UAM
Universidad Autónoma
de Madrid

FACULTAD DE
CIENCIAS

Trabajo Fin de Máster. Curso 2022 – 2023

MÁSTER ERASMUS MUNDUS EN

QUÍMICA TEÓRICA Y MODELIZACIÓN COMPUTACIONAL

Theoretical Design of Anesthetic Drugs Target to Ion Channels

Autora:

Lucía López Pacios

Directora: Lara Martínez Fernández
Co-Director: Juan José Nogueira Pérez
Lugar de realización: Facultad de Ciencias. Departamento de Química

Campus Internacional
excelencia UAM CSIC+

Acknowledgements

First of all, I would like to thank my supervisors, Juanjo and Lara, for their valuable feedback and constant support. I am also very grateful to the people in their research groups, specially to Vito, Marta and Henar, for their help and guidance during the project.

También quiero agradecer a mi familia, en especial a mis padres por haberme apoyado siempre en mis estudios, a mi hermana, que tiene más criterio estético que yo y a Julia, que además de aguantarme, me ha ayudado también a nivel científico. Mención especial a mi abuela, que se paró a mirar mis gráficos sin saber nada de química. También a Paquita, mi pájaro, por su compañía.

De plus, je veux remercier Jean-Philip Piquemal qui m'a accepté parmi son équipe à Paris, le Laboratoire de Chimie Théorique, et qui m'a accordé une gratification. Aussi à Louis, Olivier et Théo pour leur aide avec Tinker.

This work has been performed with the support of the TCCM consortium scholarship funded by the Universidad Autónoma de Madrid. Moreover, I also acknowledge the computational resources from the Centro de Computación Científica at the Universidad Autónoma de Madrid (CCC-UAM).

Abstract

Photoswitches are becoming increasingly popular in pharmacology due to the possibility of modifying their activity with light. Hence, it is crucial to understand the photophysics of these compounds and their interactions with the biological environment to identify promising light-activated drugs. Azobenzene derivatives are the most widely used class of photoswitches to date. Upon light irradiation, they photoisomerize from the *trans* to the *cis* conformation. In the context of ion channels, one of the conformations may block ion conduction, while the other allows the flow of ions. In this way, light can control ion current through the channel and, consequently, the action potential in excitable cells. In this work, two binding pockets of DADH_2^{2+}, a diprotonated azobenzene derivative, in the cardiac sodium channel ($\mathrm{Na}_V 1.5$), were obtained by means of molecular dynamics simulations and enhanced sampling techniques. The interactions of DADH_2^{2+} with the most relevant residues in the $\mathrm{Na}_V 1.5$ channel were of electrostatic nature. Moreover, the absorption spectra of DADH_2^{2+}, DAD (its neutral species) and azobenzene were calculated in the gas phase, solvent and in the $\mathrm{Na}_V 1.5$ binding pockets, using quantum (TDDFT and CASSCF/CASPT2) and hybrid QM(TDDFT)/MM and QM(TDDFT)/PCM approaches. In all cases, the bright excitation corresponded to the electronic transition to the S_2 state, of $\pi\pi^*$ character. Interestingly, the excitation energies were highly dependent on the environment and the protonation state. Then, the *trans*- *cis* photoisomerization mechanisms were computed in the gas phase and in solvent, using TDDFT and QM(TDDFT)/PCM. After excitation to $\pi\pi^*$, the system relaxes and crosses to the S_1 state, of $n\pi^*$ character. In this surface the C-N=N-C dihedral rotates until reaching the ground state, where it may convert to the *cis* isomer or go back to the *trans* conformation. As for the excitation energies, it was found that the photoisomerization was also highly dependent on the environment.

Resumen

Los *photoswitches* son cada vez más populares en farmacología debido a la posibilidad de modificar su actividad con luz. Por tanto, es crucial ententer su fotofísica y sus interacciones en medios biológicos para identificar compuestos que puedan potencialmente usarse como fármacos activados por luz. Actualmente, los derivados de azobenceno son los *photoswitches* más conocidos. Tras la irradiación con luz, los derivados de azobenceno fotoisomerizan de la conformación *trans* a la *cis*. En el contexto de los canales iónicos, una de las conformaciones puede bloquear la conducción iónica, mientras que la otra permite el paso de iones. De esta forma, la luz puede controlar el flujo de iones a través del canal y, en consecuencia, el potencial de acción en células excitables. En este trabajo, se han encontrado dos sitios de unión del DADH_2^{2+}, (derivado de azobenceno diprotonado), en el canal de sodio cardiaco ($\mathrm{Na}_V1.5$), mediante cálculos de dinámica molecular clásica y acelerada. Además, se ha obtenido el espectro de absorción del DADH_2^{2+}, DAD (su especie neutra) y azobenceno en fase gas, disolvente y en los sitios de unión del $\mathrm{Na}_V1.5$, utilizando métodos cuánticos (TDDFT y CASSCF/CASPT2) e híbridos (QM(TDDFT)/MM y QM(TDDFT)/PCM). En todos los casos, la excitación brillante se corresponde con la transición electrónica al estado excitado S_2, de carácter $\pi\pi^*$. Cabe destacar que las energías de excitación en estos sistemas dependen enormemente del medio y del estado de protonación. A continuación se calcularon los mecanismos de fotoisomerización *trans*- *cis* en fase gas y disolvente, utilizando TDDFT y QM(TDDFT)/PCM. Después de la excitación al estado $\pi\pi^*$, el sistema se relaja y cruza al estado S_1, de carácter $n\pi^*$. En esta superficie, se produce la rotación del dihedro C-N=N-C hasta alcanzar el estado fundamental S_0, donde el *photoswitch* puede convertirse en el isómero *cis* o volver a la conformación *trans*. De la misma manera que las energías de excitación, los mecanismos de isomerización también dependen del medio en el que se encuentran las moléculas.

Table of contents

1 Introduction **1**

- 1.1 Photopharmacology . . . 2
- 1.2 Azobenzenes as photoswitches . . . 3
- 1.3 Ion Channels . . . 6
- 1.4 Azobenzenes in Ion Channels . . . 9
- 1.5 Objectives . . . 10

2 Theoretical Methods **11**

- 2.1 Electronic structure methods . . . 11
  - 2.1.1 The Schrödinger equation . . . 11
  - 2.1.2 Hartree Fock . . . 13
  - 2.1.3 Multi-Configurational Self-Consistent-Field . . . 15
  - 2.1.4 Density Functional Theory . . . 20
  - 2.1.5 Basis sets . . . 26
  - 2.1.6 Wave Function Analysis . . . 27
- 2.2 Molecular Mechanics . . . 29
  - 2.2.1 Molecular Dynamics . . . 29
  - 2.2.2 Gaussian Accelerated Molecular Dynamics . . . 35
  - 2.2.3 Binding Free Energy Calculations . . . 37
- 2.3 Hybrid Quantum/Classical Methods . . . 38
  - 2.3.1 Quantum Mechanics/Molecular Mechanics . . . 38
  - 2.3.2 Polarizable Continuum Model . . . 40

3 Computational Details **42**

- 3.1 Classical Simulation Details . . . 42
  - 3.1.1 Initial structures . . . 42
  - 3.1.2 Docking . . . 43
  - 3.1.3 Molecular Dynamics Details . . . 43

3.1.4 Gaussian Accelerated Molecular Dynamics details 44
3.1.5 Molecular Mechanics Generalized Born Surface Area details 44
3.2 Quantum Mechanics calculation details 45
3.2.1 Dynamic Absorption spectra 45
3.2.2 Static Absorption spectra 45
3.2.3 Photoisomerization PES 46

4 Results 47
4.1 Binding pocket search 47
4.1.1 Na$_V$1.5 channel equilibration 47
4.1.2 Docking: preliminary binding pockets of DAD 49
4.1.3 GaMD: binding pockets of DAD 50
4.2 Absorption spectrum of DAD 54
4.3 DAD Photoisomerization 60
4.4 Outlook 62

5 Conclusions 65

References 67

Appendix A Benchmarking of excitations from a common M06-2X optimized geometry 76

