adopts the B conformation, which facilitates a direct transfer mechanism, the presence of 2-BFI

in the D1 cavity promotes A conformations, favoring the two-step hydride transfer mechanism. Despite bringing dopamine closer to FAD, as shown in the previous sections, the double-ligand systems with 2-BFI failed to yield a transition state, suggesting that 2-BFI may slow down the reaction and prevent catalytic activity. Additionally, 2-BFI, by altering dopamine's preferred conformation to A, may influence the reaction mechanism, potentially favoring the two-step hydride transfer pathway. However, our calculations did not yield transition state structures for dopamine in the presence of 2-BFI, as the scan calculations could not reach convergence when dopamine approached FAD near the TS region. These results need to be nevertheless validated and completed determining the TS from additional frames from the MD simulations, and also testing the impact of 2-BFI in the reaction when it is placed in the D2 and ArC cavities. Furthermore, the kinetic implications of these mechanistic shifts require further investigation through both computational and experimental studies.

CHAPTER 7. CONCLUSIONS

This thesis provides a detailed investigation into the dynamic properties of protein-ligand interactions, enzymatic catalysis, and photosynthetic light harvesting, employing advanced multiscale computational methodologies. Through three distinct approaches across diverse biological systems we uncover new insights into how these biological phenomena function at a molecular level.

In Chapter 4, we investigated the role of thermal effects on the spectral tuning mechanisms of PC645 and PC612 antenna complexes from cryptophyte algae. To capture the thermal effects arising from protein (slow) and pigment (fast) structural fluctuations, we applied a dual MD strategy combining classical MD simulations on the μs timescale with Born-Oppenheimer QM/MM simulations on the ps timescale. This approach enabled us to accurately describe the impact of pigment vibrations on excited states. We also explored an alternative method involving multiple QM/MM geometry optimizations along classical MD trajectories. The resulting structures from both methods were used to calculate excited states using polarizable QM/MM TD-DFT and simulate spectral properties with the FCE formalism. Our findings revealed that thermal effects only slightly influenced the energy ladder of the flexible PC612 complex, characterized by its open quaternary structure. In contrast, thermal dynamics significantly impacted the spectral properties of the closed PC645 complex due to the absence of a stabilizing hydrogen bond in PCB β82 bilins and the greater sensitivity of MBV pigments, which possess higher conjugation lengths. Overall, the dual MD approach combined with FCE spectral simulations yielded highly accurate spectral properties for PC612 and PC645, paving the way for future investigations into the influence of vibronic coherence and open-closed quaternary structures on excitation energy transfer mechanisms in light-harvesting complexes.

In Chapter 5, we focused on exploring Imidazoline type 2 receptor (I2-IR) binding sites in putative I2-IR receptor proteins, using the canonical ligand 2-BFI as case study. To achieve this, we developed and applied a hierarchical computational

protocol to systematically identify and characterize high-affinity binding pockets for 2-BFI across seventeen proteins derived from proteomics studies in a senescent mouse model. The protocol employed a combination of MDmix simulations, molecular docking, and molecular dynamics (MD) simulations to refine and validate potential binding hotspots. MDmix simulations were used to identify solvent hotspots that aligned with the pharmacophoric features of 2-BFI, such as hydrogen bonding and aromatic interactions. These hotspots were further filtered and employed as constraints in molecular docking calculations using rDock, effectively narrowing down the potential binding pockets. The most promising docked poses were then subjected to MD simulations to evaluate their structural stability and interaction patterns, ensuring the robustness of the identified binding sites. This approach led to the successful identification of high-confidence binding sites for 2-BFI in six proteins, namely Monoamine Oxidase B (MAO-B), Alpha-2A Adrenergic Receptor, S-formylglutathione Hydrolase, Hydroxymethylglutaryl-CoA Lyase (HMG-CoA Lyase), Heterogeneous Nuclear Ribonucleoprotein U, and Nischarin. Notably, Nischarin exhibited two distinct binding sites. By systematically exploring putative I2-IR receptors and characterizing their interaction sites, this study lays the groundwork for the rational design of I2 ligands as therapeutic agents for neurodegenerative diseases.

In Chapter 6, we delved into the dynamic behavior of dopamine and 2-BFI within monoamine oxidase B (MAO-B), emphasizing ligand transitions and their influence on the enzyme's catalytic activity. Through molecular dynamics (MD) simulations, we identified distinct preferences for ligand occupations across key binding cavities, including the active site (AS), D1, D2, and the aromatic cavity (ArC). Although 2-BFI and dopamine are typically positively charged, we included neutral forms in our simulations to account for potential deprotonation inside the protein. Positively charged ligands like 2-BFI+ exhibited significant mobility, particularly transitioning to the aromatic cavity from D1 (19.1%) and D2 (27.1%), where π-cation interactions appeared to play a critical role in trapping the

ligand. In contrast, neutral ligands such as neutral BFI, dopamine, and DPL demonstrated greater stability, particularly in the D1 cavity, with minimal transitions to the aromatic cavity. Charged molecules demonstrated greater stability within the D2 cavity, probably due to its proximity to the membrane-embedded tail of MAO-B. This region is likely to capture charged molecules from the negatively charged membrane surface and facilitate their transfer to inward cavities such as D1 and the active site (AS). This analysis also revealed differences between MAO-B systems based on PDB structures 2XFN and 2XFP, corresponding to open and closed I199 conformations, respectively. In 2XFP, charged BFI demonstrated transitioning more frequently to the aromatic cavity (28.8%) and D2 (42.9%), whereas no significant transition to aromatic cavity is observed for dopamine and dopal. Considering F103 as the gate, as previously described, the specific π-cation interaction of 2-BFI with the residue may block pathways differently compared to other substrates. These dynamics suggest a higher degree of ligand mobility and redistribution across cavities in systems with a closed conformation.

The findings from these transitions were complemented by ONIOM QM/MM methodologies to investigate the hydride transfer mechanism of dopamine oxidation by MAO-B and allosteric 2-BFI effect on the reaction profile. Dopamine predominantly adopted the B conformation in single-ligand systems, facilitating the direct hydride transfer pathway. However, the presence of 2-BFI in the D1 cavity shifted dopamine's preference toward the A conformation, which favors a two-step hydride transfer mechanism. This shift in mechanism, coupled with the redistribution of dopamine across binding pockets, highlights the allosteric modulation exerted by 2-BFI. Interestingly, despite the stabilizing effects of 2-BFI on dopamine's proximity to the FAD cofactor, the double-ligand systems failed to yield a transition state, indicating a potential slowing of the reaction. To further explore the effects of 2-BFI on the reaction profile of dopamine, additional frames should be incorporated to enhance the accuracy of the analysis. Furthermore, this

method should be extended to other double-ligand systems where 2-BFI is positioned in both the D2 cavity and the aromatic cavity (ArC) to provide a comprehensive understanding of its allosteric modulation.

BIBLIOGRAPHY

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* vol. 450 964–972 P (2007).
2. Zwier, M. C. & Chong, L. T. Reaching biological timescales with all-atom molecular dynamics simulations. *Current Opinion in Pharmacology* vol. 10 745–752 (2010).
3. Brunk, E. & Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chemical Reviews* vol. 115 6217–6263 (2015).
4. Adcock, S. A. & McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews* vol. 106 1589–1615 (2006).
5. Chergui, M. Ultrafast structural dynamics of biological systems. in *Comprehensive Biophysics* vol. 1 398–424 (Elsevier Inc., 2012).
6. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O. & Shaw, D. E. Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* vol. 19 120–127 (2009).
7. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *J Am Chem Soc* **140**, 2386–2396 (2018).
8. Faradjian, A. K. & Elber, R. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics* **120**, 10880–10889 (2004).
9. Hansmann, U. H. E. *Parallel Tempering Algorithm for Conformational Studies of Biological Molecules*. *Chemical Physics Letters* vol. 281 (1997).
10. Laio, A. & Parrinello, M. *Escaping Free-Energy Minima*, PNAS, (2002)
11. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *Journal of Chemical Physics* **120**, 11919–11929 (2004).
12. Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling Author links open overlay panel. *J Comput Phys* **23**, 187–199 (1977).
13. Kroese, D. P., Brereton, T., Taimre, T. & Botev, Z. I. Why the Monte Carlo method is so important today. *Wiley Interdiscip Rev Comput Stat* **6**, 386–392 (2014).
14. Levitt, M. & Warshel, A. Computer simulation of protein folding. *Nature* **253**, 694–698 (1975).
15. Levitt, M. Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). *Angewandte Chemie - International Edition* **53**, 10006–10018 (2014).
16. Duncan, A. L. & Pezeshkian, W. Mesoscale simulations: An indispensable approach to understand biomembranes. *Biophysical Journal* vol. 122 1883–1889 (2023).
17. Mirkovic, T. *et al*. Light absorption and energy transfer in the antenna complexes of photosynthetic organisms. *Chemical Reviews* vol. 117 249–293 (2017).
18. Simkin, A. J. *et al*. The role of photosynthesis related pigments in light harvesting, photoprotection and enhancement of photosynthetic yield in planta. *Photosynthesis Research* vol. 152 23–42 (2022).

19. Rabinowitch, E. & Govindjee. Photosynthesis. *American Association for the Advancement of Science* (1969).
20. Curutchet, C. & Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem Rev* **117**, 294–343 (2017).
21. Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F. & Buchleitner, A. Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ Sci* **5**, 9374 (2012).
22. Richardson, T. L. & Campbell, L. The colorful world of cryptophyte phycobiliproteins. *Journal of Plankton Research* vol. 44 806–818 (2022).
23. Kieselbach, T., Cheregi, O., Green, B. R. & Funk, C. Proteomic analysis of the phycobiliprotein antenna of the cryptophyte alga Guillardia theta cultured under different light intensities. *Photosynth Res* **135**, 149–163 (2018).
24. Doust, A. B., Wilk, K. E., Curmi, P. M. G. & Scholes, G. D. The photophysics of cryptophyte light-harvesting. *J Photochem Photobiol A Chem* **184**, 1–17 (2006).
25. Collini, E. *et al*. Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature* **463**, 644–647 (2010).
26. Mirkovic, T. *et al*. Light Absorption and Energy Transfer in the Antenna Complexes of Photosynthetic Organisms. *Chem Rev* **117**, 249–293 (2017).
27. Doust, A. B. *et al*. Developing a Structure–Function Model for the Cryptophyte Phycoerythrin 545 Using Ultrahigh Resolution Crystallography and Ultrafast Laser Spectroscopy. *J Mol Biol* **344**, 135–153 (2004).
28. Harrop, S. J. *et al*. Single-residue insertion switches the quaternary structure and exciton states of cryptophyte light-harvesting proteins. *Proceedings of the National Academy of Sciences* **111**, (2014).
29. Corbella, M., Cupellini, L., Lipparini, F., Scholes, G. D. & Curutchet, C. Spectral Variability in Phycocyanin Cryptophyte Antenna Complexes is Controlled by Changes in the α-Polypeptide Chains. *ChemPhotoChem* **3**, 945–956 (2019).
30. Michie, K. A. *et al*. Molecular structures reveal the origin of spectral variation in cryptophyte light harvesting antenna proteins. *Protein Science* **32**, (2023).
31. Curti, M. *et al*. Engineering excitonically coupled dimers in an artificial protein for light harvesting via computational modeling. *Protein Science* **32**, (2023).
32. Moya, R., Norris, A. C., Spangler, L. C., Scholes, G. D. & Schlau-Cohen, G. S. Observation of conformational dynamics in single light-harvesting proteins from cryptophyte algae. *J Chem Phys* **157**, (2022).
33. Doust, A. B. *et al*. Mediation of ultrafast light-harvesting by a central dimer in phycoerythrin 545 studied by transient absorption and global analysis. *Journal of Physical Chemistry B* **109**, 14219–14226 (2005).
34. Marin, A. *et al*. Flow of excitation energy in the cryptophyte light-harvesting antenna phycocyanin 645. *Biophys J* **101**, 1004–1013 (2011).
35. Corbella, M., Toa, Z. S. D., Scholes, G. D., Luque, F. J. & Curutchet, C. Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes. *Physical Chemistry Chemical Physics* **20**, 21404–21416 (2018).
36. Nussinov, R. & Tsai, C. J. Allostery in disease and in drug discovery. *Cell* vol. 153 293–305 (2013).
37. Du, X. *et al*. Insights into protein–ligand interactions: Mechanisms, models, and methods. *International Journal of Molecular Sciences* vol. 17 (2016).
38. Atkins, P. & De Paula. *Physical Chemistry for the Life Sciences*. (Oxford University Press, 2006).

39. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J Med Chem* **53**, 5061–5084 (2010).
40. Li, J. X. Imidazoline I2 receptors: An update. *Pharmacology and Therapeutics* vol. 178 48–56 (2017).
41. Gomez, R. E., Ernsberger, P., Feinland, G. & Reis, D. J. Rilmenidine lowers arterial pressure via imidazole receptors in brainstem C1 area. *Eur J Pharmacol* **195**, 181–191 (1991).
42. MORGAN, N. G. *et al*. Comparative Effects of Efaroxan and beta-carbolines on the Secretory Activity of Rodent and Human beta Cells. *Ann N Y Acad Sci* **1009**, 167–174 (2003).
43. Abás, S. *et al*. Neuroprotective Effects of a Structurally New Family of High Affinity Imidazoline I2 Receptor Ligands. *ACS Chem Neurosci* **8**, 737–742 (2017).
44. García-Sevilla, J. A., Escribá, P. V. & Guimón, J. Imidazoline receptors and human brain disorders. in *Annals of the New York Academy of Sciences* vol. 881 392–409 (New York Academy of Sciences, 1999).
45. Bagán, A. *et al*. Insights into the Pharmacokinetics and In Vitro Cell-Based Studies of the Imidazoline I2 Receptor Ligand B06. *Int J Mol Sci* **23**, (2022).
46. Bagán, A. *et al*. Insights into the Pharmacokinetics and In Vitro Cell-Based Studies of the Imidazoline I2 Receptor Ligand B06. *Int J Mol Sci* **23**, (2022).
47. Rodriguez-Arévalo, S. *et al*. Benzofuranyl-2-imidazoles as imidazoline I2 receptor ligands for Alzheimer's disease. *Eur J Med Chem* **222**, (2021).
48. Abás, S. *et al*. Bicyclic α-Iminophosphonates as High Affinity Imidazoline I2 Receptor Ligands for Alzheimer's Disease. *J Med Chem* **63**, 3610–3633 (2020).
49. Remaury, A. *et al*. *Analysis of the Pharmacological and Molecular Heterogeneity of I 2-Imidazoline-Binding Proteins Using Monoamine Oxidase-Deficient Mouse Models*. http://www.molpharm.org (2000).
50. Choi, D. H., Yun, J. H. & Lee, J. Protective effect of the imidazoline I2 receptor agonist 2-BFI on oxidative cytotoxicity in astrocytes. *Biochem Biophys Res Commun* **503**, 3011–3016 (2018).
51. Tian, J. S., Zhai, Q. J., Zhao, Y., Chen, R. & Zhao, L. D. 2-(2-benzofuranyl)-2-imidazoline (2-BFI) improved the impairments in AD rat models by inhibiting oxidative stress, inflammation and apoptosis. *J Integr Neurosci* **16**, 385–400 (2017).
52. Li, F. *et al*. 2-BFI ameliorates EAE-induced mouse spinal cord damage: Effective therapeutic time window and possible mechanisms. *Brain Res* **1483**, 13–19 (2012).
53. Zhu, Y. B. *et al*. Brain Protection Conferred by Long-Term Administration of 2-(2-Benzofuranyl)-2-Imidazoline Against Experimental Autoimmune Encephalomyelitis. *Neurochem Res* **40**, 572–578 (2015).
54. Siemian, J. N., Wang, K., Zhang, Y. & Li, J. X. Mechanisms of imidazoline I 2 receptor agonist-induced antinociception in rats: involvement of monoaminergic neurotransmission. *Br J Pharmacol* **175**, 1519–1534 (2018).
55. Craven, J. A. & Conway, E. L. Effects of alpha 2-adrenoceptor antagonists and imidazoline2-receptor ligands on neuronal damage in global ischaemia in the rat. *Clin Exp Pharmacol Physiol* **24**, 204–207 (1997).
56. Han, Z., Zhang, H. X., Tian, J. S., Zheng, R. Y. & Hou, S. T. 2-(-2-benzofuranyl)-2-imidazoline induces Bcl-2 expression and provides neuroprotection against transient cerebral ischemia in rats. *Brain Res* **1361**, 86–92 (2010).

57. Han, Z., Yang, J. L., Jiang, S. X., Hou, S. T. & Zheng, R. Y. Fast, Non-Competitive and Reversible Inhibition of NMDA-Activated Currents by 2-BFI Confers Neuroprotection. *PLoS One* **8**, (2013).
58. Li, J., Thorn, D. A., Qiu, Y., Peng, B. & Zhang, Y. Antihyperalgesic effects of imidazoline I2 receptor ligands in rat models of inflammatory and neuropathic pain. *Br J Pharmacol* **171**, 1580–1590 (2014).
59. Griñán-Ferré, C. *et al*. Behavioral and Cognitive Improvement Induced by Novel Imidazoline I2 Receptor Ligands in Female SAMP8 Mice. *Neurotherapeutics* **16**, 416–431 (2019).
60. Kimura, A. *et al*. Identification of an imidazoline binding protein: Creatine kinase and an imidazoline-2 binding site. *Brain Res* **1279**, 21–28 (2009).
61. Bousquet, P., Hudson, A., García-Sevilla, J. A. & Li, J. X. Imidazoline receptor system: The past, the present, and the future. *Pharmacol Rev* **72**, 50–79 (2020).
62. Mittermaier, A. & Kay, L. E. New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science (1979)* **312**, 224–228 (2006).
63. Bourgeois, D. & Royant, A. Advances in kinetic protein crystallography. *Curr Opin Struct Biol* **15**, 538–547 (2005).
64. Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat Struct Biol* **7**, 724–729 (2000).
65. Graf, C., Stankiewicz, M., Kramer, G. & Mayer, M. P. Spatially and kinetically resolved changes in the conformational dynamics of the Hsp90 chaperone machine. *EMBO J* **28**, 602–613 (2009).
66. Perozzo, R., Folkers, G. & Scapozza, L. Thermodynamics of Protein–Ligand Interactions: History, Presence, and Future Aspects. *Journal of Receptors and Signal Transduction* **24**, 1–52 (2004).
67. Ladbury, J. E. & Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem Biol* **3**, 791–801 (1996).
68. Torreri, P., Ceccarini, M., Macioce, P. & Petrucci, T. C. Biomolecular interactions by Surface Plasmon Resonance technology. *Ann Ist Super Sanita* **41**, 437–41 (2005).
69. Jönsson, U. *et al*. Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. *Biotechniques* **11**, 620–7 (1991).
70. Weber, G. Polarization of the fluorescence of macromolecules. 1. Theory and experimental method. *Biochemical Journal* **51**, 145–155 (1952).
71. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* **6**, 365–87 (2011).
72. Varadi, M. *et al*. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**, D368–D375 (2024).
73. Jumper, J. *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
74. Huang, S. Y., Grinter, S. Z. & Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics* **12**, 12899–12908 (2010).
75. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* **9**, 646–652 (2002).