Introduction

Ion transport across cell membranes underlies many essential biological processes, such as nerve impulse transmission, muscle contraction, blood pressure regulation, and fluid secretion by some organs[1]. The flow of ions between the extracellular and the intracellular environments is controlled by the pore opening/closing of the voltage-gated ion channels in response to the variation of the transmembrane electrostatic potential. The resulting ionic current in the channel induces further ionic transport in neighbouring channels and, hence, it generates an action potential that is propagated along the membrane and contiguous excitable cells.[2]

Many anaesthetics target ion channels in pain sensing neurons with the aim of blocking action potentials and hampering pain signal propagation to the brain.[3] Similarly, the treatment of certain arrhythmic disorders addresses ion channels in cardiomyocytes (cardiac muscle cells), allowing a more precise control of heart rhythm.[4] However, ion blockers usually present low selectivity and may block ion channels in undesired excitable cells (neurons and muscle cells), causing severe side effects.[5] The use of photoswitchable drugs prevents this from happening as they are activated by light when and where it is most convenient. This permits a spatio-temporal control that is not possible with conventional drugs.[6]

In this section the concepts of photopharmacology and azobenzenes, a very well known class of photoswitches, will be introduced. Then, ion channels will be presented, and more specifically, the structure of voltage-gated sodium channels and their interactions with both conventional and photoswitchable ion blockers. Finally, the main goals of the present project will be stated.

1.1 Photopharmacology

The ultimate goal of pharmacology is providing safe and effective drugs to improve health conditions. However, the environment in which drugs act after their administration is complex and may involve multiple interactions with different proteins, lipids, DNA strands and other biomolecules (broad selectivity), leading to undesirable side effects. Therefore, rational design aiming at narrowing drug selectivity is a major line of research in pharmacology nowadays.[7] Moreover, this widespread distribution of the drug activity, along with the poor control of the duration of action, often results in increased tolerance to drugs, or even drug resistance, due to excessive exposure.[8] Possibly one of the most concerning drug resistance phenomena is antibiotic resistance, which is one of the biggest threats to global health today and, although it is a natural process that we cannot avoid, it is possible to slow it down by limiting the time exposure and location to what is strictly necessary.[9]

In the last few years, photopharmacology has emerged as a way to tackle these issues.[10] It consists in the use of light-activated compounds to treat diseases, which allows a more precise spatio-temporal control given that their activation depends on the wavelength and the intensity of the irradiated light.[11] Light-activated drugs can be divided in phototriggers and photoswitches. The former are bioactive compounds that are bonded to photosensitive protecting groups. After irradiation, the photosensitive group is removed, leaving the main molecule active until degradation (irreversible change).[12] On the other hand, light induces reversible changes in photoswitches as they normally involve isomerization processes (photoisomerizations).[11] The latter, which are the subject of this work, are more convenient since phototriggers present the same disadvantages as conventional drugs after their activation (Figure 1.1).

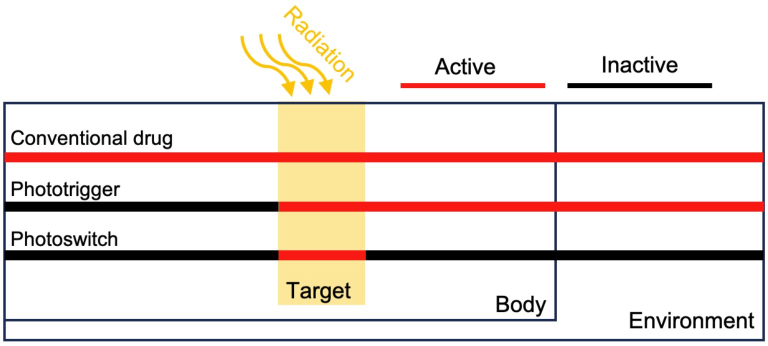

Fig. 1.1 Schematic representation of the activity of conventional drugs, phototriggers and photoswitches. Conventional drugs are always active, phototriggers become active after irradiation and photoswitches are only active while and where they are irradiated.

To ensure the effectiveness and safety of light-activated compounds, the drug conformation after irradiation should be biologically active only in the desired target, whereas the other conformation, or dark conformation, needs to be inactive and non-toxic everywhere. Moreover, the photoactivation wavelength should ideally be located in the visible region of the electromagnetic spectrum, avoiding ultraviolet (UV) light, and long enough to penetrate tissues. In addition, high oscillator strengths in the bright electronic excitations, as well as a good quantum yield of photoisomerization, are convenient to avoid the need of high intensity light for the photoactivation. Finally, back-photoisomerization from the active to the inactive conformer should take place at a wavelength different from the photoactivation or thermally. In this sense, having higher energy active conformations conveniently allows thermal relaxation to the lower energy inactive isomers without the need of light.[11]

1.2 Azobenzenes as photoswitches

Azobenzene derivatives (ABs) are the most widely used photoswitches to date as they fulfill most of the requirements stated in the previous section. Azobenzene (AB) exists in both *cis* and *trans* conformations (Figure 1.2). The *trans* conformation is more stable than the *cis* and thus the former is more abundant in the dark.[13]

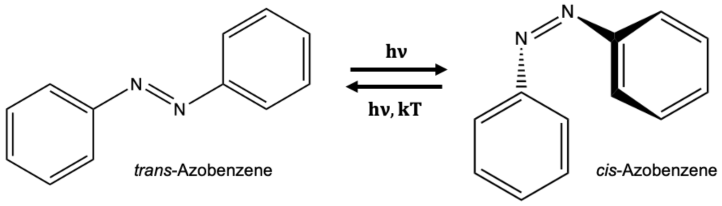

Fig. 1.2 AB isomerization reaction. The *trans*→*cis* isomerization occurs upon ligth irradiation ($h\nu$) whereas the *cis*→*trans* can occur both thermally (kT) and photochemically ($h\nu$).

The *trans*-AB absorption spectrum consists of two bands in the UV-vis region: a very strong band at λ_{max} = 330 nm corresponding to the bright $\pi\pi^*$ S_2 excitation, and a much weaker and wider one at λ around 430 nm that is attributed to the dark $n \rightarrow \pi^*$ S_1 transition. [14] On the other hand, the *cis*-AB absorption spectrum presents only one band in the visible region, around 450 nm, which corresponds to the S_1 $n \rightarrow \pi^*$ transition and whose intensity is low. Then, the S_2 $\pi\pi^*$ transition is located at 260 nm.[13]

Most studies suggest that the conformational change occurs in the dark S_1 ($n\pi^*$) state, which can be populated after direct excitation or after internal conversion from the bright S_2 ($\pi\pi^*$) state (Figure 1.3).[15] Due to its lower oscillator strength, direct excitation to the

S_1 ($n\pi^*$) state is much less probable than S_2 ($\pi\pi^*$) excitation. Moreover, different quantum yields are observed depending on the excited state that is populated first: 0.20-0.27 for S_1 ($n\pi^*$) excitation and 0.09-0.12 for S_2 ($\pi\pi^*$). This suggests that AB follows different isomerization mechanisms depending on the excitation wavelength and that more pathways may open when the photoswitch is excited to the S_2 state.[16, 17] The isomerization mechanisms are still subject of debate and the most popular ones are rotation, inversion, concerted inversion and inversion-assisted rotation (Figure 1.4).

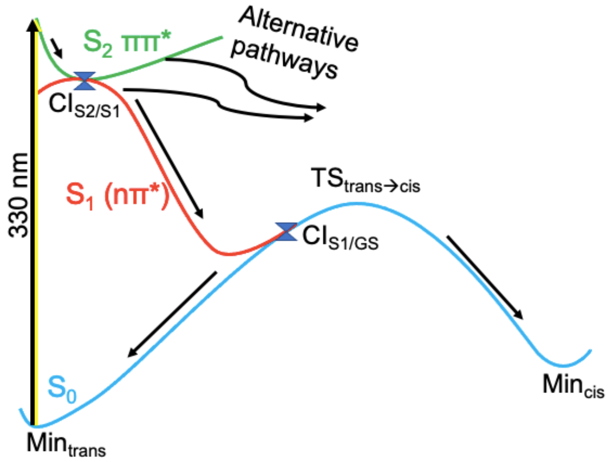

Fig. 1.3 Schematic representation of the azobenzene *trans* → *cis* photoisomerization after $\pi\pi^*$ excitation.

For a long time, it was thought that excitation to the S_1 ($n\pi^*$) state followed an in-plane inversion at one of the two nitrogens of the azobenzene (inversion), whereas excitation to the S_2 ($\pi\pi^*$) state followed rotation or twisting around the N=N bond. According to this hypothesis, the most probable excitation (S_2) lead to a rotation in the $\pi\pi^*$ state that could evolve towards a radiationless decay to the *trans*-AB ground state or to internal conversion to the S_1 ($n\pi^*$) state, where isomerization is only possible through inversion.[18] This bifurcation explained, for the first time, the difference in the quantum yields. Later on, mixed mechanisms, such as concerted inversion, inversion-assisted rotation and rotation helped by the opening of the CNN angles, were proposed. In fact, dynamic studies usually support these isomerization pathways.[19]

The most recent studies focus on the potential energy surface (PES) along the $n\pi^*$ state and, contrary to previous beliefs, they agree that rotation is the main reaction coordinate in the S_1/S_0 internal conversion, which is then followed by isomerisation in the ground state (Figure 1.3).[20] Then, *cis*-AB thermally converts back to the *trans*-AB through an inversion mechanism. This process shows a large energy barrier which results in a long *cis*-AB lifetime.

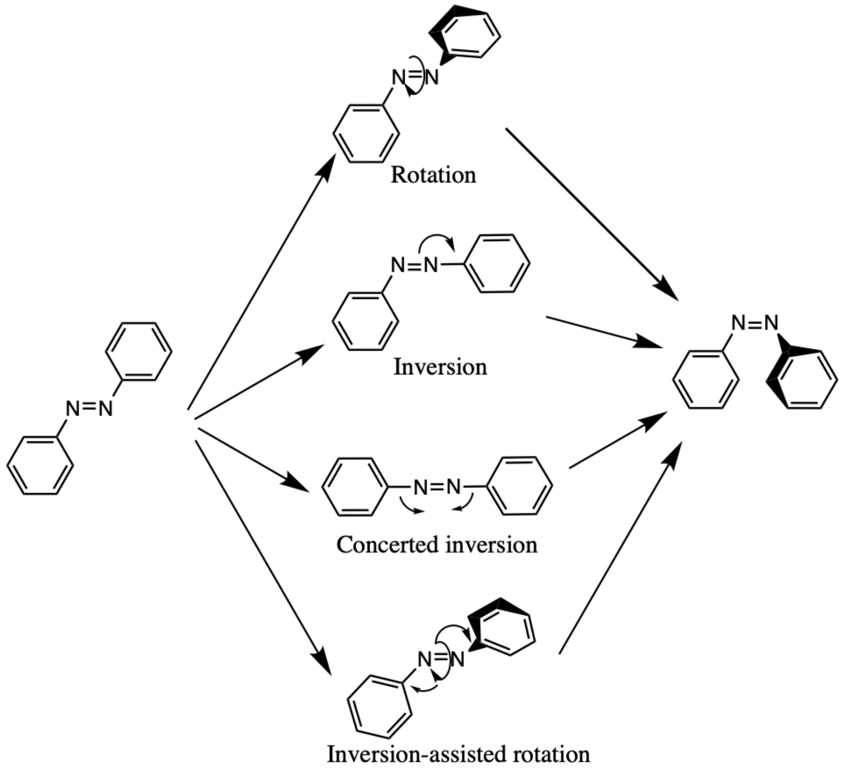

Fig. 1.4 Azobenzene *trans-cis* isomerization mechanisms.

On the other hand, the main coordinate of relaxation in S_2 leads to a minimum where the CCNN and NNCC dihedral angles are in a pedal-like conformation. In this geometry, both S_1 and S_2 are degenerate and, furthermore, its structure is very close to the Franck-Condon (FC) geometry (ground state of the *trans* isomer), explaining the high $S_2 \rightarrow S_1$ quantum yield observed in experiments. After the internal conversion, the most favorable pathway involves the same rotated S_1/S_0 conical intersection (CI) from the $n\pi^*$ PES. [15]

The S_2 PES also shows alternative deactivation pathways that may explain its lower quantum yield.[15] Moreover, the energy excess after the S_2/S_1 internal conversion also opens additional deactivation mechanisms in the $n\pi^*$ PES, decreasing the quantum yield after the bright S_2 excitation.[13]

One way of improving the yield in the photoisomerization is the functionalization of azobenzene. For example, adding a electron donor substituent, such as an amino group in ortho or para positions, would improve the quantum yield due to the S_2 red-shifting. As a result, the excitation energy decreases and, most likely, relaxation will not follow alternative pathways.

Moreover, the required radiation is less harmful, penetrates better in the tissues and, in addition, thermal isomerization is also faster because the additional donor groups decrease the energetic barrier.[13]

This work focuses on the study of the azobenzene derivative diethylamino-azo-diethylamino (DAD), depicted in Figure 1.5. DAD is a photoswitch that exists in both deprotonated or protonated forms. The deprotonated state can diffuse rapidly across biological barriers due to its non-polar nature, making it convenient to target transmembrane proteins like ion channels (more on this in the following sections). Regarding the photoisomerisation, the most abundant isomer in the dark, the *trans*-DAD, may turn into its *cis* isomer upon irradiation with 454 nm light. Then, the *cis*-DAD relaxes back to the *trans*-DAD in the dark within 200 ms.[21]

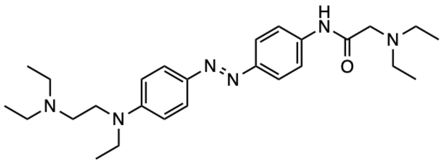

Fig. 1.5 Chemical structure of the deprotonated azobenzene derivative DAD. Protonation occurs in the tertiary amino groups, except in the one that is next to the aromatic ring.

1.3 Ion Channels

Ion channels are transmembrane proteins that selectively control ion transport (primarily K^+, Na^+, Ca^{2+}, H^+ and Cl^-) between the extracellular medium and the cytoplasm (Figure 1.6). They consist of protein domains that collectively form a selective pore filled with water through which specific ions can flow in response to mechanical forces, chemical stimuli, temperature changes or voltage variations. Voltage gated ion channels respond to changes in the membrane potential and are essential for neuronal transmission and muscle contraction, among other functions.[22] When the excitable cells are in their resting state, the pore in the ion channels is closed in order to keep the membrane potential constant. Their opening, in response to voltage variation, may lead to the membrane depolarization, which further induces the opening of neighbouring channels and results in the action potential propagation.[2]

Figure 1.7 shows the voltage profile of the membrane during the neuronal action potential. The membrane potential, i.e., the electrical potential difference across the membrane measured relative to the extracellular environment, is kept constant at around -70 mV in

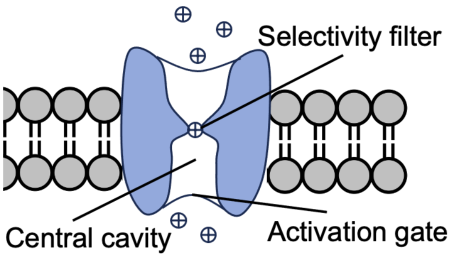

Fig. 1.6 Schematic representation of an ion channel in a lipid membrane.

the resting state of the excitable cells. When sufficient current influx raise the membrane potential up to a threshold of -55 mV, the voltage gated sodium channels (Na_V) massively open to allow the flow of Na^+ ions into the cell (depolarization). On reaching a potential of 40 mV, the Na_V channels inactivate and the voltage gated potassium channels (K_V) open, causing the efflux of K^+ ions out of the cell (repolarization). Due to their slow kinetics, the K_V channels remain opened longer than needed to reach the resting potential. This leads to the hyperpolarization of the membrane, that is later reversed thanks to the ATP dependent sodium-potassium pump.[23] Note that the Na_V channels can be in three different states: open, inactive and closed. Both inactive and closed states block the entrance of ions. However, while closed channels have the ability to open and allow the flux of Na^+, the inactive ones do not. At the resting membrane potential the channels are closed. Then, they open during depolarization and they inactivate when repolarization starts. The channels do not close until the end of the action potential. In this way, the cell membrane cannot depolarize again and it can go back to its resting state. This is known as the refractory period.[23]