76. Alvarez-Garcia, D., Schmidtke, P., Cubero, E. & Barril, X. Extracting Atomic Contributions to Binding Free Energy Using Molecular Dynamics Simulations with Mixed Solvents (MDmix). *Curr Drug Discov Technol* **19**, (2021).
77. Wolfenden, R. & Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc Chem Res* **34**, 938–945 (2001).
78. Martí, S. *et al*. Theoretical insights in enzyme catalysis. *Chem Soc Rev* **33**, 98–107 (2004).
79. Wolfenden, R. & Snider, M. J. The Depth of Chemical Time and the Power of Enzymes as Catalysts. *Acc Chem Res* **34**, 938–945 (2001).
80. Bell, E. L. *et al*. Biocatalysis. *Nature Reviews Methods Primers* vol. 1 (2021).
81. Agarwal, P. K. A Biophysical Perspective on Enzyme Catalysis. *Biochemistry* vol. 58 438–449 (2019).
82. PAULING, L. Molecular Architecture and Biological Reactions. *Chemical & Engineering News Archive* **24**, 1375–1377 (1946).
83. Warshel, A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *Journal of Biological Chemistry* **273**, 27035–27038 (1998).
84. Zhang, X. & Houk, K. N. Why Enzymes Are Proficient Catalysts: Beyond the Pauling Paradigm. *Acc Chem Res* **38**, 379–385 (2005).
85. Roduner, E. Understanding catalysis. *Chemical Society Reviews* vol. 43 8226–8239 (2014).
86. Schlexer, Philomena. Computational Modeling in Heterogeneous Catalysis. in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, (2017).
87. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Computational enzymatic catalysis - Clarifying enzymatic mechanisms with the help of computers. *Physical Chemistry Chemical Physics* vol. 14 12431–12441 (2012).
88. Lisi, G. P. & Loria, J. P. Allostery in enzyme catalysis. *Current Opinion in Structural Biology* vol. 47 123–130 (2017).
89. Novelli, F. *et al*. Time-Domain THz Spectroscopy Reveals Coupled Protein–Hydration Dielectric Response in Solutions of Native and Fibrils of Human Lysozyme. *J Phys Chem B* **121**, 4810–4816 (2017).
90. Warshel, A. *et al*. Electrostatic basis for enzyme catalysis. *Chemical Reviews* vol. 106 3210–3235 (2006).
91. Sousa, S. F. *et al*. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *WIREs Computational Molecular Science* **7**, (2017).
92. Warshel, A. & Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J Am Chem Soc* **102**, 6218–6226 (1980).
93. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers. *Physical Chemistry Chemical Physics* **14**, 12431 (2012).
94. Lonsdale, R., Harvey, J. N. & Mulholland, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chem Soc Rev* **41**, 3025 (2012).
95. Shaw, D. E. *et al*. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (1979)* **330**, 341–346 (2010).
96. Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology* **135**, 555–562 (2010).

97. Bowerman, S. & Wereszczynski, J. Detecting Allosteric Networks Using Molecular Dynamics Simulation. in 429–447 (2016).
98. González, M. A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12**, 169–200 (2011).
99. 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* **25**, 1157–1174 (2004).
100. Cieplak, P., Dupradeau, F.-Y., Duan, Y. & Wang, J. Polarization effects in molecular mechanical force fields. *J Phys Condens Matter* **21**, 333102 (2009).
101. Rick, S. W. & Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. in *Reviews in Computational Chemistry, Volume 18* 89–146 (John Wiley & Sons, Inc., Hoboken, New Jersey, USA).
102. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* **114**, 10024–10035 (1992).
103. Patel, S. & Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J Comput Chem* **25**, 1–16 (2004).
104. Lamoureux, G., Harder, E., Vorobyov, I. V., Roux, B. & MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem Phys Lett* **418**, 245–249 (2006).
105. Geerke, D. P. & van Gunsteren, W. F. Calculation of the Free Energy of Polarization: Quantifying the Effect of Explicitly Treating Electronic Polarization on the Transferability of Force-Field Parameters. *J Phys Chem B* **111**, 6425–6436 (2007).
106. Cieplak, P., Caldwell, J. & Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comput Chem* **22**, 1048–1057 (2001).
107. Wang, Z. *et al*. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J Comput Chem* **27**, 781–790 (2006).
108. Wang, Z. *et al*. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J Comput Chem* **27**, 781–790 (2006).
109. Cieplak, P., Caldwell, J. & Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comput Chem* **22**, 1048–1057 (2001).
110. Case, D. A. *et al*. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).
111. Kaminski, G. A. *et al*. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J Comput Chem* **23**, 1515–1531 (2002).

112. Ren, P. & Ponder, J. W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J Comput Chem* **23**, 1497–1506 (2002).
113. Ponder, J. W. & Case, D. A. Force Fields for Protein Simulations. in 27–85 (2003).
114. Wang, J. *et al*. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J Phys Chem B* **115**, 3100–3111 (2011).
115. Wang, J. *et al*. Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability. *J Phys Chem B* **115**, 3091–3099 (2011).
116. Debye, P. Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index. *Math Ann* **67**, 535–558 (1909).
117. Hestenes, M. R. & Stiefel, E. Methods of conjugate gradients for solving linear systems. *J Res Natl Bur Stand (1934)* **49**, 409 (1952).
118. A. R. Leach. *Molecular Modelling: Principles and Applications Second Edition*. (2001).
119. Frenkel, D. & Smit, B. *Understanding Molecular Simulation*. (2002).
120. C. J. Cramer. Essentials of Computational Chemistry: Theories and Models. (2004).
121. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23**, 327–341 (1977).
122. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* **20**, 786–798 (1999).
123. Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864–1874 (2015).
124. Lemons, D. S. & Gythiel, A. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" . *Am J Phys* **65**, 1079–1081 (1997).
125. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684–3690 (1984).
126. Cheatham, T. E. I., Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J Am Chem Soc* **117**, 4193–4194 (1995).
127. Hwang, W. *et al*. CHARMM at 45: Enhancements in Accessibility, Functionality, and Speed. *J Phys Chem B* **128**, 9976–10042 (2024).
128. Case, D. A. *et al*. The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–88 (2005).
129. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* **91**, 43–56 (1995).
130. Eastman, P. *et al*. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J Chem Theory Comput* **9**, 461–469 (2013).

131. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **25**, 247–260 (2006).
132. MacKerell, A. D. *et al*. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **102**, 3586–3616 (1998).
133. Scott, W. R. P. *et al*. The GROMOS Biomolecular Simulation Program Package. *J Phys Chem A* **103**, 3596–3607 (1999).
134. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* **118**, 11225–11236 (1996).
135. Śledź, P. & Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* **48**, 93–102 (2018).
136. Yuriev, E., Holien, J. & Ramsland, P. A. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *Journal of Molecular Recognition* **28**, 581–604 (2015).
137. Alvarez-Garcia, D. & Barril, X. Molecular Simulations with Solvent Competition Quantify Water Displaceability and Provide Accurate Interaction Maps of Protein Binding Sites. *J Med Chem* **57**, 8530–8539 (2014).
138. Ruiz-Carmona, S. *et al*. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput Biol* **10**, (2014).
139. Ghanakota, P. & Carlson, H. A. Moving beyond Active-Site Detection: MixMD Applied to Allosteric Systems. *Journal of Physical Chemistry B* **120**, 8685–8695 (2016).
140. Faller, C. E., Raman, E. P., MacKerell, A. D. & Guvench, O. Site Identification by Ligand Competitive Saturation (SILCS) Simulations for Fragment-Based Drug Design. in 75–87 (2015).
141. Cuchillo, R., Pinto-Gil, K. & Michel, J. A collective variable for the rapid exploration of protein druggability. *J Chem Theory Comput* **11**, 1292–1307 (2015).
142. Borsatto, A., Gianquinto, E., Rizzi, V. & Gervasio, F. L. SWISH-X, an Expanded Approach to Detect Cryptic Pockets in Proteins and at Protein-Protein Interfaces. *J Chem Theory Comput* **20**, 3335–3348 (2024).
143. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **28**, 849–857 (1985).
144. Boobbyer, D. N. A., Goodford, P. J., McWhinnie, P. M. & Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J Med Chem* **32**, 1083–1094 (1989).
145. Guedes, I. A., Pereira, F. S. S. & Dardenne, L. E. Empirical Scoring Functions for Structure-Based Virtual Screening: Applications, Critical Aspects, and Challenges. *Front Pharmacol* **9**, (2018).
146. McNutt, A. T. *et al*. GNINA 1.0: molecular docking with deep learning. *J Cheminform* **13**, 43 (2021).
147. Sauer, S., Matter, H., Hessler, G. & Grebner, C. Optimizing interactions to protein binding sites by integrating docking-scoring strategies into generative AI methods. *Front Chem* **10**, (2022).
148. Gentile, F. *et al*. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Cent Sci* **6**, 939–949 (2020).
149. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**, A1133–A1138 (1965).

150. R. Peverati and D. G. Truhlar. Quest for a universal density functional: The accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372**, (2014).
151. Sousa, S. F., Fernandes, P. A. & Ramos, M. J. General performance of density functionals. *Journal of Physical Chemistry A* **111**, 10439–10452 (2007).
152. Bryantsev, V. S., Diallo, M. S., Van Duin, A. C. T. & Goddard, W. A. Evaluation of B3LYP, X3LYP, and M06-Class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J Chem Theory Comput* **5**, 1016–1026 (2009).
153. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* **120**, 215–241 (2008).
154. Hehre, W. J., Ditchfield, R. & Pople, J. A. 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**, 2257–2261 (1972).
155. Jensen, F. *Introduction to Computational Chemistry*. (Wiley, Chichester, UK, 2017).
156. Curutchet, C. *et al*. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J Chem Theory Comput* **5**, 1838–1848 (2009).
157. Zhang, Y., Lee, T.-S. & Yang, W. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J Chem Phys* **110**, 46–54 (1999).
158. Antes, I. & Thiel, W. Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods. *J Phys Chem A* **103**, 9290–9295 (1999).
159. Hu, H. & Yang, W. Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annu Rev Phys Chem* **59**, 573–601 (2008).
160. Duarte, F., Amrein, B. A., Blaha-Nelson, D. & Kamerlin, S. C. L. Recent advances in QM/MM free energy calculations using reference potentials. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1850**, 954–965 (2015).
161. Maeda, S., Harabuchi, Y., Ono, Y., Taketsugu, T. & Morokuma, K. Intrinsic reaction coordinate: Calculation, bifurcation, and automated search. *Int J Quantum Chem* **115**, 258–269 (2015).
162. Car, R. & Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys Rev Lett* **55**, 2471–2474 (1985).
163. Woo, T. K., Margl, P. M., Blöchl, P. E. & Ziegler, T. A Combined Car–Parrinello QM/MM Implementation for ab Initio Molecular Dynamics Simulations of Extended Systems: Application to Transition Metal Catalysis. *J Phys Chem B* **101**, 7877–7880 (1997).
164. Gustafson, S. J. & Sigal, I. M. *Born–Oppenheimer Approximation and Adiabatic Dynamics. In: Mathematical Concepts of Quantum Mechanics*. (Springer, Cham., 2020).
165. CASIDA, M. E. Time-Dependent Density Functional Response Theory for Molecules. in 155–192 (1995).

166. Daday, C., Curutchet, C., Sinicropi, A., Mennucci, B. & Filippi, C. Chromophore–Protein Coupling beyond Nonpolarizable Models: Understanding Absorption in Green Fluorescent Protein. *J Chem Theory Comput* **11**, 4825–4839 (2015).
167. Li, Q., Mennucci, B., Robb, M. A., Blancafort, L. & Curutchet, C. Polarizable QM/MM Multiconfiguration Self-Consistent Field Approach with State-Specific Corrections: Environment Effects on Cytosine Absorption Spectrum. *J Chem Theory Comput* **11**, 1674–1682 (2015).
168. Curutchet, C. *et al*. Photosynthetic Light-Harvesting Is Tuned by the Heterogeneous Polarizable Environment of the Protein. *J Am Chem Soc* **133**, 3078–3084 (2011).
169. Pinheiro, S. & Curutchet, C. Can Förster Theory Describe Stereoselective Energy Transfer Dynamics in a Protein–Ligand Complex? *J Phys Chem B* **121**, 2265–2278 (2017).
170. Espargaró, A. *et al*. On the Binding of Congo Red to Amyloid Fibrils. *Angewandte Chemie International Edition* **59**, 8104–8107 (2020).
171. Cupellini, L., Corbella, M., Mennucci, B. & Curutchet, C. Electronic energy transfer in biomacromolecules. *WIREs Computational Molecular Science* **9**, (2019).
172. Rosnik, A. M. & Curutchet, C. Theoretical Characterization of the Spectral Density of the Water-Soluble Chlorophyll-Binding Protein from Combined Quantum Mechanics/Molecular Mechanics Molecular Dynamics Simulations. *J Chem Theory Comput* **11**, 5826–5837 (2015).
173. Jurinovich, S., Cupellini, L., Guido, C. A. & Mennucci, B. EXAT: EXcitonic analysis tool. *J Comput Chem* **39**, 279–286 (2018).
174. Ma, J. & Cao, J. Förster resonance energy transfer, absorption and emission spectra in multichromophoric systems. I. Full cumulant expansions and system-bath entanglement. *J Chem Phys* **142**, (2015).
175. Cupellini, L., Lipparini, F. & Cao, J. Absorption and Circular Dichroism Spectra of Molecular Aggregates With the Full Cumulant Expansion. *J Phys Chem B* **124**, 8610–8617 (2020).
176. Dreuw, A., Harbach, P. H. P., Mewes, J. M. & Wormit, M. Quantum chemical excited state calculations on pigment–protein complexes require thorough geometry re-optimization of experimental crystal structures. *Theor Chem Acc* **125**, 419–426 (2010).
177. Lee, M. K., Bravaya, K. B. & Coker, D. F. First-Principles Models for Biological Light-Harvesting: Phycobiliprotein Complexes from Cryptophyte Algae. *J Am Chem Soc* **139**, 7803–7814 (2017).
178. Curutchet, C. *et al*. Energy Flow in the Cryptophyte PE545 Antenna Is Directed by Bilin Pigment Conformation. *J Phys Chem B* **117**, 4263–4273 (2013).
179. Maier, J. A. *et al*. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713 (2015).
180. 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* **79**, 926–935 (1983).
181. Case, D. A. *et al*. *AMBER 20*. (San Francisco: University of California, 2020).
182. Frisch, M. J. *et al*. *Gaussian 16, Revision B.01*. (Wallingford CT: Gaussian, Inc, 2016).

183. Vreven, T. *et al*. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J Chem Theory Comput* **2**, 815–826 (2006).
184. Senn, H. M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **48**, 1198–1229 (2009).
185. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* **393**, 51–57 (2004).
186. Arpin, P. C. *et al*. Spectroscopic Studies of Cryptophyte Light Harvesting Proteins: Vibrations and Coherent Oscillations. *J Phys Chem B* **119**, 10025–10034 (2015).
187. Corbella, M., Cupellini, L., Lipparini, F., Scholes, G. D. & Curutchet, C. Spectral Variability in Phycocyanin Cryptophyte Antenna Complexes is Controlled by Changes in the α-Polypeptide Chains. *ChemPhotoChem* **3**, 945–956 (2019).
188. Mirkovic, T. *et al*. Ultrafast light harvesting dynamics in the cryptophyte phycocyanin 645. *Photochemical & Photobiological Sciences* **6**, 964–975 (2007).
189. Harrop, S. J. *et al*. Single-residue insertion switches the quaternary structure and exciton states of cryptophyte light-harvesting proteins. *Proc Natl Acad Sci U S A* **111**, E2666-75 (2014).
190. Arpin, P. *et al*. Spectroscopic Studies of Cryptophyte Light Harvesting Proteins: Vibrations and Coherent Oscillations. *J Phys Chem B* **119**, 10025–10034 (2015).
191. Tesson, F. *et al*. Localization of I2-imidazoline binding sites on monoamine oxidases. *Journal of Biological Chemistry* **270**, 9856–9861 (1995).
192. Bour, S. *et al*. The imidazoline I2-site ligands BU 224 and 2-BFI inhibit MAO-A and MAO-B activities, hydrogen peroxide production, and lipolysis in rodent and human adipocytes. *Eur J Pharmacol* **552**, 20–30 (2006).
193. Keller, B. & García-Sevilla, J. A. Effects of I2-imidazoline receptor (IR) alkylating BU99006 in the mouse brain: Upregulation of nischarin/I1-IR and μ-opioid receptor proteins and modulation of associated signalling pathways. *Neurochem Int* **108**, 169–176 (2017).
194. Bonivento, D. *et al*. Potentiation of ligand binding through cooperative effects in monoamine oxidase B. *Journal of Biological Chemistry* **285**, 36849–36856 (2010).
195. Dennis, T. N. *et al*. Human ARMT1 structure and substrate specificity indicates that it is a DUF89 family damage-control phosphatase. *J Struct Biol* **212**, 107576 (2020).
196. Meagher, K. L., Redman, L. T. & Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **24**, 1016–1025 (2003).
197. Bong, S. M. *et al*. Structural studies of human brain-type creatine kinase complexed with the ADP–Mg2+–NO3 –creatine transition-state analogue complex. *FEBS Lett* **582**, 3959–3965 (2008).
198. Fu, Z. *et al*. Crystal Structure of Human 3-Hydroxy-3-methylglutaryl-CoA Lyase. *Journal of Biological Chemistry* **281**, 7526–7532 (2006).
199. Marvin, J. ChemAxon: Chemical informatics software for industry, academia and government. *ChemAxon*.
200. Duarte, F. *et al*. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J Phys Chem B* **118**, 4351–4362 (2014).
201. Sun, Z., Chang, C. H. & Ernsberger, P. Identification of IRAS/Nischarin as an I1-imidazoline receptor in PC12 rat pheochromocytoma cells. *J Neurochem* **101**, 99–108 (2007).

202. Maziveyi, M. & Alahari, S. K. Breast cancer tumor suppressors: A special emphasis on novel protein nischarin. *Cancer Research* vol. 75 4252–4259 (2015).
203. Gupta, S. *et al*. The non-adrenergic imidazoline-1 receptor protein nischarin is a key regulator of astrocyte glutamate uptake. *iScience* **25**, (2022).
204. Zheng, P. *et al*. Contribution of Nischarin/IRAS in CNS development, injury and diseases. *Journal of Advanced Research* vol. 54 43–57 (2023).
205. Li, F. *et al*. Imidazoline receptor antisera-selected/Nischarin regulates the effect of agmatine on the development of morphine dependence. *Addiction Biology* **17**, 392–408 (2012).
206. Dontenwill, M. *et al*. IRAS, the human homologue of Nischarin, prolongs survival of transfected PC12 cells [3]. *Cell Death and Differentiation* vol. 10 933–935 (2003).
207. Lim, K. P. & Hong, W. Human nischarin/imidazoline receptor antisera-selected protein is targeted to the endosomes by a combined action of a PX domain and a coiled-coil region. *Journal of Biological Chemistry* **279**, 54770–54782 (2004).
208. Sano, H., Liu, S. C. H., Lane, W. S., Piletz, J. E. & Lienhard, G. E. Insulin receptor substrate 4 associates with the protein IRAS. *Journal of Biological Chemistry* **277**, 19439–19447 (2002).
209. Kuijl, C. *et al*. Rac and Rab GTPases dual effector Nischarin regulates vesicle maturation to facilitate survival of intracellular bacteria. *EMBO Journal* vol. 32 713–727 (2013).
210. Nikolic, K., Veljkovic, N., Gemovic, B., Srdic-Rajic, T. & Agbaba, D. Imidazoline-1 Receptor Ligands as Apoptotic Agents: Pharmacophore Modeling and Virtual Docking Study. *Comb Chem High Throughput Screen* **16**, 298–319 (2013).
211. Uotila L. & Koivtjsalo M. *Purification and Properties of S-Formylglutathione Hydrolase from Human Liver*. *THE JOURNAL OF BIOLOGICAL CHEMISTRY* vol. 249 (1974).
212. Xu, Y. & Sun, Z. *Regulation of S-Formylglutathione Hydrolase by the Anti-Aging Gene Klotho*. *Oncotarget* vol. 8 (2017).
213. Lee, W., Wheatley, W., BENEDICTt, W. F., Huang, C. & Y-h Lee, E. P. *Purification, Biochemical Characterization, and Biological Function of Human Esterase D (Genetic Marker/Esterase Antibodies/Detoxification/Phenobarbital Induction)*. *Proc. Nati. Acad. Sci. USA* vol. 83 (1986).
214. Cummins, I. *et al*. Unique Regulation of the Active site of the Serine Esterase S-Formylglutathione Hydrolase. *J Mol Biol* **359**, 422–432 (2006).
215. Lee, C. W. *et al*. Structural and functional characterization of a novel cold-active S-formylglutathione hydrolase (SfSFGH) homolog from Shewanella frigidimarina, a psychrophilic bacterium. *Microb Cell Fact* **18**, (2019).
216. Wu, D. *et al*. Crystal structure of human esterase D: a potential genetic marker of retinoblastoma. *The FASEB Journal* **23**, 1441–1446 (2009).
217. Zhang, F. *et al*. β-amyloid redirects norepinephrine signaling to activate the pathogenic GSK3β/tau cascade. *Sci Transl Med* **12**, (2020).
218. Qu, L. *et al*. Structural Basis of the Diversity of Adrenergic Receptors. *Cell Rep* **29**, (2019).
219. Carter, A. J. *Hippocampal Noradrenaline Release in Awake, Freely Moving Rats Is Regulated by Alpha-2 Adrenoceptors but Not by Adenosine Receptors*. (1997).
220. Peter, S. *et al*. Comparative Study of Allosteric GPCR Binding Sites and Their Ligandability Potential. *J Chem Inf Model* **64**, 8176–8192 (2024).

221. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851–867 (2016).
222. Miyaji, M. *et al*. Selective DNA-binding of SP120 (rat ortholog of human hnRNP U) is mediated by arginine-glycine rich domain and modulated by RNA. *PLoS One* **18**, (2023).
223. Hasegawa, Y. *et al*. The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. *Dev Cell* **19**, 469–476 (2010).
224. Han, S. P., Tang, Y. H. & Smith, R. Functional diversity of the hnRNPs: past, present and perspectives. *Biochemical Journal* **430**, 379–392 (2010).
225. Marenda, M., Lazarova, E. & Gilbert, N. The role of SAF-A/hnRNP U in regulating chromatin structure. *Curr Opin Genet Dev* **72**, 38–44 (2022).
226. Brandão-Teles, C., Antunes, A. S. L. M., de Moraes Vrechi, T. A. & Martins-de-Souza, D. The Roles of hnRNP Family in the Brain and Brain-Related Disorders. *Mol Neurobiol* **61**, 3578–3595 (2024).
227. Stegink, L. D. & Coon, M. J. Stereospecificity and Other Properties of Highly Purified β-Hydroxy-β-methylglutaryl Coenzyme A Cleavage Enzyme from Bovine Liver. *Journal of Biological Chemistry* **243**, 5272–5279 (1968).
228. Fu, Z. *et al*. Crystal structure of human 3-hydroxy-3-methylglutaryl-CoA lyase: Insights into catalysis and the molecular basis for hydroxymethylglutaric aciduria. *Journal of Biological Chemistry* **281**, 7526–7532 (2006).
229. Puisac, B. *et al*. HMG–CoA Lyase Deficiency. in *Advances in the Study of Genetic Disorders* (2011)
230. Yang, H., Shan, W., Zhu, F., Wu, J. & Wang, Q. Ketone bodies in neurological diseases: Focus on neuroprotection and underlying mechanisms. *Frontiers in Neurology* vol. 10 (2019).
231. Jensen, N. J., Wodschow, H. Z., Nilsson, M. & Rungby, J. Effects of ketone bodies on brain metabolism and function in neurodegenerative diseases. *International Journal of Molecular Sciences* vol. 21 1–17 Preprint at (2020).
232. Yin, J., Han, P., Tang, Z., Liu, Q. & Shi, J. Sirtuin 3 Mediates Neuroprotection of Ketones against Ischemic Stroke. *Journal of Cerebral Blood Flow & Metabolism* **35**, 1783–1789 (2015).
233. Yao, J., Chen, S., Mao, Z., Cadenas, E. & Brinton, R. D. 2-Deoxy-D-Glucose Treatment Induces Ketogenesis, Sustains Mitochondrial Function, and Reduces Pathology in Female Mouse Model of Alzheimer's Disease. *PLoS One* **6**, (2011).
234. Srikanth, V. *et al*. Advanced glycation endproducts and their receptor RAGE in Alzheimer's disease. *Neurobiol Aging* **32**, 763–777 (2011).
235. Fu, Z., Runquist, J. A., Montgomery, C., Miziorko, H. M. & Kim, J.-J. P. Functional Insights into Human HMG-CoA Lyase from Structures of Acyl-CoA-containing Ternary Complexes. *Journal of Biological Chemistry* **285**, 26341–26349 (2010).
236. Manos, P. & Bryan, G. K. Cellular and Subcellular Compartmentation of Creatine Kinase in Brain. *Dev Neurosci* **15**, 271–279 (1993).
237. Kuiper, J. W. P. *et al*. Local ATP Generation by Brain-Type Creatine Kinase (CK-B) Facilitates Cell Motility. *PLoS One* **4**, (2009).
238. Hettling, H. & van Beek, J. H. Analyzing the Functional Properties of the Creatine Kinase System with Multiscale 'Sloppy' Modeling. *PLoS Comput Biol* **7**, (2011).
239. Meiser, J., Weindl, D. & Hiller, K. *Complexity of Dopamine Metabolism. Cell Communication and Signaling* vol. 11 (2013).

240. Muñoz, P., Huenchuguala, S., Paris, I. & Segura-Aguilar, J. Dopamine oxidation and autophagy. *Parkinson's Disease* Preprint (2012).
241. Moraes, F. P. & De Azevedo, W. F. Targeting imidazoline site on monoamine oxidase B through molecular docking simulations. *J Mol Model* **18**, 3877–3886 (2012).
242. Chapman, S. K., Perham, R. N. & Scrutton, N. S. *Flavins and Flavoproteins 2002*. (2002).
243. MacMillar, S., Edmondson, D. E. & Matsson, O. Nitrogen kinetic isotope effects for the monoamine oxidase B-catalyzed oxidation of benzylamine and (1,1-(2)H2)benzylamine: nitrogen rehybridization and CH bond cleavage are not concerted. *J Am Chem Soc* **133**, 12319–12321 (2011).
244. Akyüz, M. A. & Erdem, S. S. Computational modeling of the direct hydride transfer mechanism for the MAO catalyzed oxidation of phenethylamine and benzylamine: ONIOM (QM/QM) calculations. *J Neural Transm* **120**, 937–945 (2013).
245. Repič, M. *et al*. Empirical valence bond simulations of the hydride transfer step in the monoamine oxidase B catalyzed metabolism of dopamine. *Proteins: Structure, Function and Bioinformatics* **82**, 3347–3355 (2014).
246. Mavri, J., Matute, R. A., Chu, Z. T. & Vianello, R. Path Integral Simulation of the H/D Kinetic Isotope Effect in Monoamine Oxidase B Catalyzed Decomposition of Dopamine. *Journal of Physical Chemistry B* **120**, 3488–3492 (2016).
247. Edmondson, D. E., Binda, C., Wang, J., Upadhyay, A. K. & Mattevi, A. Molecular and Mechanistic Properties of the Membrane-Bound Mitochondrial Monoamine Oxidases. *Biochemistry* **48**, 4220–4230 (2009).
248. Rajić, M., Prah, A. & Stare, J. Deciphering the Two-Step Hydride Mechanism of Monoamine Oxidase Flavoenzymes. *ACS Omega* **9**, 43046–43057 (2024).
249. Frisch, M. J. *et al*. Gaussian 16, Revision. C.01. *Gaussian, Inc.: Wallingford, CT* (2016).
250. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **32**, 1456–1465 (2011).

Appendix 1: Multiscale modelling of photosynthetic complexes

Table A1.1. Root-mean square fluctuations (RMSF) calculated along classical MD trajectories for the pigments in PC612 and PC645.

PC612	RMSF (Å)	PC645	RMSF (Å)
$DBV_{50/61B}$	0.96	$DBV_{50/61B}$	0.76
$DBV_{50/61D}$	0.94	$DBV_{50/61D}$	0.84
PCB_{158B}	0.93	PCB_{158B}	0.87
PCB_{158D}	1.03	PCB_{158D}	0.85
PCB_{82B}	0.78	PCB_{82B}	0.80
PCB_{82D}	0.91	PCB_{82D}	0.78
PCB_{20A}	2.02	MBV_{19A}	0.81
PCB_{20C}	1.30	MBV_{19C}	0.79

Table A1.2. Average MD-OPT and MD-BOMD site energies (cm^{-1}) computed along the classical MD trajectories of PC612 and PC645.

PC612	MD-BOMD	MD-OPT	PC645	MD-BOMD	MD-OPT
$DBV_{50/61B}$	19812	20452	$DBV_{50/61B}$	19614	20166
$DBV_{50/61D}$	20050	20598	$DBV_{50/61D}$	19674	20407
PCB_{158B}	18290	18702	PCB_{158B}	18134	18555
PCB_{158D}	18331	18760	PCB_{158D}	17924	18484
PCB_{82B}	17817	18794	PCB_{82B}	17942	18657
PCB_{82D}	17932	18719	PCB_{82D}	17675	18196
PCB_{20A}	17964	18926	MBV_{19A}	17497	18228
PCB_{20C}	18102	18876	MBV_{19C}	17629	18175

Table A1.3. Standard deviation (cm^{-1}) of MD-OPT and MD-BOMD site energies computed along the classical MD trajectories of PC612 and PC645.

PC612	MD-BOMD	MD-OPT	PC645	MD-BOMD	MD-OPT
$DBV_{50/61B}$	429	259	$DBV_{50/61B}$	441	243
$DBV_{50/61D}$	414	299	$DBV_{50/61D}$	308	315
PCB_{158B}	370	258	PCB_{158B}	531	305
PCB_{158D}	412	305	PCB_{158D}	397	294
PCB_{82B}	364	240	PCB_{82B}	606	419
PCB_{82D}	331	185	PCB_{82D}	459	241
PCB_{20A}	568	305	MBV_{19A}	346	146
PCB_{20C}	455	352	MBV_{19C}	301	244

Table A1.4. Average MD-OPT and MD-BOMD electronic transition dipole moments (Debye) computed along the classical MD trajectories of PC612 and PC645.

PC612	MD-BOMD	MD-OPT	PC645	MD-BOMD	MD-OPT
$DBV_{50/61B}$	12.9	13.1	$DBV_{50/61B}$	13.1	13.2
$DBV_{50/61D}$	12.6	12.8	$DBV_{50/61D}$	13.0	13.1
PCB_{158B}	13.5	14.0	PCB_{158B}	14.1	14.3
PCB_{158D}	13.8	14.0	PCB_{158D}	14.2	14.4
PCB_{82B}	14.4	14.4	PCB_{82B}	14.2	14.3
PCB_{82D}	14.4	14.4	PCB_{82D}	14.5	14.7
PCB_{20A}	13.6	13.6	MBV_{19A}	14.6	14.5
PCB_{20C}	14.0	14.2	MBV_{19C}	14.3	14.5

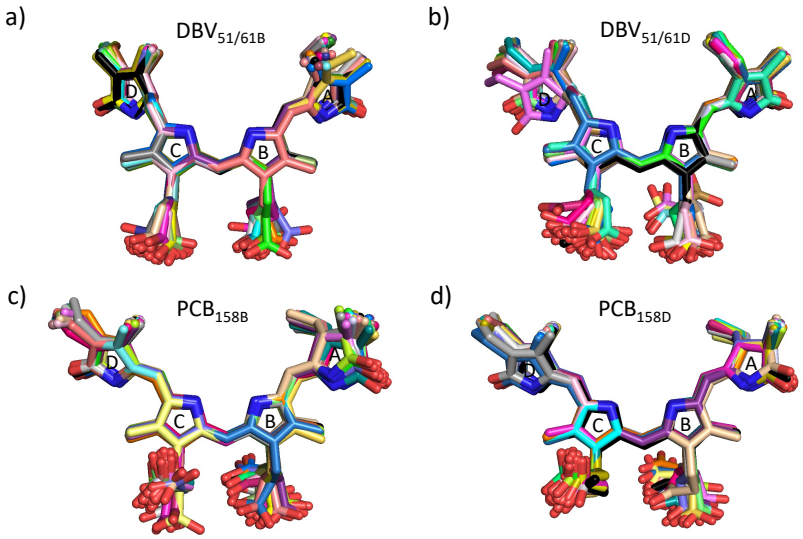


Figure A1.1. Alignment of PC645 pigment structures in the crystal structure (black) and optimized along the MD simulation (colors): a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

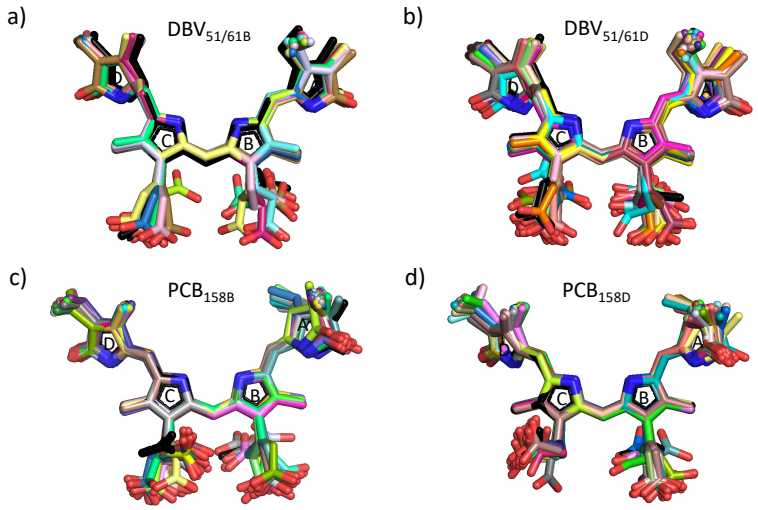


Figure A1.2. Alignment of PC612 pigment structures in the crystal structure (black) and optimized along the MD simulation (colors): a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

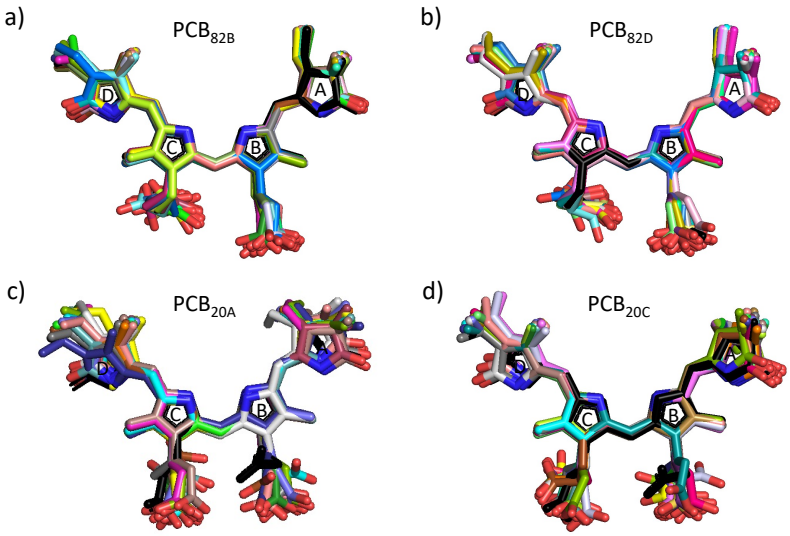


Figure A1.3. Alignment of PC612 pigment structures in the crystal structure (black) and optimized along the MD simulation (colors): a) PCB_{82B}, b) PCB_{82D}, c) PCB_{20A} and d) PCB_{20C}.

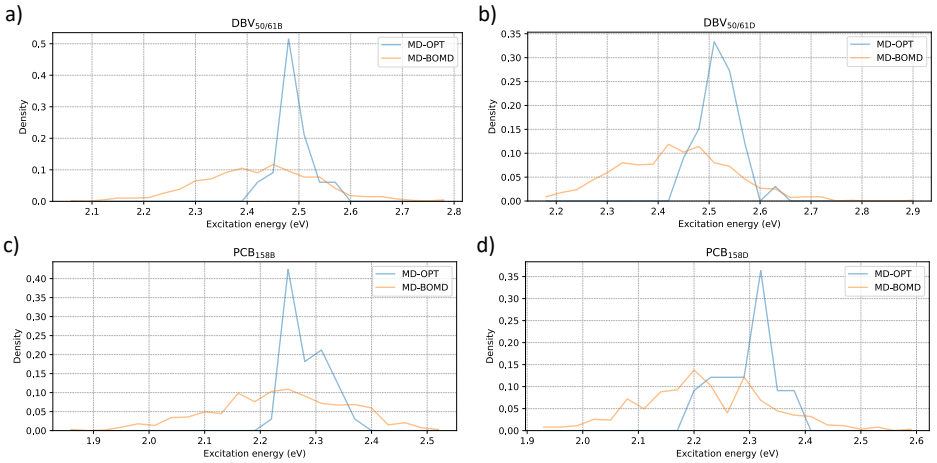


Figure A1.4. Distribution of site energies for PC645 pigments computed from MD-OPT and MD-BOMD calculations: a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

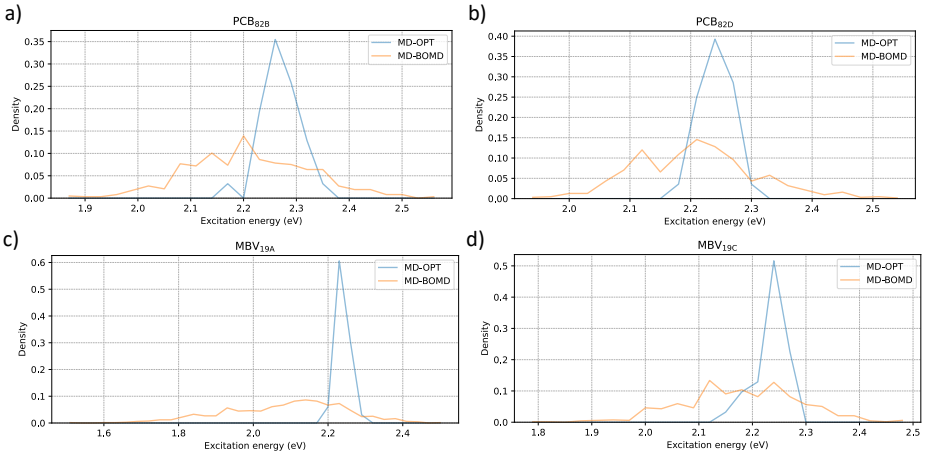


Figure A1.5. Distribution of site energies for PC645 pigments computed from MD-OPT and MD-BOMD calculations: a) PCB_{82B}, b) PCB_{82D}, c) MBV_{19A} and d) MBV_{19C}.

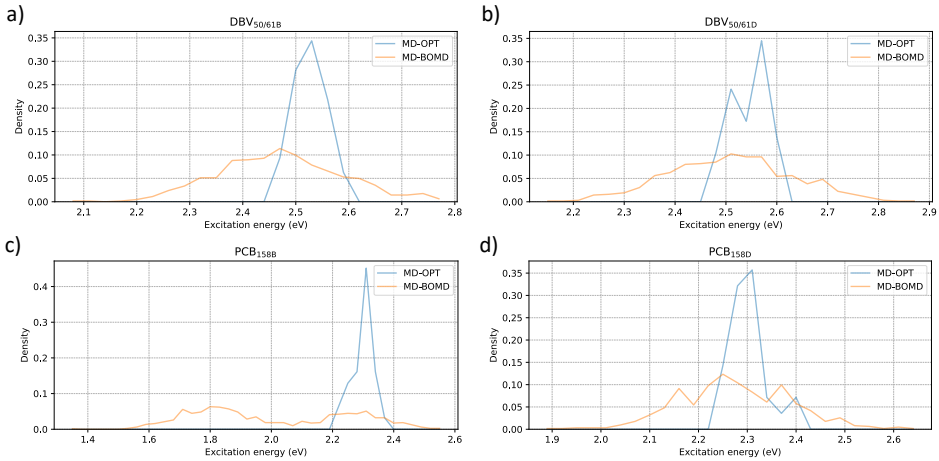


Figure A1.6. Distribution of site energies for PC612 pigments computed from MD-OPT and MD-BOMD calculations: a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

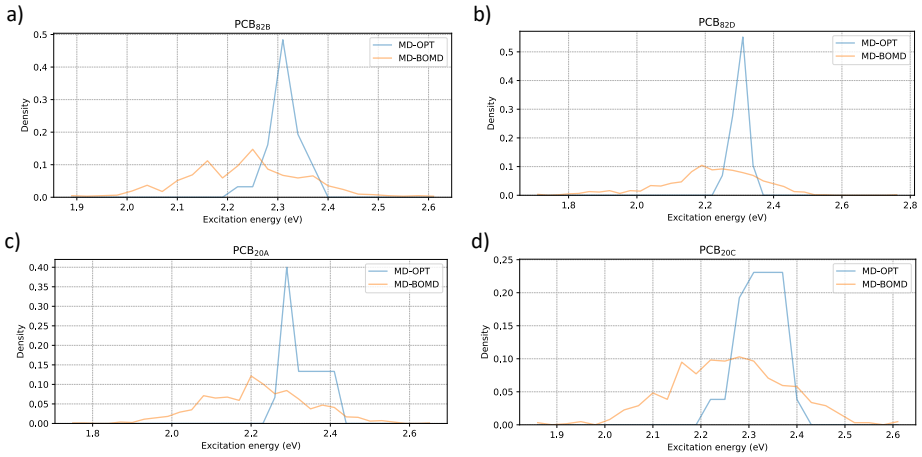


Figure A1.7. Distribution of site energies for PC612 pigments computed from MD-OPT and MD-BOMD calculations: a) PCB_{82B}, b) PCB_{82D}, c) PCB_{20A} and d) PCB_{20C}.