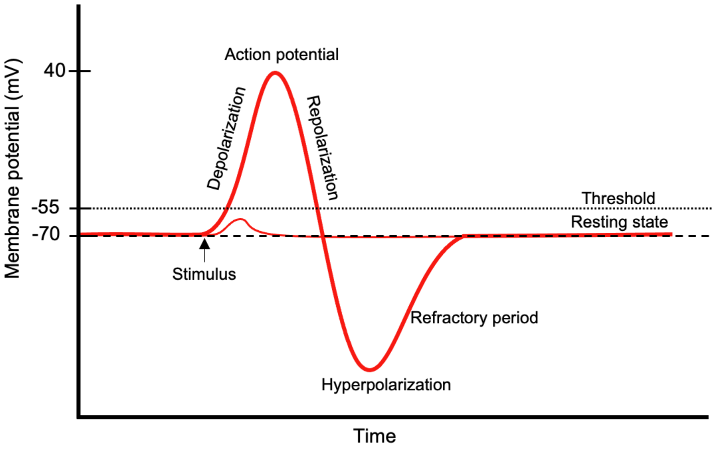

Fig. 1.7 Voltage profile of the action potential in neurons. The smaller bump represents a failed action potential due to insufficient current to reach the threshold. Action potentials in other excitable cells have different voltage profiles although they all start by depolarization.

Blocking Na$_V$ channels would hinder the action potential in cell membranes and, thus, stop its propagation through nerves or muscle fibres. For this reason, this work focuses on the study of Na$_V$ channels as controlling their pore opening allows a precise control of signal propagation. Eukaryotic Na$_V$ channels share a common structure composed of a large pore-forming α subunit and smaller β subunits with auxiliary functions. The pore-forming subunit (Figure 1.8) consists of four homologous domains (DI-DIV) that contain 6 transmembrane segments each (S1-S6). Segments S1-S4 in all domains form the voltage sensors (VSs), whereas S5-S6 form the pore module (PM). The PM also contains a pore loop between S5 and S6 (P loop).[24]

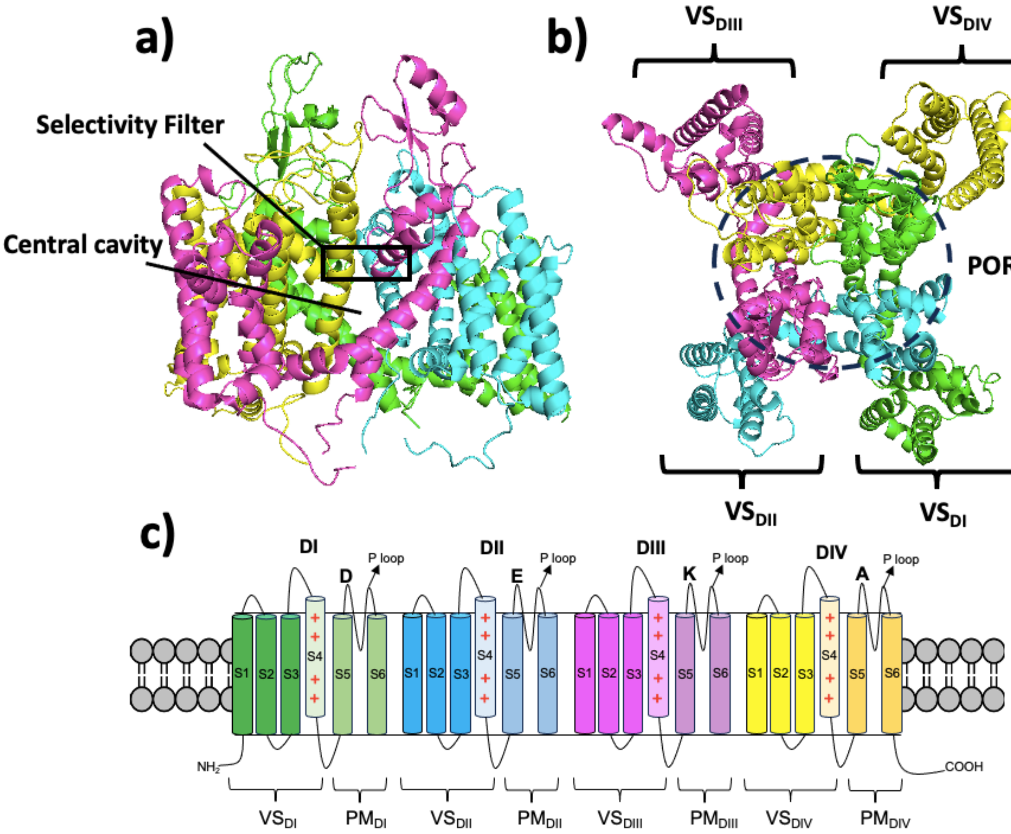
a)
Selectivity Filter
Central cavity
b)
VS$_{DIII}$
VS$_{DIV}$
PORE
VS$_{DII}$
VS$_{DI}$
c)
DI
DII
DIII
DIV
D
E
K
A
P loop
S1
S2
S3
S4
S5
S6
NH$_2$
COOH
VS$_{DI}$
PM$_{DI}$
VS$_{DII}$
PM$_{DII}$
VS$_{DIII}$
PM$_{DIII}$
VS$_{DIV}$
PM$_{DIV}$

Fig. 1.8 a) Side view of the four domains of the Na$_V$1.5 channel. b) Top view of the four domains of the Na$_V$1.5 channel. c) 2D structure representation of any Na$_V$ channel. Domains DI, DII, DIII and DIV are green, blue, pink and yellow respectively.

Variation in the membrane potential induces a conformational change in the S4 segments of the VSs, leading to the pore opening/closing. S4 contains positively charged aminoacids that interact with negative residues in the neighbouring S1-S3. Upon depolarization, this electrostatic force is relieved and, thus, S4 slides towards the extracellular medium and opens the pore of the ion channel.[25]

One of the most important regions in the Na_V channels is the selectivity filter (SF). It is located in the P-loop and right before the central cavity of the pore (Figure 1.8). As shown in Figures 1.9 and 1.8C, it is composed by the aminoacids Asp, Glu, Lys and Ala (DEKA) in eukaryotic cells and it is responsible for the selectivity of the ion channel, although the reasons are still unknown.[25]

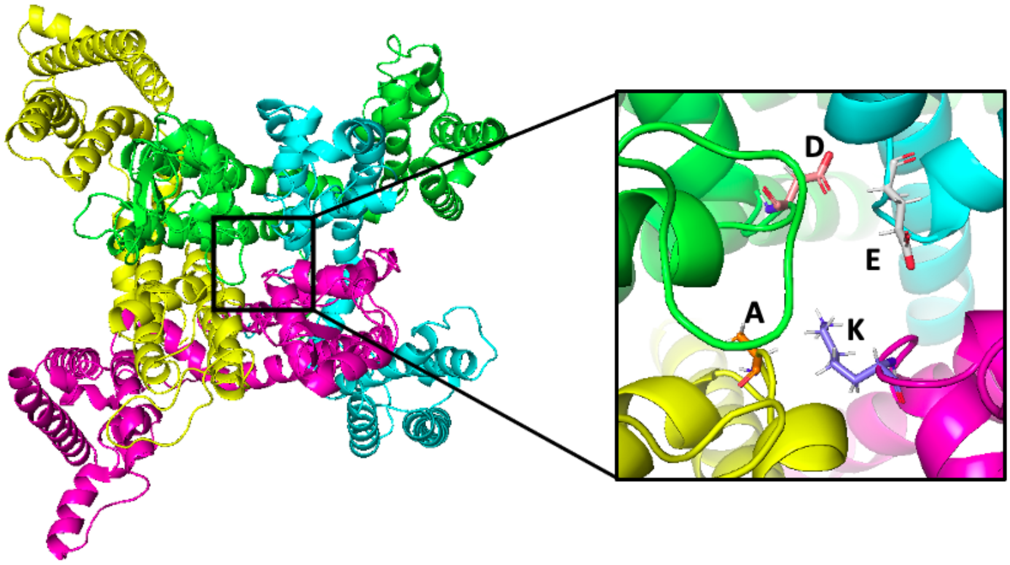

Fig. 1.9 DEKA aminoacids forming the selectivity filter in the $Na_V1.5$ voltage-gated ion channel. Domains DI, DII, DIII and DIV are green, blue, pink and yellow respectively.

This study will focus on the $Na_V1.5$ channel, whose open crystal structure was determined in 2021 by means of cryo-electron microscopy.[24] The $Na_V1.5$ channel triggers the cardiac action potential and initiates the heartbeat. Certain mutations in these channels have been associated to arrhythmic disorders and, in this sense, some antiarrhythmic drugs have been designed to block the $Na_V1.5$ channel.[24]

1.4 Azobenzenes in Ion Channels

The treatment of channelopathies (disorders in ion channels) usually involves drugs that hinder ion conduction through their pore. After forming non-covalent bonds with the ion channel, the drug can block the flow of ions by direct or indirect inhibition. In the former, the drug directly interferes with the protein selectivity filter, whereas in the latter the inactive state of the channel is stabilised.[26]

In the same way as conventional drugs, azobenzenes can follow both direct and indirect blocking strategies. The challenge lays on developing azobenzene derivatives whose photophysics assures a precise spatio-temporal control of ion channels. As mentioned earlier, they also need to be non-toxic and fulfill the requirements of light-activated drugs. Moreover, their structure should be capable of penetrating cell membranes, they must not degrade in the cytoplasm, and they should selectively interact with the desired ion-channels.[27, 28]

DAD is a promising photoswitch to target ion channels as it has already been used to restore retinal light responses and light-driven behavior to blind mice.[21] According to this experimental study, the dark *trans*-DAD (protonated or deprotonated) blocks the K_V channel and then, upon 460 nm irradiation, it photoisomerizes to *cis*-DAD and unblocks the channel. Furthermore, the existence of deprotonated DAD (neutral species), allows the rapid diffusion across the membrane and its entrance to the ion channel.

1.5 Objectives

The present work includes the first steps of a more ambitious project, whose main final goal is simulating the phoisomerization of DAD inside the $Na_V1.5$ channel to understand the ion conduction blocking mechanism of this photoswitch. More specifically, the aims of this project are:

- Finding potential binding pockets of *trans*-DAD in the $Na_V1.5$ channel and understanding the nature of its interactions with the most relevant residues in the binding modes.
- Calculating the absorption spectrum of *trans*-DAD in vacuum, in water and in the most stable pockets of the $Na_V1.5$ channel to characterise the excited states involved in the photophysics of DAD, and to investigate the effect of the environment on the excitations.
- Computing the *trans*-*cis* photoisomerization of DAD in vacuum and in water to deepen into the photochemical aspects of these mechanisms and to assess the effect of the solvent on the deactivation process.

Theoretical Methods

In this section, the theoretical methods that have been used during this work will be introduced. They are divided in electronic structure methods, molecular mechanics and hybrid Quantum/Classical Methods. The first ones are based in the resolution of the Schrödinger equation. They have been used to model the photophysics of DAD in vacuum and in solution, in combination with hybrid methods. Conversely, Molecular Mechanics uses empirical equations (Force Fields) to approximate the energy of the system. They are often combined with the classical equations of motion to investigate the time evolution of the system. This approach has been used to find the binding pockets of DAD in the $\mathrm{Na}_V 1.5$ channel.

2.1 Electronic structure methods

2.1.1 The Schrödinger equation

The time-dependent Schrödinger equation [29] is of central importance in Quantum Chemistry as it describes the temporal evolution of the wave functions of particles. For this reason, most methods in Quantum Chemistry aim at finding approximate solutions to the Schrödinger equation, which is given as:

$$\frac{\partial \Psi(r,t)}{\partial t} = -\frac{i}{\hbar}\hat{H}(r,t)\Psi(r,t) \tag{2.1}$$

Where $\Psi(r,t)$ is the wave function, that contains all the information of the system, t is the time, r are the coordinates of $\Psi(r,t)$ and $\hat{H}$ is the Hamiltonian operator, which is expressed as the sum of the kinetic ($\hat{T}$) and potential ($\hat{V}$) energy operators:

$$\hat{H}(r,t) = \hat{T}(r) + \hat{V}(r,t) = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V}(r,t) \tag{2.2}$$

$\hat{T}(r)$ depends on the second order derivative ∇^2. When $\hat{V}$ does not depend on time, the Hamiltonian becomes time-independent. Then, based on the Born-Huang formulation, i.e., the separation of the fast and slow degrees of freedom, $\Psi(r,t)$ can be expressed as the product of its spatial and temporal parts. As a result, equation 2.1 reduces to:

$$\hat{H}(r)\,|\Psi(r)\rangle = E\,|\Psi(r)\rangle \tag{2.3}$$

Where E is the energy of the $|\Psi\rangle$ eigenfunction. Equation 2.3 is known as the Time-Independent Schrödinger Equation (TISE) and its $|\Psi\rangle$ solutions correspond to the stationary states of the system. It is impossible to find analytical solutions of the TISE for systems involving more than two particles and, hence, several approximations are needed to treat most chemical problems. The $\hat{H}$ operator in molecules consists of five terms:

$$\hat{H} = -\sum_{i=1}^{n}\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N}\frac{1}{2M_A}\nabla_A^2 + \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{r_{ij}} + \sum_{A=1}^{N}\sum_{B>A}^{N}\frac{Z_A Z_B}{R_{AB}} - \sum_{i=1}^{n}\sum_{A=1}^{N}\frac{Z_A}{r_{ij}} \tag{2.4}$$

Where Z denotes the atomic number, and r_{ij}, R_{AB} and R_{iA} are the electron-electron, nucleus-nucleus and electron-nucleus distances, respectively. The first two terms are the kinetic energy operators of the n electrons ($\hat{T}_e$) and the N nuclei ($\hat{T}_N$) respectively. Similarly, the last three terms correspond to the potential between electrons (V_{ee}), the potential between nuclei (V_{NN}), and the potential between electrons and nuclei (V_{eN}). Note that equation 2.4 is expressed in atomic units and, thus, $\hbar$, the charge and the mass of the electron are equal to 1.

Considering that nuclei move several orders of magnitude slower than electrons due to its heavier mass, the Born-Oppenheimer approximation[30] states that the total wave function can be split in its nuclear and electronic contributions:

$$|\Psi(r,R)\rangle = |\psi_e(r,R)\rangle\,|\psi_N(R)\rangle \tag{2.5}$$

Where $|\psi_e(r,R)\rangle$ and $|\psi_N(R)\rangle$ are the electronic and nuclear wave functions respectively. Note that $|\psi_e(r,R)\rangle$ has a parametrical dependance on the nuclear coordinates R. In this way, the electronic Schrödinger equation (Equation 2.6) is solved for fixed nuclear geometries, whereas Equation 2.3 includes all nuclear configurations.

$$\hat{H}_e\,|\Psi_e\rangle = E_e\,|\Psi_e\rangle \tag{2.6}$$

The electronic Hamiltonian $\hat{H}_e$ only consists of $\hat{T}_e$, $\hat{V}_{ee}$ and $\hat{V}_{eN}$. As the nuclear geometry is fixed, $\hat{T}_N = 0$ and $\hat{V}_{NN}$ is constant. Then, the total energy of the system is obtained as the sum of the electronic energy (E_e) and the Coulomb potential between the nuclei (V_{NN}).[31]

However, due to the $\hat{V}_{ee}$ term, the electronic Schrödinger equation (Equation 2.6) is still impossible to solve for systems containing more than one electron. In this sense, the following sections focus on further approximations that theoretical chemists use to deal with this problem.