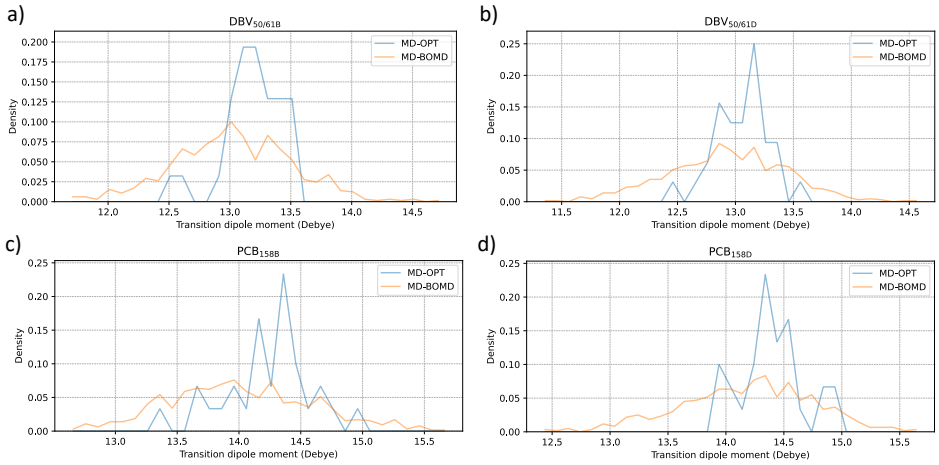


Figure A1.8. Distribution of electronic transition dipole moments for PC645 pigments computed from MD-OPT and MD-BOMD calculations: a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

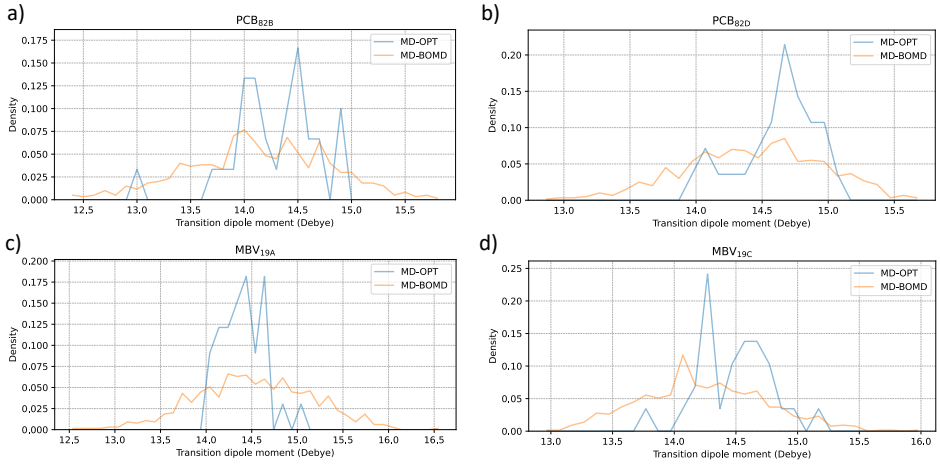


Figure A1.9. Distribution of electronic transition dipole moments for PC645 pigments computed from MD-OPT and MD-BOMD calculations: a) PCB_{82B}, b) PCB_{82D}, c) MBV_{19A} and d) MBV_{19C}.

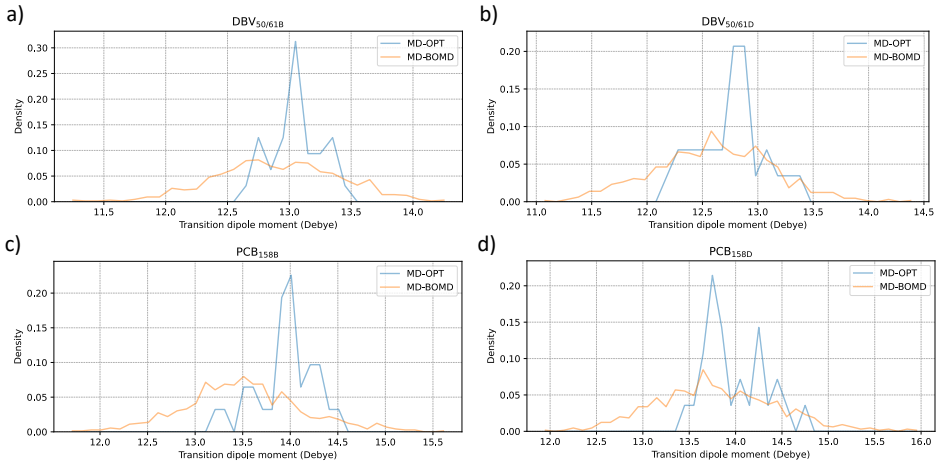


Figure A1.10. Distribution of electronic transition dipole moments for PC612 pigments computed from MD-OPT and MD-BOMD calculations: a) $DBV_{51/61B}$, b) $DBV_{51/61D}$, c) PCB_{158B} and d) PCB_{158D}.

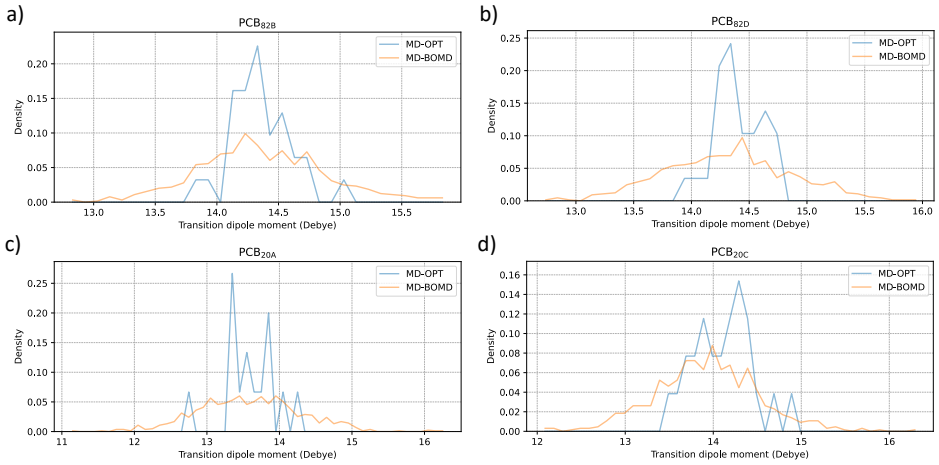


Figure A1.11. Distribution of electronic transition dipole moments for PC612 pigments computed from MD-OPT and MD-BOMD calculations: a) PCB_{82B}, b) PCB_{82D}, c) PCB_{20A} and d) PCB_{20C}.

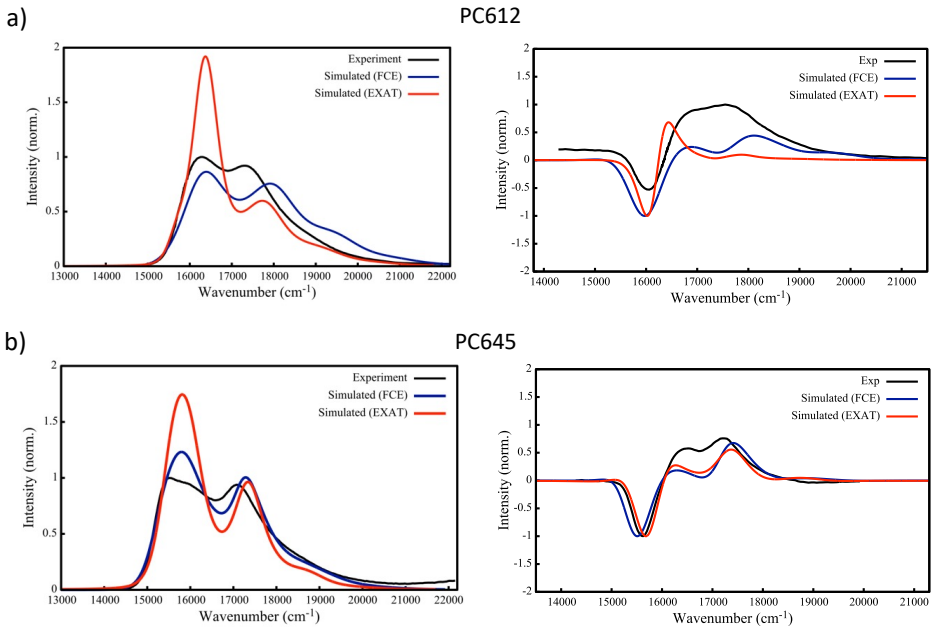


Figure A1.12. Experimental[188–190] and simulated absorption (left) and circular dichroism (right) spectra of a) PC612 and b) PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-OPT geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 2180 cm^{-1} (EXAT PC612), 1540 cm^{-1} (FCE PC612), 2300 cm^{-1} (EXAT PC645) and 2080 cm^{-1} (FCE PC645). Static disorder was modeled by applying a common σ = 100 cm^{-1}.

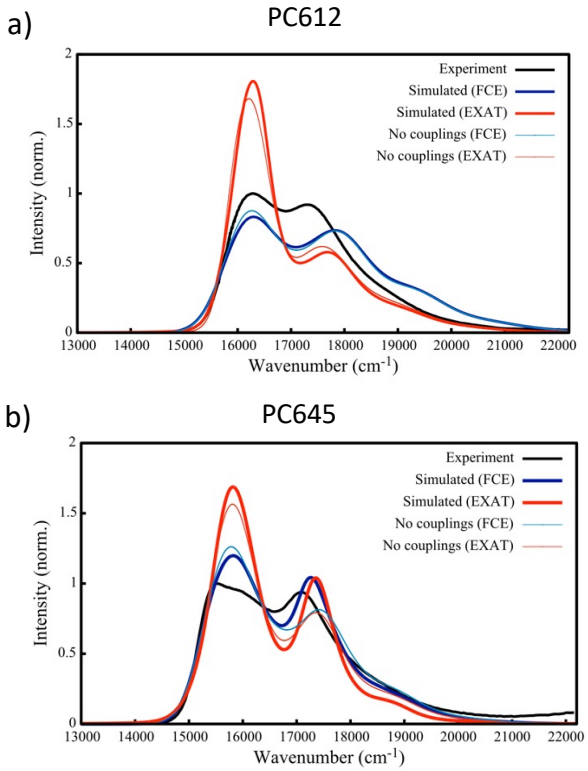


Figure A1.13. Comparison of experimental[188–190] and simulated absorption spectra of a) PC612 and b) PC645 cryptophyte antenna complexes obtained including or neglecting electronic couplings between bilin pigments. Simulations are based on QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-BOMD geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands (couplings/no couplings values are indicated): 1600/1500 cm^{-1} (EXAT PC612), 1000/980 cm^{-1} (FCE PC612), 1700/1640 cm^{-1} (EXAT PC645) and 1480/1500 cm^{-1} (FCE PC645). Static disorder was modeled by applying a common σ = 100 cm^{-1}.

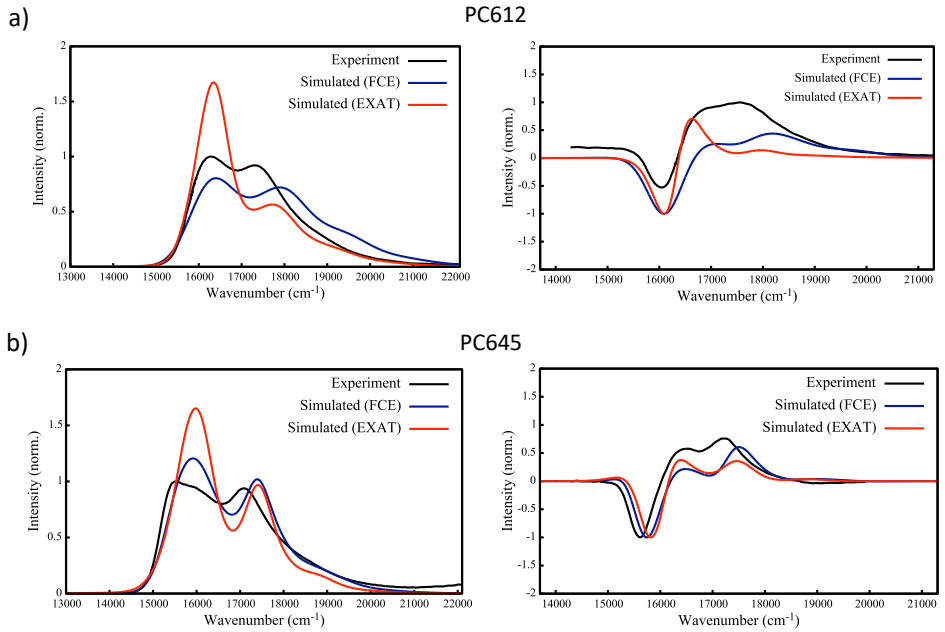


Figure A1.14. Experimental[188–190] and simulated absorption (left) and circular dichroism (right) spectra of a) PC612 and b) PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-BOMD geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 1560 cm^{-1} (EXAT PC612), 920 cm^{-1} (FCE PC612), 1620 cm^{-1} (EXAT PC645) and 1400 cm^{-1} (FCE PC645). Static disorder based on the MD-OPT σ values reported for each bilin in Table S3.

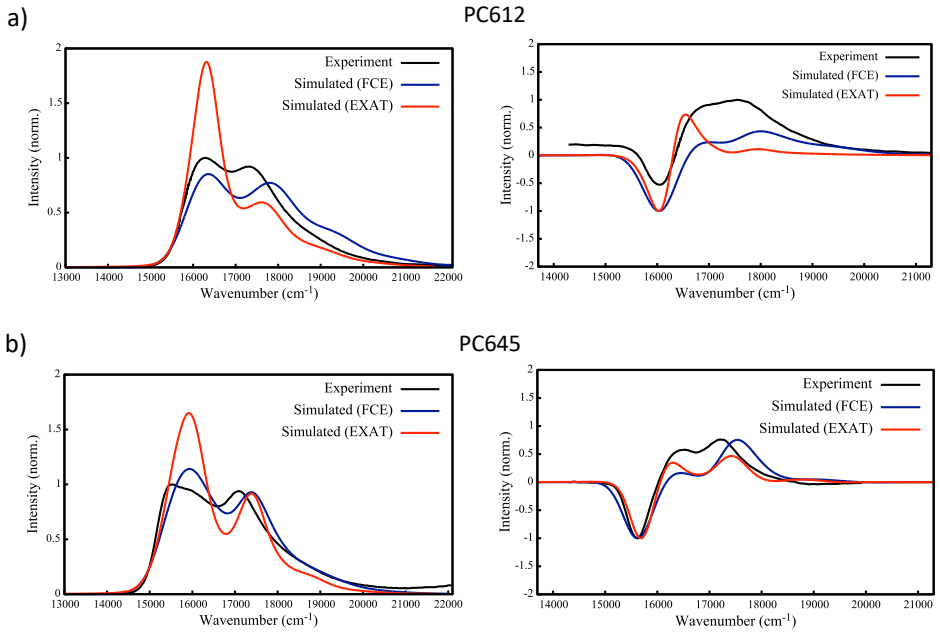


Figure A1.15. Experimental[188–190] and simulated absorption (left) and circular dichroism (right) spectra of a) PC612 and b) PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-OPT geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 2320 cm^{-1} (EXAT PC612), 1660 cm^{-1} (FCE PC612), 2260 cm^{-1} (EXAT PC645) and 1980 cm^{-1} (FCE PC645). Static disorder based on the MD-OPT σ values reported for each bilin in Table S3.

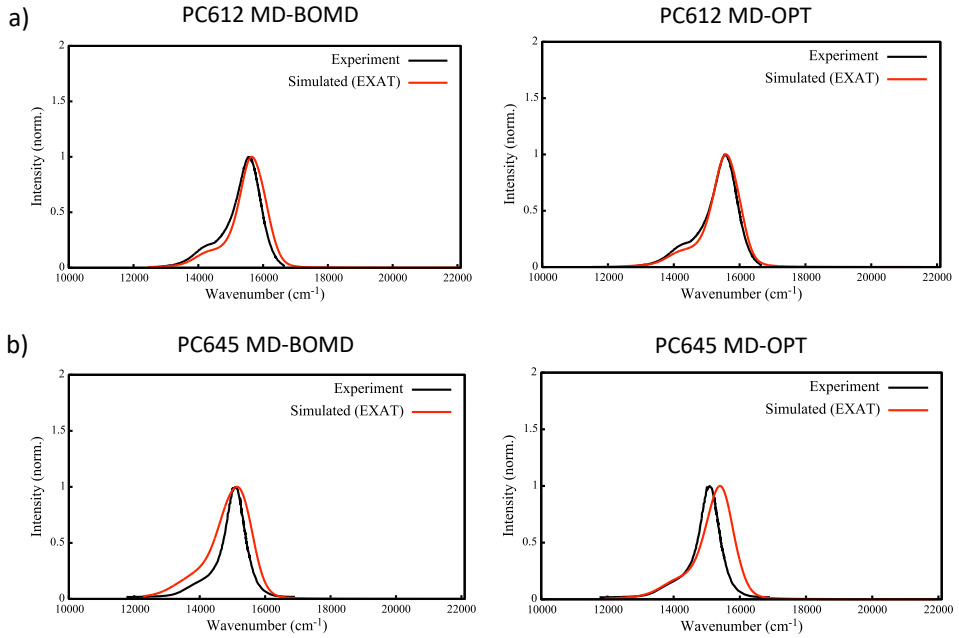


Figure A1.16. Experimental[188–190] and simulated fluorescence spectra of a) PC612 and b) PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-BOMD (left) and MD-OPT (right) geometries using the EXAT code to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 1620 cm^{-1} (PC612 MD-BOMD), 2480 cm^{-1} (PC612 MD-OPT), 1740 cm^{-1} (PC645 MD-BOMD) and 2100 cm^{-1} (PC645 MD-OPT). Static disorder based on the MD-OPT σ values reported for each bilin in Table S3.

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fmolb.2023.1268278

Check for updates

OPEN ACCESS

EDITED BY
Carla Molteni,
King's College London, United Kingdom

REVIEWED BY
Francesco Buda,
Leiden University, Netherlands
Riccardo Nifosì,
National Research Council (CNR), Italy

*CORRESPONDENCE
Carles Curutchet,
carles.curutchet@ub.edu

RECEIVED 28 July 2023
ACCEPTED 30 August 2023
PUBLISHED 18 September 2023

CITATION
Ozaydin B and Curutchet C (2023), Unraveling the role of thermal fluctuations on the exciton structure of the cryptophyte PC612 and PC645 photosynthetic antenna complexes.
Front. Mol. Biosci. 10:1268278.
doi: 10.3389/fmolb.2023.1268278

COPYRIGHT
© 2023 Ozaydin and Curutchet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Unraveling the role of thermal fluctuations on the exciton structure of the cryptophyte PC612 and PC645 photosynthetic antenna complexes

Beste Ozaydin[1,2] and Carles Curutchet[1,2]*

[1]Departament de Farmàcia i Tecnologia Farmacèutica, i Fisicoquímica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona (UB), Barcelona, Spain, [2]Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona (UB), Barcelona, Spain

Protein scaffolds play a crucial role in tuning the light harvesting properties of photosynthetic pigment-protein complexes, influencing pigment-protein and pigment-pigment excitonic interactions. Here, we investigate the influence of thermal dynamic effects on the protein tuning mechanisms of phycocyanin PC645 and PC612 antenna complexes of cryptophyte algae, featuring closed or open quaternary structures. We employ a dual molecular dynamics (MD) strategy that combines extensive classical MD simulations with multiple short Born-Oppenheimer quantum/molecular mechanical (QM/MM) simulations to accurately account for both static and dynamic disorder effects. Additionally, we compare the results with an alternative protocol based on multiple QM/MM geometry optimizations of the pigments. Subsequently, we employ polarizable QM/MM calculations using time-dependent density functional theory (TD-DFT) to compute the excited states, and we adopt the full cumulant expansion (FCE) formalism to describe the absorption and circular dichroism spectra. Our findings indicate that thermal effects have only minor impacts on the energy ladder in PC612, despite its remarkable flexibility owing to an open quaternary structure. In striking contrast, thermal effects significantly influence the properties of PC645 due to the absence of a hydrogen bond controlling the twist of ring D in PCB β82 bilins, as well as the larger impact of fluctuations on the excited states of MBV pigments, which possess a higher conjugation length compared to other bilin types. Overall, the dual MD protocol combined with the FCE formalism yields excellent spectral properties for PC612 and PC645, and the resultant excitonic Hamiltonians pave the way for future investigations concerning the implications of open and closed quaternary structures on phycocyanin light harvesting properties.

KEYWORDS

photosynthetic light harvesting, pigment-protein complexes, cryptophytes, multiscale QM/MM models, exciton states, molecular dynamics

1 Introduction

Photosynthetic organisms rely on specialized pigment-protein antenna complexes to capture sunlight. The absorbed energy then undergoes electronic energy transfer (EET) steps, eventually reaching reaction centers with high quantum efficiencies (Croce and van Amerongen, 2014), despite the considerable diversity in antenna complexes in terms of size, structure and pigment composition (Curutchet and Mennucci, 2017; Mirkovic et al., 2017).

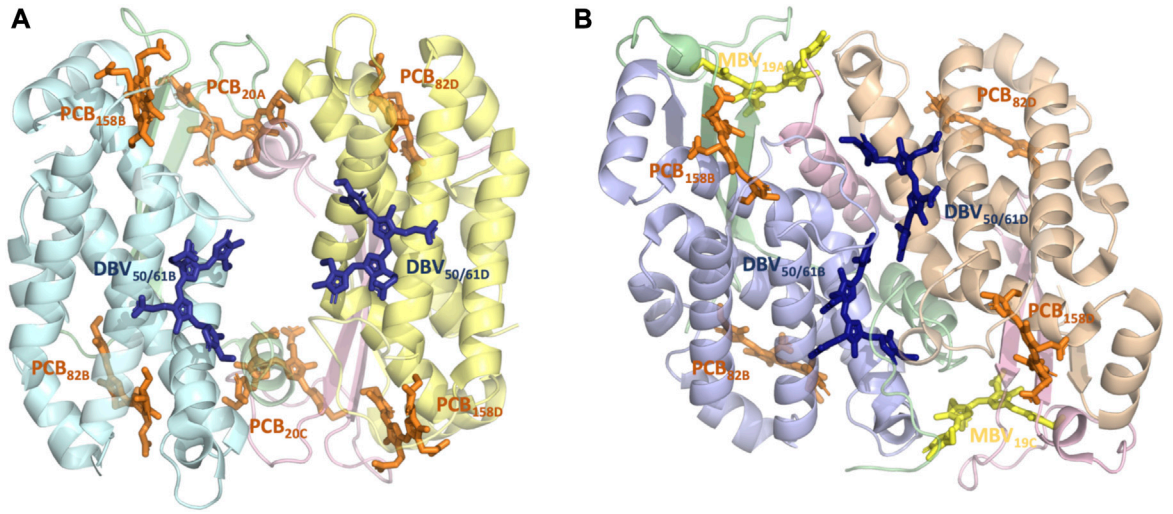


FIGURE 1
Structure and pigment composition of the phycocyanin cryptophyte antenna complexes **(A)** PC612 and **(B)** PC645.

Cryptophyte algae represent a unique group of photosynthetic organisms that originated from secondary endosymbiosis between an unknown eukaryotic host and a red algal symbiont. They replaced the primary red algal antenna, the phycobilisome, with a single antenna composed of two highly conserved β subunits originating from the phycobilisome and two more divergent α subunits, whose origin was largely unknown until recently, when they were linked to scaffolding proteins related to phycobilisome linker proteins (Greenwold et al., 1902; Rathbone et al., 2021). Remarkably, this algae display optimal photosynthetic activities at very low light conditions in marine and freshwater environments (Scholes et al., 2012). Despite this capacity, the pigments in these antenna complexes exhibit remarkable interpigment separations compared to the main antennae of other alga, bacteria or higher plants. Interestingly, over the last decade, a novel "open" quaternary structure of cryptohpyte antennae was discovered in various species of the *Hemiselmis* lineage. This structure involves the insertion of a single aspartic acid residue into the α subunit sequence, leading to a ~73° rotation of the two αβ protomers, which translates into a significant decrease in the electronic interaction between central bilin pigments. Intriguingly, this structural change has also been linked to a decrease in observed quantum beatings on 2D electronic spectra, with potential implications in the involvement of vibronic coherence in the EET mechanisms (Harrop et al., 2014).

Beyond the switch between open and closed structures and variations in bilin composition, sequence variations in the α subunits provide further spectral tuning mechanisms that allow cryptophytes to adapt to different environments, as suggested by recent X-ray crystallography and multiscale simulations (Corbella et al., 2019; Michie et al., 2023). However, quantitative studies on the impact of α subunit sequences and quaternary structures displayed by different species on the excitonic structure and light harvesting mechanisms of cryptophyte antennae have been impeded by uncertainties in the precise location of the exciton states, especially regarding the lowest-energy ones in closed phycocyanins like PC630 and PC645 (Mirkovic et al., 2007; Lee et al., 2017).

Recently, we clarified the protonation properties of different bilin pigments present in cryptophyte complexes, including 15,16-dihydrobiliverdin (DBV), phycocyanobilin (PCB), mesobiliverdin (MBV) and phycoerythrobilin (PEB), and investigated the exciton structure in PC577, PC612, PC630, and PC645 by multiscale polarizable quantum/molecular mechanical (QM/MMPol) calculations (Corbella et al., 2018; Corbella et al., 2019). Underlying inaccuracies related to the choice of input geometries to be used in excited state calculations however severely hampers these protocols (Dreuw et al., 2010; Lee et al., 2017; Corbella et al., 2019; Cupellini et al., 2019). Inclusion of thermal effects can be achieved by performing classical molecular dynamics (MD) simulations of the protein complex, followed by post-processing to compute excite states with QM/MM models (Curutchet et al., 2013). Alternatively, QM/MM geometry optimization of the pigments can be done, which drastically improved results in PC577, PC612, PC630 and PC645 complexes, although at the price of neglecting thermal effects (Corbella et al., 2019). Indeed, important discrepancies with experiment remained, like unbalanced descriptions of the relative intensity of low and high-energy bands in PC577 and PC612, or a strong underestimation of the MBV bands in PC630 and PC645, which points to the impact of thermal effects and the limits of rationalizing spectral properties in terms of X-ray crystal structures (Curutchet et al., 2013; Corbella et al., 2019; Michie et al., 2023).

In this study, we investigate the exciton properties of the *Hemiselmis virescens* PC612 and the *Chroomonas* PC645 antenna complexes shown in Figure 1 using a dual-MD multiscale protocol based on the combination of classical MD and Born-Oppenheimer (BOMD) simulations based on density functional theory (DFT). QM/MM MD simulations allow to describe with high accuracy the spectral density of electron-vibrational coupling in photosynthetic complexes and thus allow an accurate treatment of thermal effects (Rosnik and Curutchet, 2015; Blau et al., 2018; Loco et al., 2018; Maity et al., 2020; Maity et al., 2021a; Maity et al., 2021b; Sarngadharan et al., 2022). We

also compare this approach with an alternative protocol based on multiple QM/MM geometry optimizations along the trajectories. Moreover, we adopt the full Cumulant Expansion (FCE) formalism to describe the absorption and circular dichroism spectra (Ma and Cao, 2015; Cupellini et al., 2020), which allows the inclusion of non-Markovian and non-secular effects that are neglected in modified Redfield theory. Overall, this new protocol yields absorption spectra with unprecedented accuracy for PC612 and PC645, providing an accurate excitonic Hamiltonian for future investigations addressing the implications of open and closed quaternary structures on phycocyanin light harvesting properties.

2 Materials and methods

2.1 Simulated systems, MD simulations and geometry optimizations

We based our simulations on high-quality crystal structures obtained for *Chroomonas* PC645 (PDB 4LMS, resolution 1.35 Å) and *Hemiselmis virescens* PC612 (PDB 4LM6, resolution 1.8 Å) reported by Curmi and co-workers, with the PC numbering indicating their maximum absorption wavelength (Harrop et al., 2014). The preparation of the systems and the details of the classical MD simulations are described in Corbella et al. (2018), Corbella et al. (2019). Both complexes have four polypeptide chains, but PC612 is organized as an $(\alpha\beta)_2$ homodimer, while PC645 displays a $(\alpha_L\beta)\cdot(\alpha_S\beta)$ organization, where α_L and α_S indicate long and short α subunits, respectively (Michie et al., 2023). Additionally, PC612 features an open quaternary structure due to a ~73° rotation of the αβ protomers, whereas PC645 exhibits a closed configuration more commonly found in cryptophyte antennae. In terms of pigment composition, PC612 contains six PCBs and two DBVs, while in PC645, two of the PCBs are replaced by MBVs. These pigments are covalently linked to a Cys, except for DBVs, which involve two covalent links with Cys residues. Therefore, pigments are denoted as PCB_{82D} or $DBV_{50/61B}$, A/C and B/D indicating α and β chains, respectively, with numbering referring to the Cys attachment residues in that chain. All bilins were modelled with a fully protonated tetrapyrrole backbone and anionic propionic side chains, while all amino acids were considered in their standard protonation state, except His21 C in PC645.

We started from the 200 ns classical MD simulations reported in Ref. Corbella et al. (2019), performed using the ff14SB, (Maier et al., 2015), TIP3P (Jorgensen et al., 1983) and GAFF (Wang et al., 2004) force fields for the protein, water and bilin chromophores, respectively. Classical *NVT* simulations at 300 K were extended to a total of 1 μs, using a time step of 2 fs, the SHAKE algorithm, periodic boundary conditions, the particle-mesh Ewald approach and a non-bonded cutoff of 8 Å. From the last 500 ns, we extracted 34 frames at 14 ns intervals to be used as starting conditions for BOMD QM/MM simulations. Then, from each frame we run eight independent 1 ps BOMD simulations, where one of the bilin pigments was described at the B3LYP/6-31G level of theory, using a 0.5 fs time step. From the last 500 fs of BOMD trajectories, we extracted 25 frames at regular time intervals to be used as input geometries in QM/MMPol excited state calculations. This protocol is referred to as MD-BOMD. All runs were performed with the Amber20 (Case et al., 2020)—Gaussian 16 interface (Frisch et al., 2016).

On the other hand, QM/MM optimizations were independently performed for each chromophore geometry starting from the 34 frames extracted from the classical trajectories. We adopted the ONIOM substractive scheme with electrostatic embedding as implemented in Gaussian (Vreven et al., 2006). In each optimization, a given pigment and selected amino acids interacting with them were fully relaxed and described at the B3LYP/6-31G(d) level of theory, while the rest of the system was kept frozen and described with the classical force field adopted in the MD. For PCBs and DBVs in both complexes, we included the Asp or Glu side chains coordinating the central rings, whereas for MBVs in PC645 we included selected amino acids (Asn22 A for MBV19A and His21 C and Glu25 C for MBV19C). QM/MM boundaries were defined at the residue-residue and Cys-bilin bonds using the link atom scheme (Senn and Thiel, 2009). The optimized geometries were used as inputs in subsequent QM/MMPol excited state calculations. This protocol is denoted as MD-OPT.

2.2 QM/MMPol calculations

Excited state calculations were performed using the QM/MMPol model (Curutchet et al., 2009) in its time dependent density functional theory (TD-DFT) linear response formulation. In the MMPol model, the environment is described using a polarizable force field based on the induced point dipole model, which implies assignment of point charges and isotropic polarizabilities to the environment atoms, and accounts full mutual polarization effects among the QM and MM regions. QM/MMPol calculations on structures extracted from BOMD trajectories and ONIOM optimizations were performed at the TD-CAM-B3LYP/6-31G(d) level of theory (Yanai et al., 2004). Previous work showed that different global and range-separated hybrid functionals with different degrees of exact exchange (M06, M06-2X and ωB97XD) led to similar results in phycobiliproteins (Corbella et al., 2019). The MMPol region was described using the Amber pol12 AL parameters (Wang et al., 2011a; Wang et al., 2011b), and QM/MM boundaries in the bilin-Cys bonds were treated using the link atom scheme (Senn and Thiel, 2009). Atomic charges for water and bilins, consistent with the Amber pol12 AL polarizabilities, were derived as described in Ref. Corbella et al. (2019). Explicit polarization was limited to residues within a cutoff radius of 12 Å from the QM heavy atoms (MMPol region), whereas residues up to a 35 Å were also included but adopting the additive force field (MM region). MMPol calculations were performed using a locally modified development version of the Gaussian package (Frisch et al., 2010).

2.3 Modelling of steady-state spectra

Simulations of absorption (OD), circular dichroism (CD) and fluorescence (FLU) spectra were performed using the EXAT (Jurinovich et al., 2018) and the FCE codes (Arpin et al., 2015; Cupellini et al., 2020). EXAT simulations were based on modified

Redfield theory, whereas FCE ones include non-Markovian and non-secular effects as described in Refs. Ma and Cao (2015), Cupellini et al. (2020). We used the following excitonic Hamiltonian describing the multichromophoric system:

$$\hat{H} = \sum_{n=1}^{N} \varepsilon_n |n\rangle\langle n| + \sum_{n \neq m}^{N} V_{nm} |n\rangle\langle m| \quad (1)$$

Where N is the number of interacting chromophores, ε_n is the excitation energy of the nth chromophore and V_{nm} is the coupling between the nth and mth chromophores. Exciton states $|k\rangle$ were obtained from diagonalization of the excitonic Hamiltonian:

$$\hat{H} = \sum_{k=1}^{N} \varepsilon_k |k\rangle\langle k|; \; |k\rangle = \sum_{n=1}^{N} C_n^k |n\rangle \quad (2)$$

where ε_k is the energy of the kth exciton state and C_n^k describes the participation of the nth chromophore to the kth exciton state.

We used the site energies averaged for each pigment over the 850 or 34 frames calculated using the dual quantum/classical MD (MD-BOMD set) or the ONIOM optimization protocols (MD-OPT set), respectively, and results were also compared to the energies estimated previously based on classical MD (MD set) or a single optimization performed on the crystal structure (Crystal-OPT set) in Ref. Corbella et al. (2019). Electronic couplings, transition dipole moments and pigment center coordinates were taken from previous QM/MMPol/ddCOSMO calculations based on the Crystal-OPT protocol and include dielectric screening effects exerted by the surrounding environment. Beyond the DBV central pair in PC645, the pigment pairs in these complexes are not strongly coupled, and the adoption of the MD and Crystal-OPT sets of couplings and dipoles led to negligible differences in predicted spectra. Even the complete neglect of couplings leads to small changes, as shown in Supplementary Figure S13. For this reason, we do not recompute the couplings along the many MD-OPT optimizations. We also adopt the transition dipole moments from the Crystal-OPT set for consistency, but these values are very similar to those obtained by the MD-OPT and MD-BOMD protocols, as shown in Supplementary Table S4, so this choice leads to negligible changes in predicted spectra. On the other hand, BOMD simulations are performed for single pigments, so recalculation of electronic couplings on those geometries would lead to an unbalanced description of the pigment pair.

Electronic-vibrational coupling was accounted for using spectral densities for the intramolecular part calculated using the Vertical Gradient approach, while the continuous part due to slow environmental motions was added *a posteriori* using an overdamped Brownian oscillator with parameters adjusted to reproduce the Stokes shift, as described in Ref. Corbella et al. (2019). Finally, Static disorder was modeled by averaging realizations of the spectra over a random distribution of site energies characterized by a given standard deviation σ, which was adjusted to reproduce the broadening of the experimental emission lineshapes for each complex or estimated from the variation of site energies sampled from the MD-Opt set.

3 Results and discussion

3.1 Structural flexibility of open and closed phycocyanins

Recent studies have suggested that the sequence diversity of α subunits plays a significant role in determining the chromophore conformations observed in X-ray crystal structures and, consequently, the spectral properties of cryptophyte complexes (Michie et al., 2023). To investigate the impact of thermal effects and quaternary structure on the light-harvesting properties of PC612 and PC645 complexes, we conducted extensive classical and QM/MM MD simulations. These simulations allowed us to analyze and compare the structural flexibility of the open PC612 and closed PC645 complexes.

Figure 2 shows the root-mean-square fluctuations (RMSF) calculated along the classical MD trajectories for the α and β subunits of PC612 and PC645. The open arrangement of the two αβ protomers in PC612 leads to higher flexibility in all subunits compared to PC645, which has a closed structure. Notably, in the β subunits B and D of PC612, there is increased flexibility in the N-terminal regions near PCB_{20A} and PCB_{20C}, where they interact with Tyr18 B and Tyr18 D through their propionic groups. In contrast, the C-terminal regions of the α subunits A and C in PC612 appear more rigid, despite the open quaternary arrangement. However, regions containing residues ~20–40 in chains A and C, starting at the anchoring points of PCB_{20A} and PCB_{20C} (Cys20 A

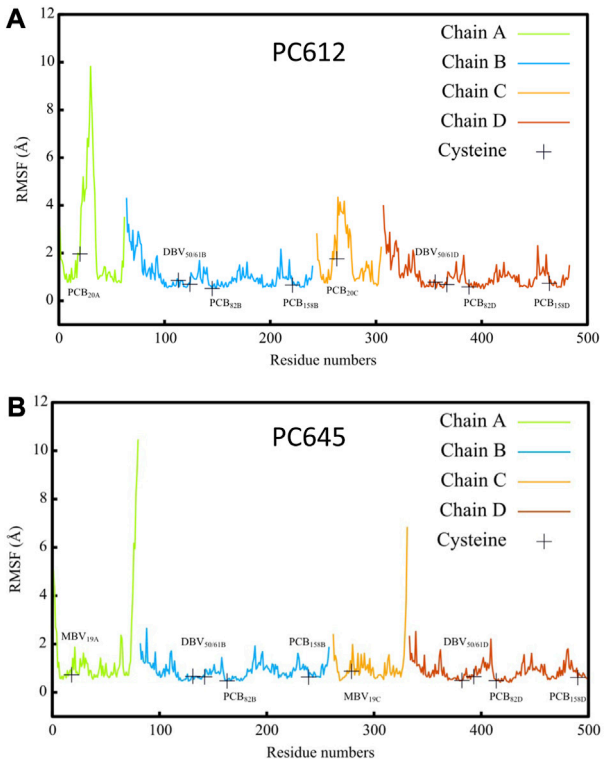


FIGURE 2
Root mean square fluctuation (RMSF) of atomic positions calculated along classical MD trajectories for α A/C and β B/D subunits of antenna complexes **(A)** PC612 and **(B)** PC645.

and Cys20 C), show considerably more flexibility in PC612 compared to PC645. In PC612, these regions form loops that mostly interact with PCB_{20A} and PCB_{20C}, partially shielding them from the solvent. This disposition is similar to that observed in PC645, which has the closed structure. However, in PC612, PCB_{20A} and PCB_{20C} are considerably more exposed to the solvent than MBV_{19A} and MBV_{19C} in PC645. Consequently, increased flexibility of PCBs in the open PC612 complex, likely due to higher solvent exposure, leads to more flexible protein loops. The RMSF values corresponding to the pigments in PC612 and PC645 (Supplementary Table S1 of the Supplementary Material), confirm that increased fluctuations in the open protein PC612 result in more flexible pigments, particularly PCB_{20} bilins in α subunits, which display significantly larger fluctuations than the MBVs in PC645.