2.1.2 Hartree Fock

The Hartree Fock approximation (HF) [31] is the basis of most ab initio methods in quantum chemistry. It expresses the many-electron wave function of the system as a Slater determinant (equation 2.7) of one-electron wave functions, which are the spin orbitals $|\chi_i(x_i)\rangle$.

$$|\Psi(x_1, x_2, \cdots, x_n)\rangle = \begin{vmatrix} |\chi_1(x_1)\rangle & |\chi_2(x_1)\rangle & \cdots & |\chi_n(x_1)\rangle \\ |\chi_1(x_2)\rangle & |\chi_2(x_2)\rangle & \cdots & |\chi_n(x_2)\rangle \\ \vdots & \vdots & \ddots & \vdots \\ |\chi_1(x_n)\rangle & |\chi_2(x_n)\rangle & \cdots & |\chi_n(x_n)\rangle \end{vmatrix} \tag{2.7}$$

Where x_i are the spatial and spin coordinates of electron i. The spin orbitals $|\chi_i(x_i)\rangle$ have the following form:

$$|\chi_i(x_i)\rangle = |\psi_i(r_i)\rangle\,|\sigma_i\rangle \tag{2.8}$$

$\psi_i(r_i)$ is the spatial orbital, that depends on the spatial coordinates r_i, and $|\sigma_i\rangle$ corresponds to the spin, which can be either α or β.

HF is a variational method, i.e., the energy of a trial wave function (Ψ_t) is always higher than the exact ground state energy unless Ψ_t is equal to the exact wave function. Therefore, the solutions can be obtained iteratively by varying the parameters of an initial trial wave function until the lowest possible energy is reached. HF minimizes the energy of its wave function (the Slater determinant) by optimizing the spin orbitals $|\chi(x_i)\rangle$ according to the following equation:

$$\hat{f}(x_i)\,|\chi(x_i)\rangle = \varepsilon_v\,|\chi(x_i)\rangle \tag{2.9}$$

Where ε_v are the orbital energies and $\hat{f}(x_i)$ is the Fock operator:

$$\hat{f}(x_i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{N} \frac{Z_A}{r_{iA}} + v^{HF}(x_i) \tag{2.10}$$

$v^{HF}(x_i)$ acts as an average potential on electron i due to the rest of the electrons in the system and it is expressed as:

$$v^{HF}(x_i) = \sum_{b=1}^{n} \left(\hat{J}_b(x_i) - \hat{K}_b(x_i)\right) \tag{2.11}$$

Where $\hat{J}_b(x_i)$ and $\hat{K}_b(x_i)$ are the coulomb and exchange operators respectively. Both $\hat{J}_b(x_i)$ and $\hat{K}_b(x_i)$ depend on the spin orbitals and, thus, the Fock operator depends on its own eigenfunctions. As a result, equations 2.9 can only be solved iteratively.

The restricted Hartree Fock (RHF) formulation assumes that each spatial orbital $|\psi_i\rangle$ can be occupied by two electrons that differ on their spin. Taking them as a linear combination of atomic orbitals (LCAO) results in:

$$|\psi_i(r_i)\rangle = \sum_{\mu=1}^{k} c_{\mu i} \left|\phi_\mu(r_i)\right\rangle \tag{2.12}$$

$\left|\phi_\mu\right\rangle$ is a set of predefined basis functions centered on k atoms (more on this later) and $c_{\mu i}$ are their corresponding coefficients. Substituting equation 2.12 on equation 2.9 and multiplying both sides by $\langle\phi_\nu(r_i)|$:

$$\sum_{\mu=1}^{k} c_{\mu i} \langle\phi_\nu(r_i)| \hat{f}(r_i) \left|\phi_\mu(r_i)\right\rangle = \varepsilon_\nu \sum_{\mu=1}^{k} c_{\mu i} \left\langle\phi_\nu(r_i)\middle|\phi_\mu(r_i)\right\rangle \tag{2.13}$$

These are known as the Roothaan-Hall equations, that can be written in matrix form as:

$$FC = SCE \tag{2.14}$$

Where F is the Fock matrix, S is the overlap matrix, C is the matrix containing the coefficients of the spatial orbitals, and E contains the orbital energies. The Self-Consistent-Field (SCF) procedure solves the Roothaan-Hall equations and obtains the coefficients in C and the orbital energies iteratively. The final HF wave function consists of the $n/2$ lowest lying spatial orbitals, and the total HF energy is then calculated from these results.

On the other hand, the unrestricted Hartree Fock (UHF) formulation allows the opposite-spin spin orbitals to differ on their spatial parts. In this case, the Pople-Nesbet equations need to be solved:

$$F^\alpha C^\alpha = SC^\alpha E^\alpha \tag{2.15}$$

$$F^\beta C^\beta = SC^\beta E^\beta \tag{2.16}$$

HF fails to give the exact energy and it is not suitable to study many chemical problems. The difference between the exact non-relativistic energy and the lowest HF energy is the correlation energy, which accounts for the electron-electron interaction that the HF average potential does not consider. Many methods aim at recovering this correlation energy. In particular, we can distinguish between dynamic and static correlation. The former is related with the instant correlation of electrons, while the latter is associated with a more permanent avoidance between electrons.

2.1.3 Multi-Configurational Self-Consistent-Field

Certain chemical problems, such as electronic excited states calculations, require multiple Slater determinants to get an accurate description of the system. In these cases, some configurations are nearly degenerate and, hence, they contribute equally to the wave function. Multi-Configurational Self-Consistent-Field (MCSCF) methods [31] deal with this kind of systems. Their wave function consists of a linear combination of configurations (equation 2.17), and, in this way, they recover the static correlation.

$$|\Psi_{MC}\rangle = \sum_{m=1}^{N_{CSF}} c_m |\Psi_m\rangle \tag{2.17}$$

Where c_m are the expansion coefficients and $|\Psi_m\rangle$ are Configuration State Functions (CSF), i.e., Slater determinants built from orthonormal orbitals where opposite spin electrons share the same spatial part:

$$|\Psi_m\rangle = |\psi_i\bar{\psi}_i, \cdots, \psi_{N_e}\bar{\psi}_{N_e}\rangle_m \tag{2.18}$$

ψ_i are the spatial orbitals $\psi_i = \sum_{\mu=1}^{N_{AO}} c_{\mu i} \left|\phi_\mu\right\rangle$, where $c_{\mu i}$ are the expansion coefficients of the atomic orbitals $\left|\phi_\mu\right\rangle$. In MCSCF methods, both c_m and $c_{\mu i}$ are optimized simultaneously via an iterative process.

Complete Active Space Self-Consistent-Field

The computational cost of MCSCF methods depends on the number of configurations that are considered in Equation 2.17. One of the most common strategies to reduce the number of CSFs, and hence, the computational cost, is the Complete Active Space Self-Consistent-Field (CASSCF) method.[32, 33]

As shown in Figure 2.1, the orbitals in the CASSCF wave function are divided in three different regions: inactive, active and virtual. The inactive and virtual orbitals are always doubly occupied and empty, respectively. On the other hand, the occupation of the active orbitals varies between 0 and 2. These orbitals constitute the most important region: the Active Space (AS). The CASSCF wave function includes all the possible combinations of excitations (single, double, triple, etc) involving the orbitals in the AS. In this context, single (doubly, triply, etc) excited configurations are built from single (double, triple, etc) excitations in which one (two, three, etc) electron(s) move from an occupied orbital to an unoccupied orbital with respect to the HF wave function. Then, the excited configurations resulting from all the possible combinations are included in Equation 2.17, performing what is known as a full Configuration Interaction (CI) calculation in this region. However, as the $c_{\mu i}$ expansion coefficients of the orbitals are optimized in all subsets, the orbitals can exchange between the AS and the inactive/virtual regions in some cases. To further reduce the computational cost, some frozen orbitals, which are usually core orbitals, can be specified. In this region, the $c_{\mu i}$ orbital coefficients are taken from HF.