3.2 Spectral tuning of bilins by the protein scaffold

Next, we investigated how thermal fluctuations and protein flexibility influence the electronic excited states of the bilin pigments. Figure 3 presents the site energies (uncoupled transition energies) of the pigments computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations for PC612 and PC645. We compare the results obtained from classical MD geometries (MD set) or a single optimization of the bilins in the protein scaffold based on the crystal structure (Crystal-OPT set) (Corbella et al., 2019) with those obtained from multiple geometry optimizations of the pigments performed along classical MD trajectories (MD-OPT set) and short BOMD DFT-based simulations (MD-BOMD set). In addition, in the Supplementary Tables S2, S4 and Supplementary Figures S4–S11 we provide the averaged values and the complete distributions of site energies and transition dipole moments obtained from both sets.

In PC612, both the MD-BOMD and MD-OPT methodologies yielded similar site energies, closely matching those obtained using the Crystal-OPT strategy. This indicates that the energy ladder of PC612 is well explained by the crystal structure and the pigment-protein interactions, which constrain the bilin pigments to specific conformations. On the other hand, in PC645, the MD-BOMD and MD-OPT protocols significantly modify the picture provided by previous studies. While the energies of DBV and PCB_{158} pairs agree with previous values based on the crystal, the energetic position of PCB_{82} and MBV_{19} pairs is significantly modified, suggesting a crucial role of thermal effects. In terms of transition dipole

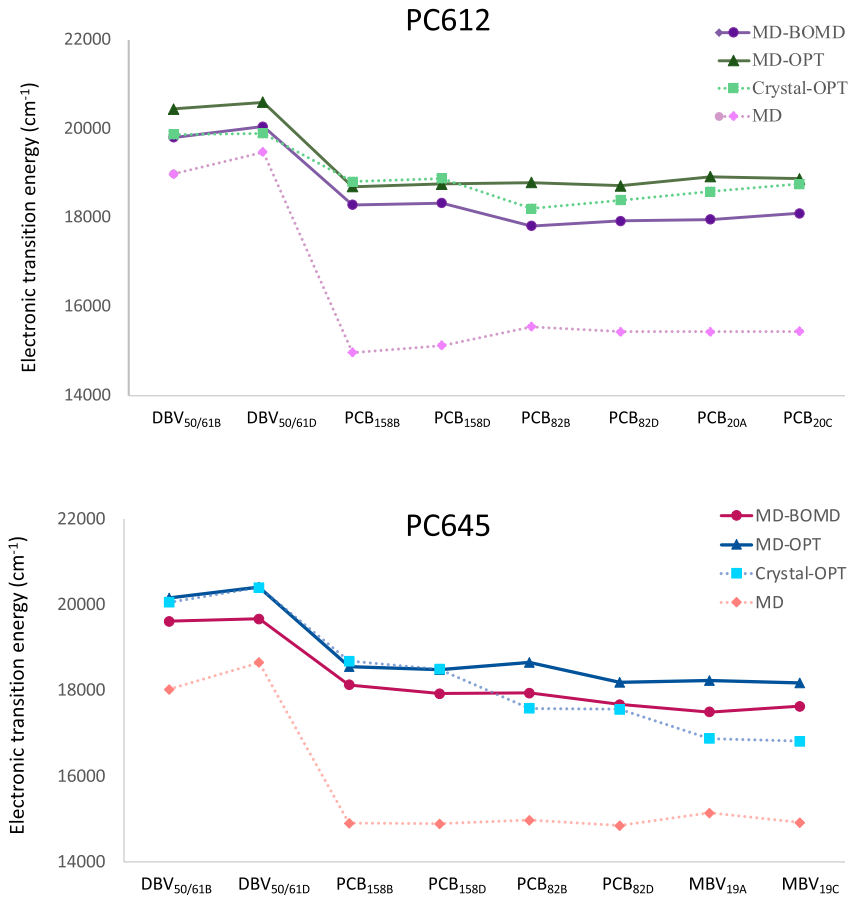


FIGURE 3
Electronic transition energies computed at the QM/MMPol TD-CAM-B3LYP/6-31G(d) level of theory for cryptophyte complexes using geometries obtained using the MD-BOMD and MD-OPT protocols compared to previous results based on classical MD or a single optimization of the bilins in the crystal structure (Corbella et al., 2019): **(A)** PC612 and **(B)** PC645.

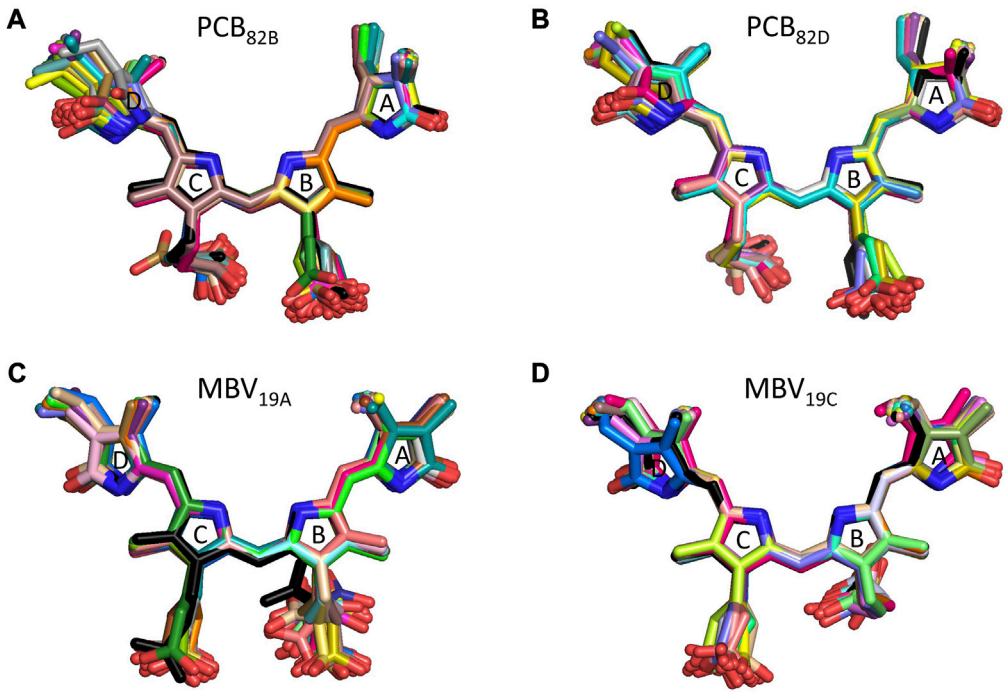


FIGURE 4
Alignment of PC645 pigment structures in the crystal structure (black) and optimized along the MD simulation (colors): **(A)** PCB_{82B}, **(B)** PCB_{82D}, **(C)** MBV_{19A} and **(D)** MBV_{19C}.

TABLE 1 Averaged dihedral angles and their standard deviations computed from DFT-based BOMD simulations characterizing the twists among pyrrole units in the pigments of PC612. For comparison, the values corresponding to the crystal structure are provided in parentheses.

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$DBV_{50/61B}$	183 ± 9 (182)	32 ± 11 (20)	172 ± 10 (190)	182 ± 11 (168)	—	—
$DBV_{50/61D}$	178 ± 9 (178)	34 ± 11 (23)	181 ± 9 (186)	179 ± 9 (171)	—	—
PCB_{158B}	178 ± 6 (171)	−36 ± 14 (−20)	180 ± 7 (176)	181 ± 6 (178)	−32 ± 12 (−40)	178 ± 5 (183)
PCB_{158D}	177 ± 7 (171)	−39 ± 13 (−34)	184 ± 10 (179)	178 ± 8 (177)	−40 ± 11 (−43)	173 ± 9 (181)
PCB_{82B}	186 ± 7 (186)	−20 ± 11 (−24)	179 ± 9 (173)	187 ± 8 (189)	−33 ± 10 (−37)	170 ± 9 (175)
PCB_{82D}	186 ± 7 (189)	−18 ± 12 (−25)	177 ± 8 (174)	188 ± 9 (188)	−33 ± 11 (−38)	176 ± 8 (186)
PCB_{20A}	184 ± 6 (173)	11 ± 19 (−36)	184 ± 8 (190)	190 ± 10 (178)	−41 ± 11 (−33)	180 ± 9 (181)
PCB_{20C}	175 ± 8 (174)	−34 ± 15 (−37)	184 ± 10 (187)	179 ± 9 (179)	−40 ± 10 (−33)	176 ± 8 (186)

moments, however, both MD-OPT and MD-BOMD sets provide results very similar to the Crystal-OPT ones on both complexes, as shown in Supplementary Table S4. Thus, despite the considerable distributions of dipoles obtained, which are comparable to those observed in site energies, as shown in Supplementary Figures S4–S11, the averaged values turn to be very close in all cases to those based on the crystal structure.

To explain the impact of thermal effects in the site energies of the PCB_{82} and MBV_{19} pairs in PC645, it is interesting to note that comparison of the X-ray crystal structures of PC645 and a similar complex, PC630, showed that in the latter there is a hydrogen bond between ring D in PCB β82 chromophores and Gln6 in the α_L and α_S subunits, whereas in PC645 Gln6 is replaced by Leu5, and the lack of this hydrogen bond is the responsible for the red shift of the absorption spectra of PC645 compared to PC630 (Michie et al., 2023). Here, we show that beyond this analysis, thermal fluctuations of the protein environment are key to obtain a balanced description of β82 PCB chromophores and to explain the intriguing question of why α19 MBVs in PC645 have site energies very close to the PCBs, despite their larger degree of conjugation. Indeed, bilin excitation energies are strongly controlled by the conformation of the tetrapyrrole backbone, and hydrogen bonds and steric clashes with the protein scaffold can twist the bonds connecting pyrrole rings, thus decreasing the

TABLE 2 Averaged dihedral angles and their standard deviations computed from DFT-based BOMD simulations characterizing the twists among pyrrole units in the pigments of PC645. For comparison, the values corresponding to the crystal structure are provided in parentheses.

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
$DBV_{50/61B}$	190 ± 8 (195)	27 ± 11 (20)	190 ± 9 (187)	168 ± 8 (174)	—	—
$DBV_{50/61D}$	193 ± 8 (194)	33 ± 11 (34)	176 ± 9 (185)	187 ± 9 (179)	—	—
PCB_{158B}	176 ± 8 (174)	−37 ± 12 (−34)	182 ± 9 (181)	182 ± 8 (179)	−34 ± 11 (−27)	168 ± 9 (168)
PCB_{158D}	171 ± 7 (171)	−36 ± 10 (−35)	187 ± 8 (189)	179 ± 8 (172)	−34 ± 11 (−25)	168 ± 9 (168)
PCB_{82B}	183 ± 8 (187)	−24 ± 11 (−29)	179 ± 8 (181)	185 ± 9 (188)	27 ± 19 (−12)	189 ± 9 (201)
PCB_{82D}	182 ± 8 (191)	−23 ± 11 (−28)	179 ± 8 (179)	186 ± 8 (182)	11 ± 16 (−6)	188 ± 9 (203)
MBV_{19A}	182 ± 8 (184)	31 ± 11 (29)	181 ± 9 (185)	181 ± 8 (183)	−39 ± 10 (−48)	182 ± 3 (187)
MBV_{19C}	186 ± 8 (187)	29 ± 11 (23)	178 ± 9 (179)	184 ± 9 (183)	−43 ± 11 (−45)	175 ± 8 (176)

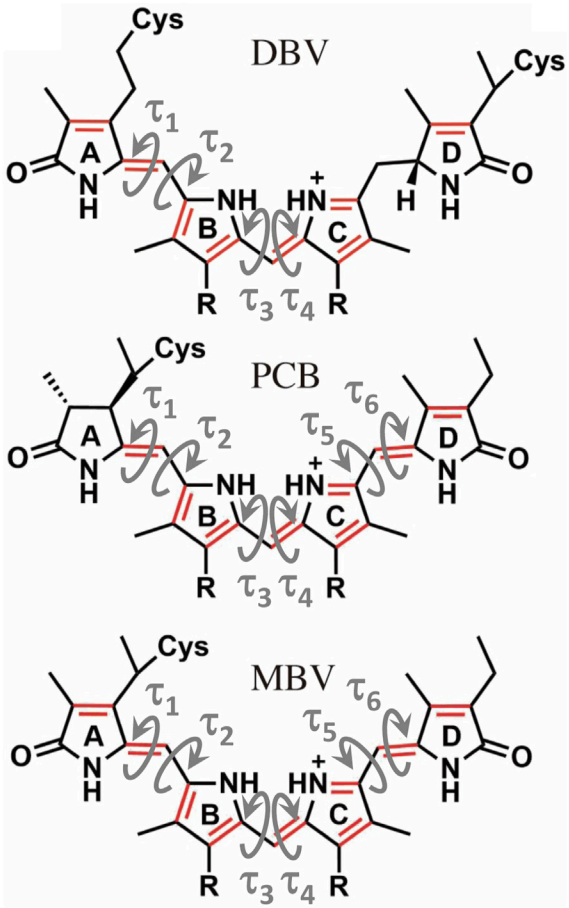


FIGURE 5
Structure of the linear tetrapyrrole arrangement of DBV, PCB and MBV bilin pigments indicating the key torsions that define the conformation of the conjugated region that characterizes the excited state.

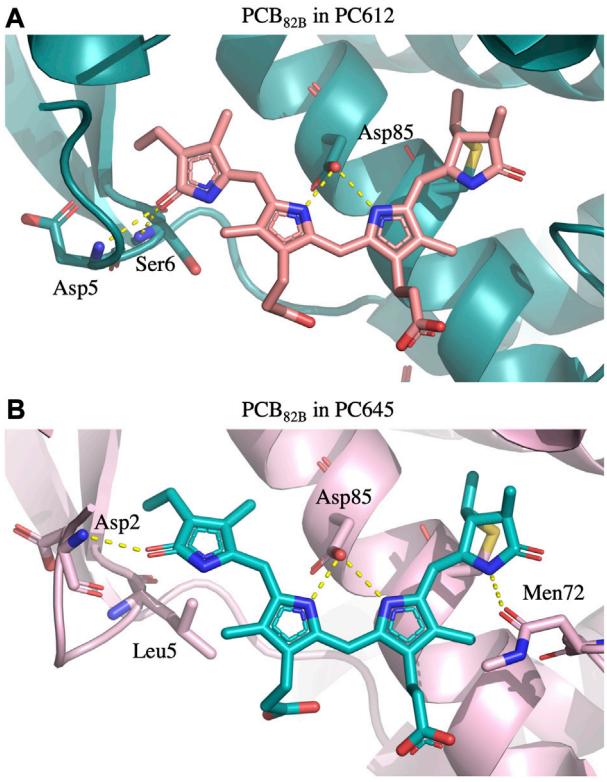


FIGURE 6
Hydrogen bonds established by PCB β82 bilins with the protein scaffold in **(A)** PC612 and **(B)** PC645. In PC612, the carbonyl group of ring D interacts with Asp5 and Ser6, stabilizing a considerable twist between rings C and D. In contrast, in PC645 the presence of Leu5 allows significantly larger fluctuations, which are not completely stabilized by the interaction with Asp2.

conjugation of the π system, which increases the transition energies of the $\pi \rightarrow \pi^*$ states.

To understand the large impact of thermal effects on β82 PCB and α19 MBVs energy levels in PC645, in Figure 4 we show an alignment of their structures as found in the crystal structure or optimized along the MD simulation following the MD-OPT protocol. In addition, in Table 1 and Table 2 we report the dihedral angles of the bonds connecting the bilin pyrrole rings (defined in Figure 5) obtained in the MD-BOMD calculations of PC612 and PC645 compared to the crystal references, as has been done in previous analyses of biliproteins either based on

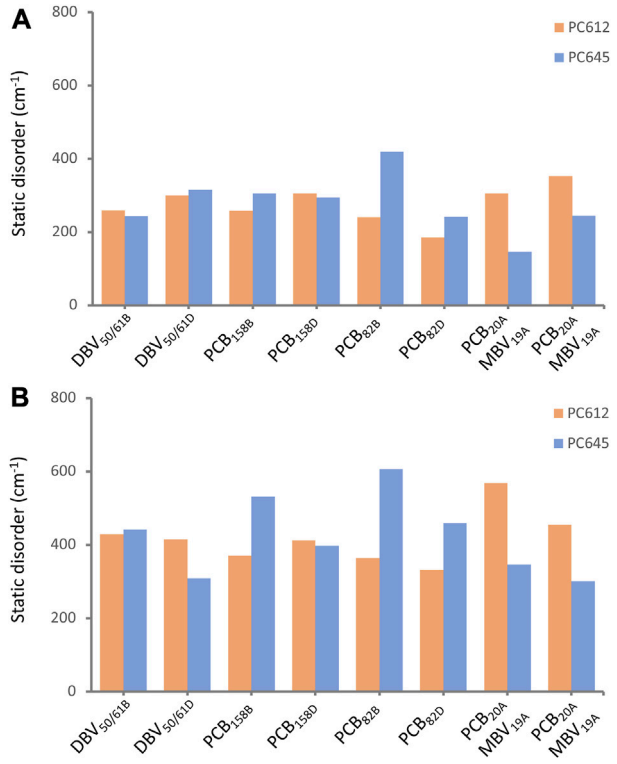


FIGURE 7
Static disorder (cm^{-1}) estimated from **(A)** MD-OPT and **(B)** MD-BOMD averaged transition energies computed along the classical MD trajectories of PC612 and PC645.

simulations or crystal structures (Curutchet et al., 2013; Michie et al., 2023).

Overall, Figure 4 indicates that there are no large conformational changes of the β82 PCB and α19 MBVs bilins in PC645 compared to the crystal reference, especially regarding the central B and C pyrrole rings. However, rings A and D and the propionic side chains display remarkable fluctuations compared to the crystal reference. Remarkable fluctuations are especially observed for ring D (shown at the left) in PCB_{82} bilins, which is confirmed by a standard deviation of ~16–19 degrees in the torsion τ_5 reported in Table 2, which is almost twice the value observed in other bilins. As previously discussed, in PC645 the nitrogen atom of ring D in PCB β82 bilins interacts with Leu5, missing the two hydrogen bonds with Ser6 and Asp5 found in PC612 shown in Figure 6. In PC645 there is indeed a hydrogen bond with Asp2, but its arrangement does not prevent considerable twists among the planes of C and D rings. Thus, in PC645 the lack of a hydrogen bond controlling the twist of ring D leads to larger fluctuations in PCB_{82} bilins that break the conjugation of the tetrapyrrole backbone, leading to a significant increase in the electronic transition energies of the $\pi \rightarrow \pi^*$ states, as shown in Figure 3. In MBVs, even stronger thermal effects are observed than in PCB_{82}s. In this case, however, the fluctuations of rings A and D are comparable, and the strong impact of thermal effects seems to be linked to the fact that MBVs have a higher conjugation length that extends to ring A, so thermal fluctuations in the twist of that ring necessarily lead to a stronger impact than that found for DBVs or PCBs. Indeed, beyond the case of PCB_{82}s in PC645 discussed above, Table 1 and Table 2 show similar standard deviations for the torsions of MBVs and the other bilins in PC612 and PC645, with somewhat larger values for PCB_{20} and PCB_{158} pigments in PC612, which can be explained by the overall larger flexibility of the open structure of this complex.

In addition to intramolecular fluctuations of the pigments, we analyzed how the quaternary structure of PC612 and PC645 and collective protein motions impact static disorder. Figure 7 shows an estimate of static disorder obtained as the standard deviation of site energies computed along the many MD-OPT optimized geometries. We also provide a similar analysis based on the energies averaged over different BOMD QM/MM simulations. In the latter, short BOMD runs necessarily translate into overestimated disorder values, but the comparison between both protocols shows similar trends that suggests they both capture similar physics in the complexes. Indeed, both analyses show a similar disorder for $DBV_{50/61}$ and PCB_{158} pairs in the two proteins, but opposite trends for PCB_{82} and the PCB/MBV α19/20 pairs. The increased flexibility and solvent exposure of PCB_{20} in the open structure of PC612 compared to MBV_{19} in PC645 seems to be the reason for the increased disorder. In contrast, in PC645, despite the closed quaternary structure leading to an overall more rigid protein, PCB_{82} pigments consistently display larger disorder compared to PC612. As discussed previously, PCB_{82} bilins in PC645 lack a key hydrogen bond with α_L and α_S subunits that lead to larger fluctuations in the twist of the bonds linking rings C and D, which explains such increased disorder. We note that our best estimates for disorder, shown in Figure 7, are larger than the value of 100 cm^{-1} we derived previously from a fit between experiment and simulations of absorption and emission spectra (Corbella et al., 2019). In the next section we discuss the reliability of the site energies and the disorder values in describing the optical spectra of the complexes.

3.3 Absorption and circular dichroism spectra

To validate the excitonic Hamiltonian derived for PC612 and PC645 based on the MD-BOMD and MD-OPT site energies, we simulated OD and CD spectra of the complexes and compared them to experiment. In Figure 8 we report our most accurate estimates, based on the dual classical/quantum MD strategy, the MD-BOMD protocol, which, in contrast to previous studies, allows to accurately account for thermal effects by avoiding the geometry mismatch problem related to the use of MD trajectories based on classical force fields (Corbella et al., 2019). In addition, here we adopt a more rigorous theory to simulate excitonic spectra, the full Cumulant Expansion (FCE) formalism (Ma and Cao, 2015; Cupellini et al., 2020), which allows including non-Markovian and non-secular effects neglected in modified Redfield theory. To gauge the impact of these effects, we also report spectra based on modified Redfield theory obtained using the EXAT (Jurinovich et al., 2018) code.

The performance of the MD-BOMD set of site energies, combined with the FCE formalism, in describing the OD spectrum is astonishing. For both complexes, we obtain an excellent agreement with experiment, and our simulations reproduce with unprecedented accuracy the position of the two

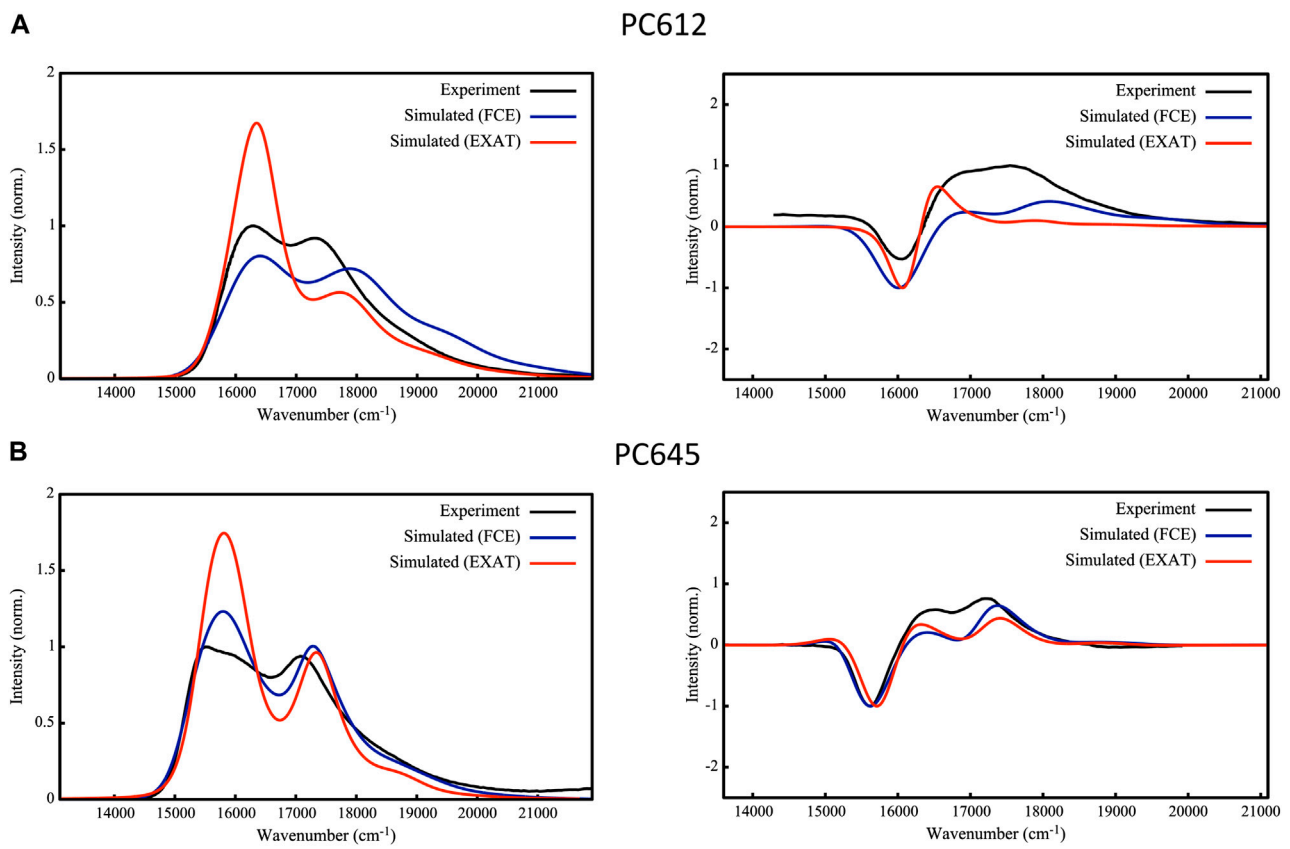


FIGURE 8
Experimental (Mirkovic et al., 2007; Harrop et al., 2014; Arpin et al., 2015) and simulated absorption (left) and circular dichroism (right) spectra of **(A)** PC612 and **(B)** PC645 cryptophyte antenna complexes. Simulations are based on energies and couplings computed from QM/MMPol TD-CAM-B3LYP/6-31G(d) calculations performed on MD-BOMD geometries using the EXAT or FCE codes to calculate the spectra. We applied the following shifts to simulated spectra to fit the experimental bands: 1,600 cm^{-1} (EXAT PC612), 1,000 cm^{-1} (FCE PC612), 1,700 cm^{-1} (EXAT PC645) and 1,480 cm^{-1} (FCE PC645). Static disorder was modeled by applying a common σ = 100 cm^{-1}.

main peaks, their relative intensities, and the overall broadening in the spectra. This is drastic improvement over previous simulations reported on these systems (Corbella et al., 2019), in which the absence of thermal effects led, for example, to a strong underestimation of MBV energies in PC645, too large differences among PCB_{158} and PCB_{82} bands, and overestimated DBV energies. For PC612, previous reports were more accurate, but both peaks were too close leading to an overall large band. In terms of CD, again the MD-BOMD simulations led to the correct positions of the three bands observed experimentally, which are instead obscured in the OD spectra dominated by two bands. Interestingly, the MD-OPT protocol also leads to accurate spectra very similar to that obtained using the MD-BOMD benchmark protocol, as shown in Supplementary Figure S12 of the Supporting Information. Because the multiple geometry optimizations in this protocol do not allow to fully account for thermal effects associated to intramolecular vibrations of the pigments, we conclude that accounting for the collection of protein conformations is the key aspect to be accounted for, which is not included in analysis limited to the crystal structure.

As we extensively discuss in the previous section, thermal effects seem to be key to correctly describe the energies of PCB_{82} and MBV bilins in PC645. Accounting for non-Markovian and non-secular effects through the FCE formalism however seems to be also important to correctly describe the relative intensities of the two main absorption peaks in OD spectra, as shown in Figure 8, where EXAT simulations overestimate the intensity of the low energy band. A similar improvement is also obtained in the relative intensities of CD bands, especially in PC612. In addition, whereas for PC612 FCE and EXAT lead to similar OD lineshapes in the blue edge of the spectrum, for PC645 the FCE formalism provides a more accurate lineshape. This can be ascribed to a better description of the coupling between vibrations and the exciton states delocalized on the central $DBV_{50/61B}$—$DBV_{50/61D}$ dimer in PC645, characterized by a strong electronic coupling of 238 cm^{-1} (Corbella et al., 2019). In the open quaternary structure of PC612 this coupling is attenuated to a 16 cm^{-1}, leading to mostly localized states. This is clearly visible in Supplementary Figure S13, where we compare the OD spectra to simulations in which all couplings are zeroed. Whereas neglecting excitonic couplings do not change much PC612 spectra, characterized by localized states, the impact on PC645 is large, namely, in the high-energy band characterized by the $DBV_{50/61B}$—$DBV5_{0/61D}$ strongly coupled dimer.

Finally, we recall that the theoretical estimates of static disorder shown in Figure 7 are larger than the value of 100 cm^{-1} used in the simulations reported in Figure 8, previously fitted by comparison to absorption and emission experimental spectra (Corbella et al., 2019). In Supplementary Figures S14–S16 in the Supplementary Material, we report OD, CD and FLU spectra obtained adopting our theoretical estimates of disorder based on MD-OPT simulations. In all cases the OD spectra displays minor changes. For CD and FLU, in PC612 we also obtain a reasonable broadening. However, the CD and FLU spectra of PC645 is slightly too broad, suggesting our methodology still has limitations to estimate this important parameter, probably because simulations are based on MDs spanning a limited μs timescale.

4 Conclusion

In this study, we investigated the role of thermal effects on the spectral tuning mechanisms of PC645 and PC612 antenna complexes from cryptophyte algae. To accurately account for thermal effects arising from protein (slow) and pigment (fast) structural fluctuations, we have applied a dual MD strategy to determine the site energies that combines classical MD simulations to sample the conformational space of the complex on the μs timescale, with short Born-Oppenheimer QM/MM simulations on the ps timescale, which allowed us to accurately describe the impact of pigment's vibrations on the relevant excited states. We have also investigated the performance of an alternative method, based on multiple QM/MM geometry optimizations of the pigments performed along the classical MD trajectories. The resulting structures of both protocols were then used to evaluate the excited states of the pigments using polarizable QM/MM TD-DFT calculations, and the excitonic Hamiltonians were used in subsequent spectral simulations of absorption and circular dichroism based on the FCE formalism. Interestingly, we found that thermal effects only lead to small corrections to the energy ladder in PC612, despite the remarkable flexibility of the PC612 complex, characterized by an open quaternary structure. In contrast, thermal effects are key to describe the properties of the closed PC645 complex due to the lack of a hydrogen bond controlling the twist of ring D in PCB β82 bilins, as well as the larger impact of fluctuations on the excited states of MBV pigments, characterized by a higher conjugation length compared to PCB and DBV bilins found in PC612. Overall, the dual MD protocol combined with FCE spectral simulations led to excellent spectral properties for PC612 and PC645 of unprecedented accuracy. The resulting excitonic Hamiltonian paves the way for future investigations addressing the implications of open and closed quaternary structures on phycocyanin light harvesting properties, with particular interest in addressing the role of vibronic coherences in their excitation transfer mechanisms.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

BO: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. CC: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Financial support from the State Research Agency/Spanish Ministry of Science and Innovation (AEI/10.13039/501100011033; grants PID 2020-115812GB-I00 and CEX 2021-001202-M) and the Catalan Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; 2021SGR00671) are acknowledged. BO thanks a fellowship funded by the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia and the European Social Fund (Ref. 2021 FI_B 00463).

Acknowledgments

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1268278/full#supplementary-material

References

Arpin, P., Turner, D. B., McClure, S. D., Jumper, C. C., Mirkovic, T., Challa, J. R., et al. (2015). Spectroscopic studies of cryptophyte light harvesting proteins: vibrations and coherent oscillations. *J. Phys. Chem. B* 119, 10025–10034. doi:10.1021/acs.jpcb.5b04704

Blau, S. M., Bennett, D. I. G., Kreisbeck, C., Scholes, G. D., and Aspuru-Guzik, A. (2018). Local protein solvation drives direct down-conversion in phycobiliprotein PC645 via incoherent vibronic transport. *Proc. Natl. Acad. Sci.* 115 (15), E3342–E3350. doi:10.1073/pnas.1800370115

Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, et al. (2020). *AMBER 20*. San Francisco: University of California.

Corbella, M., Cupellini, L., Lipparini, F., Scholes, G. D., and Curutchet, C. (2019). Spectral variability in phycocyanin cryptophyte antenna complexes is controlled by changes in the α-polypeptide chains. *ChemPhotoChem* 3 (9), 945–956. doi:10.1002/cptc.201900045

Corbella, M., Toa, Z. S. D., Scholes, G. D., Luque, F. J., and Curutchet, C. (2018). Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes. *Phys. Chem. Chem. Phys.* 20 (33), 21404–21416. doi:10.1039/C8CP02541J

Croce, R., and van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. *Nat. Chem. Biol.* 10 (7), 492–501. doi:10.1038/nchembio.1555

Cupellini, L., Corbella, M., Mennucci, B., and Curutchet, C. (2019). Electronic energy transfer in biomacromolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 9 (2), e1392. doi:10.1002/wcms.1392

Cupellini, L., Lipparini, F., and Cao, J. (2020). Absorption and circular dichroism spectra of molecular aggregates with the full cumulant expansion. *J. Phys. Chem. B* 124 (39), 8610–8617. doi:10.1021/acs.jpcb.0c05180

Curutchet, C., and Mennucci, B. (2017). Quantum chemical studies of light harvesting. *Chem. Rev.* 117 (2), 294–343. doi:10.1021/acs.chemrev.5b00700

Curutchet, C., Munoz-Losa, A., Monti, S., Kongsted, J., Scholes, G. D., and Mennucci, B. (2009). Electronic energy transfer in condensed phase studied by a polarizable QM/MM model. *J. Chem. Theory Comput.* 5 (7), 1838–1848. doi:10.1021/ct9001366

Curutchet, C., Novoderezhkin, V. I., Kongsted, J., Muñoz-Losa, A., Van Grondelle, R., Scholes, G. D., et al. (2013). Energy flow in the cryptophyte PE545 antenna is directed by bilin pigment conformation. *J. Phys. Chem. B* 117 (16), 4263–4273. doi:10.1021/jp305033d

Dreuw, A., Harbach, P. H. P., Mewes, J. M., and Wormit, M. (2010). Quantum chemical excited state calculations on pigment–protein complexes require thorough geometry Re-optimization of experimental crystal structures. *Theor. Chem. Acc.* 125 (3), 419–426. doi:10.1007/s00214-009-0680-3

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2010). *Gaussian development version, revision H.36*. Wallingford CT: Gaussian, Inc.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16, revision B.01*. Wallingford CT: Gaussian, Inc.

Greenwold, M. J., Cunningham, B. R., Lachenmyer, E. M., Pullman, J. M., Richardson, T. L., and Dudycha, J. L. (1902). Diversification of light capture ability was accompanied by the evolution of phycobiliproteins in cryptophyte algae. *Proc. R. Soc. B Biol. Sci.* 2019, 20190655. doi:10.1098/rspb.2019.0655

Harrop, S. J., Wilk, K. E., Dinshaw, R., Collini, E., Mirkovic, T., Teng, C. Y., et al. (2014). Single-residue insertion switches the quaternary structure and exciton states of cryptophyte light-harvesting proteins. *Proc. Natl. Acad. Sci. U. S. A.* 111 (26), E2666–E2675. doi:10.1073/pnas.1402538111

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79 (2), 926–935. doi:10.1063/1.445869

Jurinovich, S., Cupellini, L., Guido, C. A., and Mennucci, B. (2018). EXAT: EXcitonic analysis tool. *J. Comput. Chem.* 39 (5), 279–286. doi:10.1002/jcc.25118

Lee, M. K., Bravaya, K. B., and Coker, D. F. (2017). First-principles models for biological light-harvesting: phycobiliprotein complexes from cryptophyte algae. *J. Am. Chem. Soc.* 139 (23), 7803–7814. doi:10.1021/jacs.7b01780

Loco, D., Jurinovich, S., Cupellini, L., Mennucci, B., and Menger, M. F. S. J. (2018). The modeling of the absorption lineshape for embedded molecules through a polarizable QM/MM approach. *Photochem. Photobiological Sci.* 17, 552–560. doi:10.1039/c8pp00033f

Ma, J., and Cao, J. (2015). Förster resonance energy transfer, absorption and emission spectra in multichromophoric systems. I. Full cumulant expansions and system-bath entanglement. *J. Chem. Phys.* 142 (9), 094106. doi:10.1063/1.4908599

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). Ff14SB: improving the accuracy of protein side chain and backbone parameters from Ff99SB. *J. Chem. Theory Comput.* 11 (8), 3696–3713. doi:10.1021/acs.jctc.5b00255

Maity, S., Bold, B. M., Prajapati, J. D., Sokolov, M., Kubař, T., Elstner, M., et al. (2020). DFTB/MM molecular dynamics simulations of the FMO light-harvesting complex. *J. Phys. Chem. Lett.* 11 (20), 8660–8667. doi:10.1021/acs.jpclett.0c02526

Maity, S., Daskalakis, V., Elstner, M., and Kleinekathöfer, U. (2021a). Multiscale QM/MM molecular dynamics simulations of the trimeric major light-harvesting complex II. *Phys. Chem. Chem. Phys.* 23 (12), 7407–7417. doi:10.1039/d1cp01011e

Maity, S., Sarngadharan, P., Daskalakis, V., and Kleinekathöfer, U. (2021b). Time-dependent atomistic simulations of the CP29 light-harvesting complex. *J. Chem. Phys.* 155 (5), 055103. doi:10.1063/5.0053259

Michie, K. A., Harrop, S. J., Rathbone, H. W., Wilk, K. E., Teng, C. Y., Hoef-Emden, K., et al. (2023). Molecular structures reveal the origin of spectral variation in cryptophyte light harvesting antenna proteins. *Protein Sci.* 32 (3), e4586. doi:10.1002/pro.4586

Mirkovic, T., Doust, A. B., Kim, J., Wilk, K. E., Curutchet, C., Mennucci, B., et al. (2007). Ultrafast light harvesting dynamics in the cryptophyte phycocyanin 645. *Photochem. Photobiological Sci.* 6 (9), 964–975. doi:10.1039/b704962e

Mirkovic, T., Ostroumov, E. E., Anna, J. M., van Grondelle, R., Govindjee, and Scholes, G. D. (2017). Light absorption and energy transfer in the antenna complexes of photosynthetic organisms. *Chem. Rev.* 117 (2), 249–293. doi:10.1021/acs.chemrev.6b00002

Rathbone, H. W., Michie, K. A., Landsberg, M. J., Green, B. R., and Curmi, P. M. G. (2021). Scaffolding proteins guide the evolution of algal light harvesting antennas. *Nat. Commun.* 12 (1), 1890. doi:10.1038/s41467-021-22128-w

Rosnik, A. M., and Curutchet, C. (2015). Theoretical characterization of the spectral density of the water-soluble chlorophyll-binding protein from combined quantum mechanics/molecular mechanics molecular dynamics simulations. *J. Chem. Theory Comput.* 11 (12), 5826–5837. doi:10.1021/acs.jctc.5b00891

Sarngadharan, P., Maity, S., and Kleinekathöfer, U. (2022). Spectral densities and absorption spectra of the core antenna complex CP43 from photosystem II. *J. Chem. Phys.* 156 (21), 215101. doi:10.1063/5.0091005

Scholes, G. D., Mirkovic, T., Turner, D. B., Fassioli, F., and Buchleitner, A. (2012). Solar light harvesting by energy transfer: from ecology to coherence. *Energy Environ. Sci.* 5 (11), 9374. doi:10.1039/c2ee23013e

Senn, H. M., and Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angew. Chemie-International Ed.* 48 (7), 1198–1229. doi:10.1002/anie.200802019

Vreven, T., Byun, K. S., Komáromi, I., Dapprich, S., Montgomery, J. A., Morokuma, K., et al. (2006). Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* 2 (3), 815–826. doi:10.1021/ct050289g

Wang, J., Cieplak, P., Li, J., Hou, T., Luo, R., and Duan, Y. (2011a). Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* 115 (12), 3091–3099. doi:10.1021/jp112133g

Wang, J., Cieplak, P., Li, J., Wang, J., Cai, Q., Hsieh, M., et al. (2011b). Development of polarizable models for molecular mechanical calculations II: induced dipole models significantly improve accuracy of intermolecular interaction energies. *J. Phys. Chem. B* 115 (12), 3100–3111. doi:10.1021/jp1121382