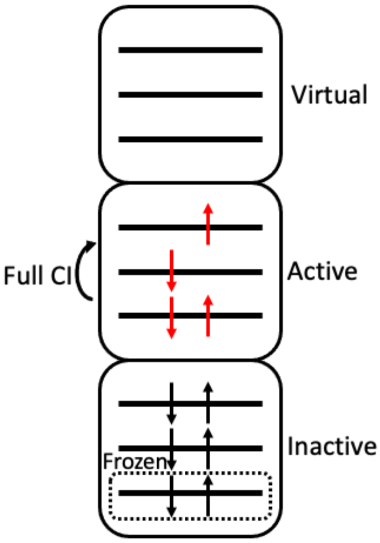

Fig. 2.1 Representation of the inactive, active and virtual regions in CASSCF calculations. The dotted region represents the frozen orbitals, which are usually core orbitals.

A major drawback of CASSCF is that it is not a "black box" method. The user needs to specify the orbitals in the AS and the results are highly dependent on this choice. There is no systematic procedure to select the most appropriate AS, however, there are some general rules:

- Orbitals with occupation numbers very close to 2 or 0 should belong to the inactive and virtual regions, respectively. On the other hand, orbitals with occupation numbers significantly different from 0 or 2 should be considered in the AS.

- The HF highest energy occupied and lowest energy unoccupied orbitals should also belong to the AS.

- The AS must contain the most important orbitals to describe the chemical problem (chemical intuition). For example, in electronic excited states calculations, π and π^* orbitals are usually relevant. Note that, if a particular π orbital is in the AS, the corresponding π^* antibondig should also be included. The same applies to σ and σ^* orbitals, which are often involved in bond breaking.

The CASSCF wave function for excited states is expressed as:

$$\left|\Psi_{root}^{MC}\right\rangle = \sum_{m=1}^{N_{CSF}} c_{m,root} \left|\Psi_m\right\rangle \tag{2.19}$$

Where the $c_{m,root}$ coefficients differ for each electronic state (root). Then, the electronic states can be obtained from independent CASSCF calculations, in which, in each one of them, a single electronic state is optimized. However, the individual optimization of the electronic states may result in a change in their order (root flipping), leading to convergence problems. The State Average CASSCF (SA-CASSCF) method can avoid this complication. In this formalism, the root average energy is optimized, and the wave function consists of a linear combination of the electronic states wave functions:

$$\left|\Psi_{SA}^{MC}\right\rangle = \sum_{root=1}^{N_{roots}} \omega_{root} \left|\Psi_{root}^{MC}\right\rangle \tag{2.20}$$

Where ω_{root} is the weight of each electronic state. Specific c_m coefficients are obtained for each of the states. On the contrary, all states in SA-CASSCF share a common set of orbitals and, thus, the $c_{\mu i}$ are not optimized for the individual states. As a result, their description is poorer when the number of states in the SA-CASSCF calculation is higher.

Complete Active Space Second Order Perturbation Theory

MCSCF methods fail to include the dynamic correlation, which appears in configurations having a small contribution to the wave function. In this sense, the Complete Active Space Second Order Perturbation Theory (CASPT2) [34, 35] aims at recovering the missing dynamic correlation by means of second order perturbation theory.[36] The Hamiltonian in perturbation theory consists of an unperturbed Hamiltonian $\hat{H}^{(0)}$ and a perturbed Hamiltonian $\hat{H}'$ of small magnitude:

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}' \tag{2.21}$$

λ ($0 \leq \lambda \leq 1$) is an artifact that quantifies the extent of the perturbation. Exact solutions $\left|\Psi_i^{(0)}\right\rangle$ exist for the unperturbed, or zeroth order Hamiltonian:

$$\hat{H}^{(0)}\left|\Psi_i^{(0)}\right\rangle = E_i^{(0)}\left|\Psi_i^{(0)}\right\rangle \tag{2.22}$$

Where $E_i^{(0)}$ is the zeroth order energy. To find the perturbed solutions, the wave function and the energy are expanded as a Taylor series:

$$E_i(\lambda) = E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \lambda^3 E_i^{(3)} + \cdots \tag{2.23}$$

$$|\Psi_i(\lambda)\rangle = \left|\Psi_i^{(0)}\right\rangle + \lambda\left|\Psi_i^{(1)}\right\rangle + \lambda^2\left|\Psi_i^{(2)}\right\rangle + \lambda^3\left|\Psi_i^{(3)}\right\rangle + \cdots \tag{2.24}$$

Substituting into the time-independent Schrödinger equation and reordering according to the powers of λ:[37]

$$\lambda^0 : \hat{H}^{(0)}\left|\Psi_i^{(0)}\right\rangle = E_i^{(0)}\left|\Psi_i^{(0)}\right\rangle \tag{2.25}$$

$$\lambda^1 : \hat{H}^{(0)}\left|\Psi_i^{(1)}\right\rangle + \hat{H}'\left|\Psi_i^{(0)}\right\rangle = E_i^{(0)}\left|\Psi_i^{(1)}\right\rangle + E_i^{(1)}\left|\Psi_i^{(0)}\right\rangle \tag{2.26}$$

$$\lambda^2 : \hat{H}^{(0)}\left|\Psi_i^{(2)}\right\rangle + \hat{H}'\left|\Psi_i^{(1)}\right\rangle = E_i^{(0)}\left|\Psi_i^{(2)}\right\rangle + E_i^{(1)}\left|\Psi_i^{(1)}\right\rangle + E_i^{(2)}\left|\Psi_i^{(0)}\right\rangle \tag{2.27}$$

$$\lambda^k : \hat{H}^{(0)}\left|\Psi_i^{(k)}\right\rangle + \hat{H}'\left|\Psi_i^{(k-1)}\right\rangle = \sum_{j=0}^{k} E_i^{(j)}\left|\Psi_i^{(k-j)}\right\rangle \tag{2.28}$$

Then, the kth order energy correction can be obtained as:

$$E_i^{(k)} = \left\langle\Psi_i^{(0)}\middle|\hat{H}'\middle|\Psi_i^{(k-1)}\right\rangle \tag{2.29}$$

Taking into account the intermediate normalization condition $\left\langle\Psi_i\middle|\Psi_i^{(0)}\right\rangle = 1$, the second order correction energy is:

$$E_i^{(2)} = \sum_{k\neq i} \frac{\left\langle\Psi_i^{(0)}\middle|\hat{H}'\middle|\Psi_k^{(0)}\right\rangle\left\langle\Psi_k^{(0)}\middle|\hat{H}'\middle|\Psi_i^{(0)}\right\rangle}{E_i^{(0)} - E_k^{(0)}} \tag{2.30}$$

CASPT2 uses the CASSCF wave function as the unperturbed $\left|\Psi_i^{(0)}\right\rangle$, and the CASSCF Fock operator as the zeroth order Hamiltonian $\hat{H}^{(0)}$. Equation 2.30 in CASPT2 excited states calculations translates as:

$$E_{root}^{(2)} = \sum_{k} \frac{\left\langle \Psi_{root}^{(0)} \middle| \hat{H}' \middle| \Psi_{k}^{(0)} \right\rangle \left\langle \Psi_{k}^{(0)} \middle| \hat{H}' \middle| \Psi_{root}^{(0)} \right\rangle}{E_{root}^{(0)} - E_{k}^{(0)}} \tag{2.31}$$

Where k are all the considered single and double excitations and $E_{root}^{(2)}$ is the State Specific CASPT2 (SS-CASPT2) energy. The SS-CASPT2 wave function is the first order wave function:

$$\left| \Psi_{root}^{(SS-CASPT2)} \right\rangle = c_0 \left| \Psi_{root}^{(0)} \right\rangle + c_1 \left| \Psi_{root}^{(1)} \right\rangle \tag{2.32}$$