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25 (9), 1157–1174. doi:10.1002/jcc.20035

Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange–correlation functional using the coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 393 (1–3), 51–57. doi:10.1016/j.cplett.2004.06.011

Appendix 2: Computational screening of putative I2 receptors

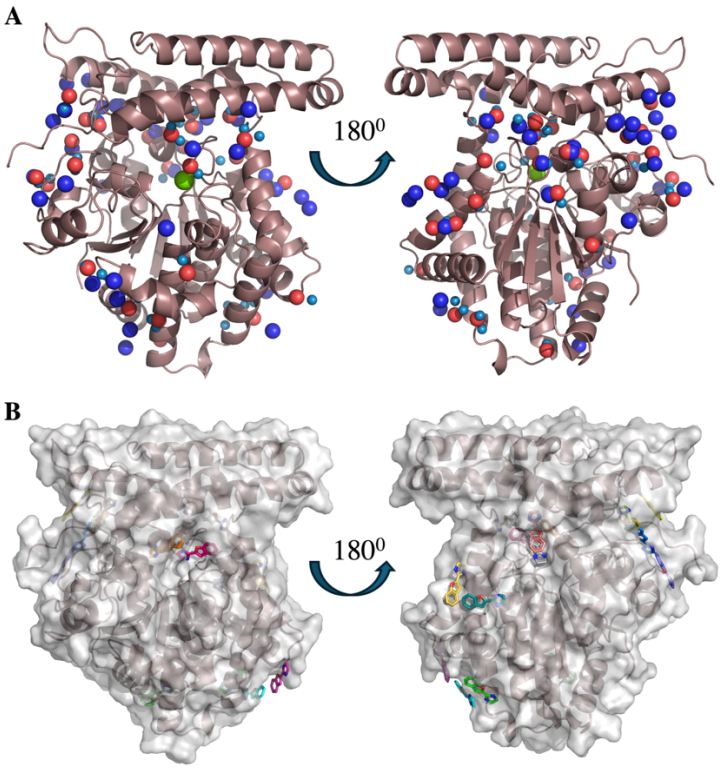


Figure A2.1 (A) Analysis of MDmix simulations for Acidic Residue Methyltransferase 1 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

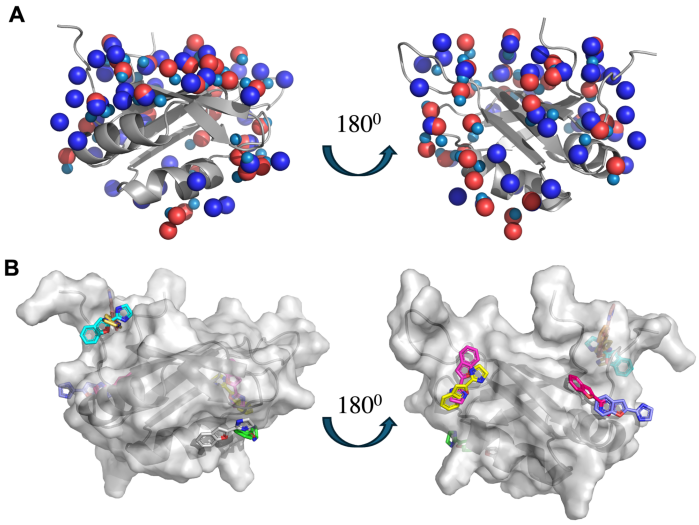


Figure A2.2 (A) Analysis of MDmix simulations for Ribonucleoprotein G-like protein highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

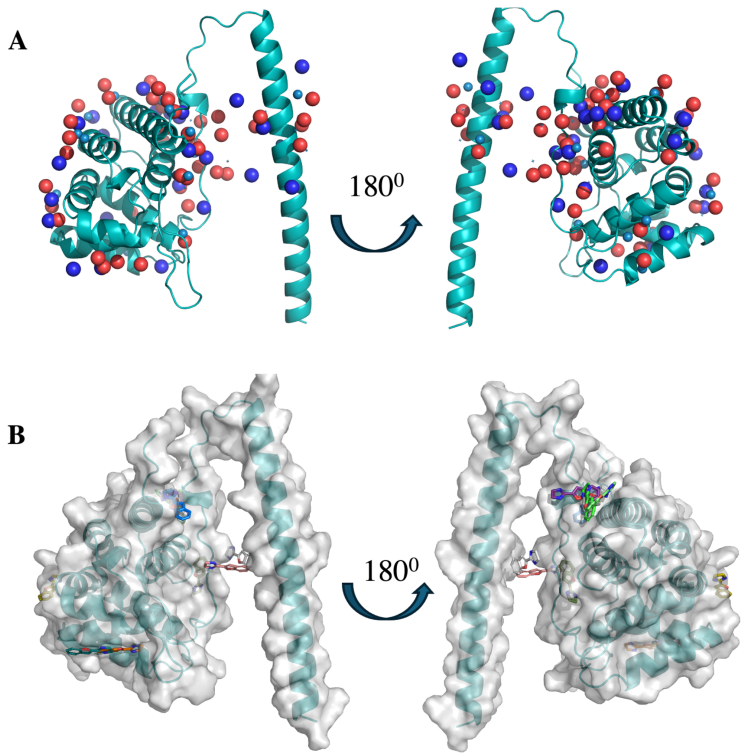


Figure A2.3 (A) Analysis of MDmix simulations for RUN and FYVE domain-containing protein 2 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

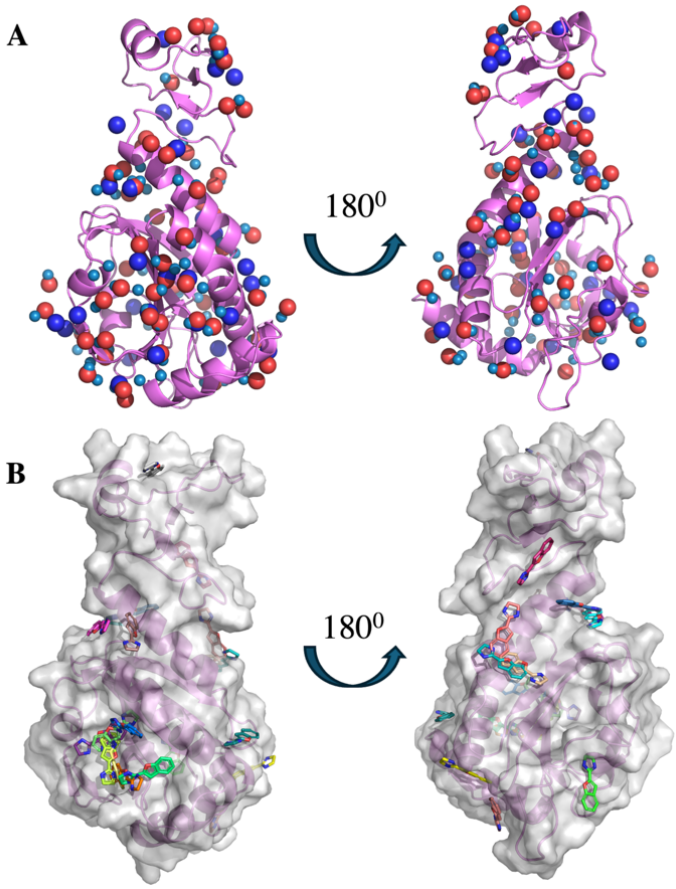


Figure A2.4 (A) Analysis of MDmix simulations for Peroxiredoxin-6 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

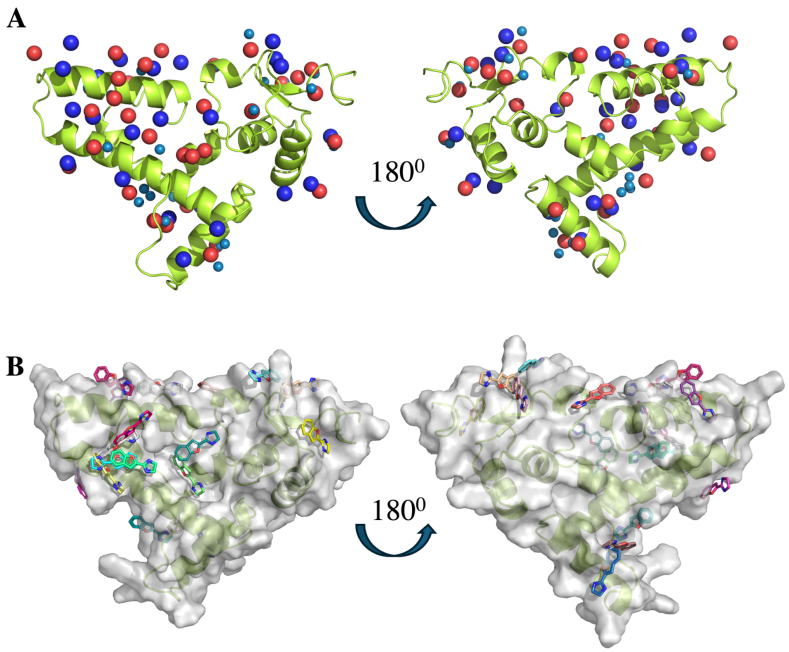


Figure A2.5 (A) Analysis of MDmix simulations for 0S ribosomal protein S9 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

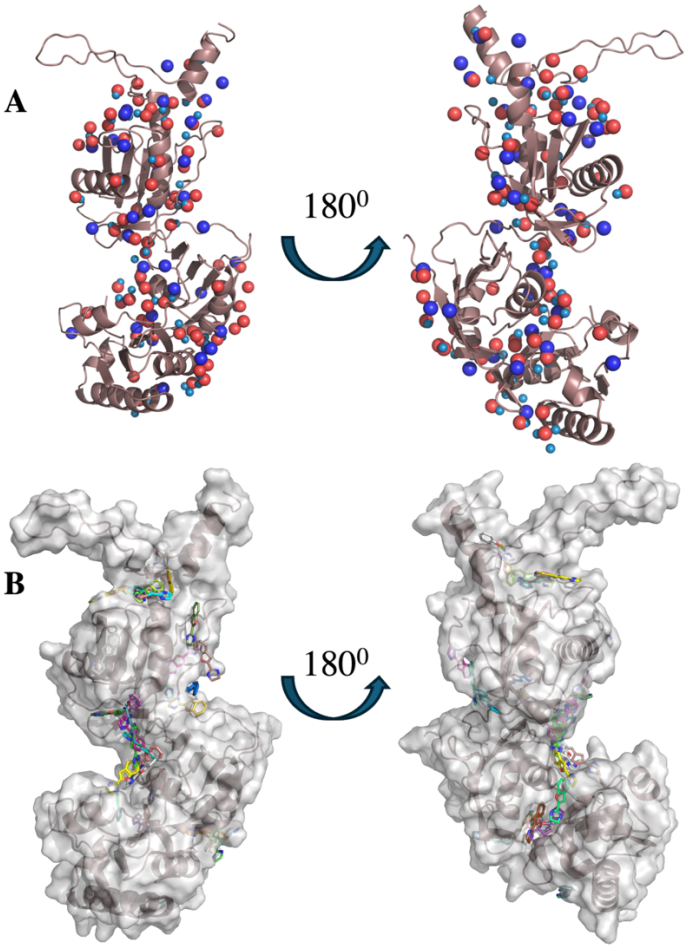


Figure A2.6 (A) Analysis of MDmix simulations for Splicing factor U2AF 65 kDa subunit highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

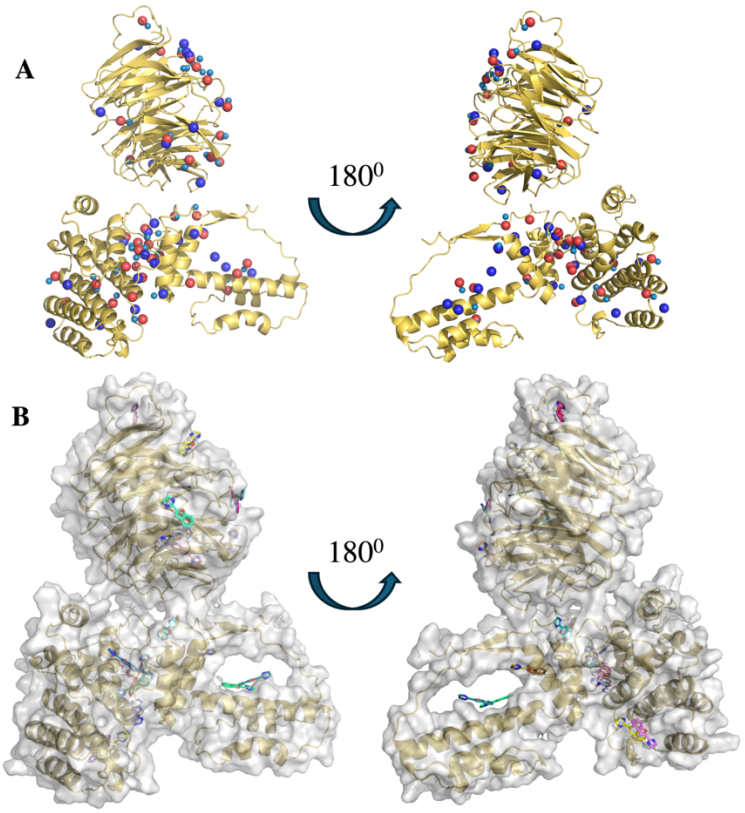


Figure A2.7 (A) Analysis of MDmix simulations for WD repeat-containing protein 47 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

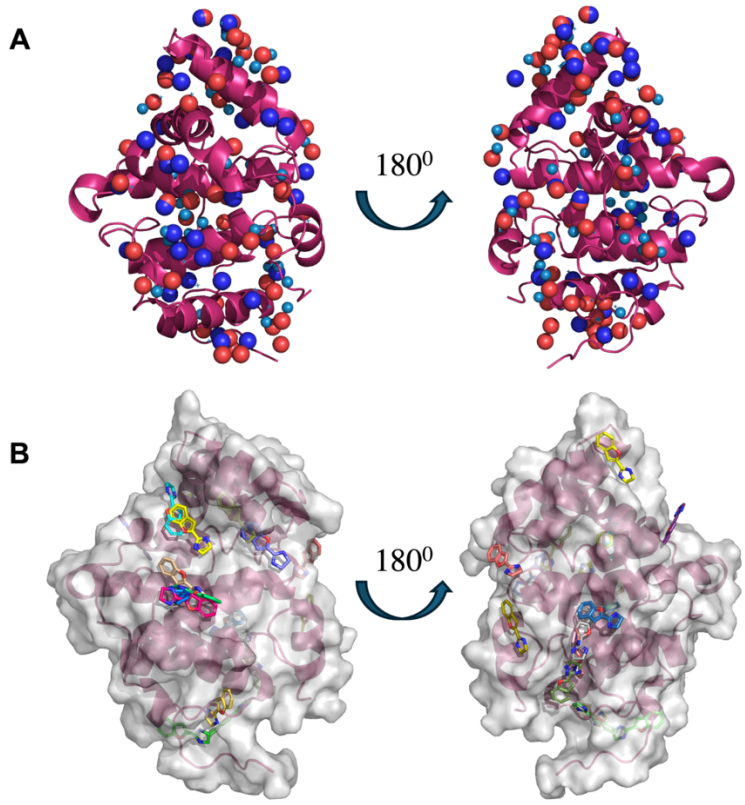


Figure A2.8 (A) Analysis of MDmix simulations for Rho GTPase-activating protein 44 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

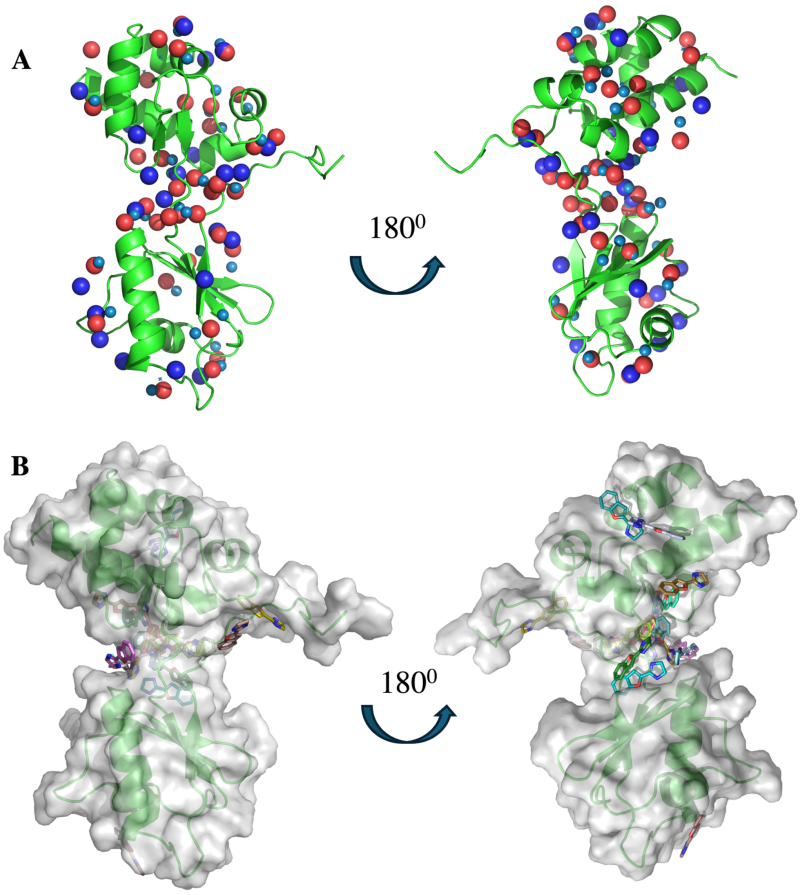


Figure A2.9 (A) Analysis of MDmix simulations for La-related protein 4B highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).

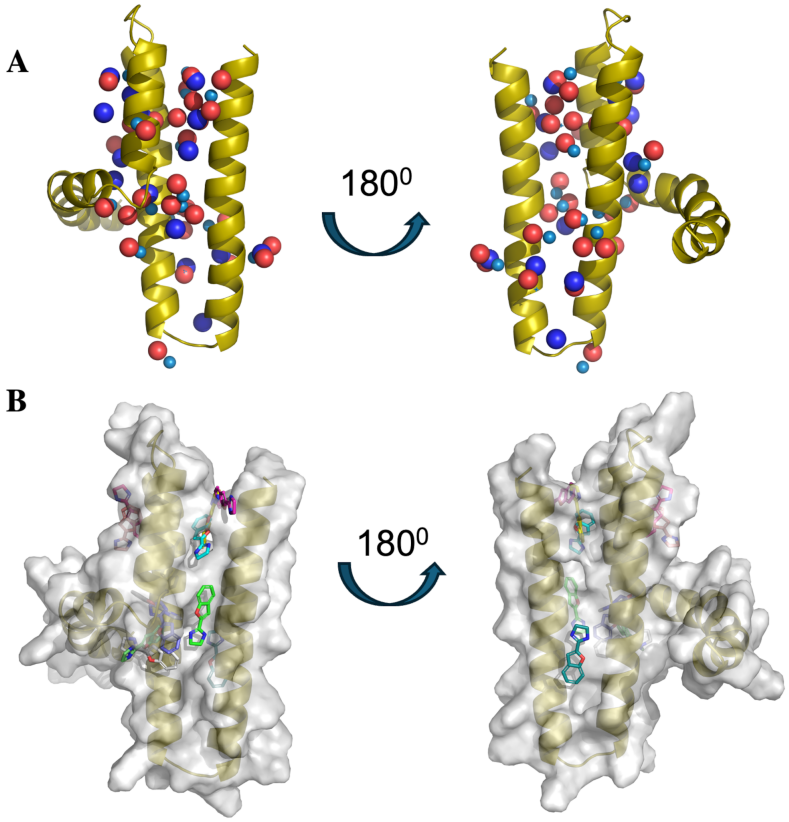


Figure A2.10 (A) Analysis of MDmix simulations for receptor expression-enhancing protein 1 highlighting probe hotspots for pyridine (navy blue), ethanol/isopropanol (OH-red), and water (light blue). (B) Docking poses of 2-BFI obtained using these hotspots as pharmacophoric restraints (only docking scores below zero are shown).