Where c_0 and c_1 are the weights of the reference wave function and the first order correction, respectively. The reference weight (c_0) is used to asses the CASPT2 calculation as it should be large enough to consider perturbation theory a good approximation. Moreover, it should also have a similar magnitude for all the electronic states because the perturbation is supposed to be comparable in all roots. When the reference weights significantly differ from one another, there may be "intruder states"[38] in the AS of the CASPT2 calculation. The corresponding $E_{k}^{(0)}$ energy of the intruder states is very close to $E_{root}^{(0)}$ and, hence, the denominator in Equation 2.31 tends to 0. When there is only one intruder state, the best approach to solve this issue is increasing the active space to include the orbitals that this state involves. Otherwise, a small constant, known as the level shift [38], can be added to the zeroth order Hamiltonian. In this way, the CASPT2 correction energy:

$$E_{root}^{(2)} = \sum_{k} \frac{\left\langle \Psi_{root}^{(0)} \middle| \hat{H}' \middle| \Psi_{k}^{(0)} \right\rangle \left\langle \Psi_{k}^{(0)} \middle| \hat{H}' \middle| \Psi_{root}^{(0)} \right\rangle}{E_{root}^{(0)} - E_{k}^{(0)} + \varepsilon} \tag{2.33}$$

Where ε can be both a real or an imaginary shift. Similarly, it was shown that introducing an empirical parameter, the IPEA shift, in the denominator improves the dissociation energies of diatomic molecules and the electronic excitations in some polyatomic molecules.[39]

In the same way as CASSCF, SS-CASPT2 only calculates the second order correction of a single root, without considering its interactions with the other electronic states. On the other hand, Multi-State CASPT2 (MS-CASPT2)[40] takes all the excited states of interest into account. To do this, it uses an effective Hamiltonian whose diagonal elements are the SS-CASPT2 energies, and the r, s non-diagonal elements correspond to the $\left\langle \Psi_{r}^{(0)} \middle| \hat{H} \middle| \Psi_{s}^{(1)} \right\rangle$ coupling between states r and s. Its diagonalization leads to the MS-CASPT2 energies and the Perturbated Modified CASSCF (PMCASSCF) wave function:

$$\left|\Psi_{root}^{(PMCASSCF)}\right\rangle = \sum_{i=1}^{N_{roots}} c_i^{(0)} \left|\Psi_i^{(0)}\right\rangle \tag{2.34}$$

2.1.4 Density Functional Theory

Similarly to MCSCF methods, Density Functional Theory (DFT)[41] also aims to include the correlation energy that is missing in HF. However, instead of optimizing the wave function, it focuses on the electron density. The electron density is the number of electrons per unit volume and, thus, it is expressed as:

$$\rho(r) = n \int_0^{\infty} |\Psi(r_1, \cdots, r_n)|^2 dr_1 \cdots dr_n \tag{2.35}$$

No matter the number of electrons in the system, the optimization of the electron density is a one-particle problem that depends only on 3 spatial coordinates. Therefore, calculations are much faster compared to wave function methods including correlation and, in practice, the computational cost is similar to HF, which makes calculations affordable for bigger chemical systems. Moreover, the electron density contains the same information as the wave function since, most properties, such as the number of electrons, the nuclei position and the nuclear charges, can be extracted from it.

Hohenberg-Kohn Theorems

Hohenberg and Kohn [42] proposed two theorems that are the foundations of DFT. The first theorem, also known as the proof of existence, considers that the electron density creates an external potential V_{ext} in which the electrons move. By reductio ad absurdum, it is possible to prove that there cannot be two different V_{ext} that yield exactly the same $\rho(r)$. This means that V_{ext} is an unique functional of $\rho(r)$. In addition, as V_{ext} fixes the Hamiltonian ($\hat{H} = \hat{T} + \hat{V}_{ee} + \hat{V}_{ext}$), the many particle ground state is also a unique functional of $\rho(r)$.[41] Therefore, all the properties of the system can be obtained from $\rho(r)$. In particular, the ground state energy as a function of the electron density is defined as:

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] + V_{nn}[\rho] \tag{2.36}$$

Where $T[\rho]$, $V_{ne}[\rho]$, $V_{ee}[\rho]$ and $V_{nn}[\rho]$ are the kinetic energy, nuclear-electron, electron-electron and nuclear-nuclear potentials respectively. Within the Born-Oppenheimer approximation, $V_{nn}[\rho]$ is a constant. In addition, $V_{ne}[\rho]$ can be expressed in terms of the potential created by the nuclei (v_{ext}):

$$V_{ne}[\rho] = \int v_{ext}(r)\rho(r)dr \tag{2.37}$$

$T[\rho]$ and $V_{ee}[\rho]$ do not depend on v_{ext} so they are both included in the Hohenberg-Kohn functional ($F_{HK}[\rho]$) according to $F_{HK}[\rho] = T[\rho] + V_{ee}[\rho]$. In this way, the electronic ground state energy as a function of the density reduces to:

$$E_v[\rho] = \int v_{ext}(r)\rho(r)dr + F_{HK}[\rho] \tag{2.38}$$

The second Hohenberg-Kohn theorem states that DFT is variational: a trial electron density will always provide a higher energy than the exact ground-state. Hence, a more accurate density can be obtained by minimizing $E_v[\rho]$:

$$\left[\frac{\partial E_v[\rho]}{\partial \rho}\right] = v_{ext}(r) + \frac{\partial F_{HK}[\rho]}{\partial \rho} \tag{2.39}$$

Solving Equation 2.39 would result in the exact ground state energy. However, as the exact F_{HK} is unknown, DFT is not variational in practice because of the use of approximate functionals.

The Kohn-Sham Approach

In order to develop approximate functionals, Kohn and Sham defined a fictitious system of non-interacting particles whose electron density is exactly the same as in the real system of interacting particles.[37] The non-interacting Hamiltonian is given as:

$$\hat{H}_{KS} = -\frac{1}{2}\nabla^2 + V_{KS}(r) \tag{2.40}$$

Where the first term is the electronic kinetic energy of the non-interacting particles, and $V_{KS}(r)$ is the Kohn-Sham potential: the constant external potential that determines the density of both fictitious and real systems. This leads to:

$$\hat{H}_{KS}\chi_i^{KS} = \varepsilon_i \chi_i^{KS} \tag{2.41}$$

In the same way as in HF, equation 2.41 is solved self-consistently and the spatial part of the χ_i^{KS} solutions are the Kohn-Sham orbitals. χ_i^{KS} are equivalent to the HF Slater determinants and describe the orbitals of the non-interacting system. As a result, the computational problem depends on 3N variables. Then, the electron density can be obtained according to:

$$\rho(r) = \sum_{i=1}^{N_{elec}} |\chi_i^{KS}|^2 \tag{2.42}$$

The procedure explained above is only possible if the exact Hamiltonian in equation 2.40 is known. Therefore, the form of $V_{KS}(r)$ needs to be determined. In this context, equation 2.36 translates as:

$$E[\rho] = T_{KS}[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.43}$$

$T_{KS}[\rho]$, $V_{ne}[\rho]$, and $J[\rho]$ can be calculated exactly from the Kohn-Sham orbitals. According to equation 2.40, the kinetic energy of the non-interacting system is:

$$T_{KS}[\rho] = -\frac{1}{2}\sum_{i=1}^{N_{elec}} \left\langle \chi_i^{KS} \middle| \nabla^2 \middle| \chi_i^{KS} \right\rangle \tag{2.44}$$

Regarding the nuclear-electron potential $V_{ne}[\rho]$, it is calculated as:

$$V_{ne}[\rho] = -\sum_{i=1}^{N_{nuc}} Z_i \int \frac{\rho(r)}{|r - R_i|} dr \tag{2.45}$$

Then, $J[\rho]$, which is part of the $V_{ee}[\rho]$ electron interaction, corresponds to the classical Coulomb electron-electron interaction:

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 \tag{2.46}$$

Finally $E_{xc}[\rho]$ includes the exchange-correlation non-classical effects, correction to self-interaction (the interaction of one electron with itself), and the kinetic energy that is not included in the non-interacting system:

$$E_{xc}[\rho] = (T[\rho] - T_{KS}[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{2.47}$$

Therefore, the Hohenberg-Kohn functional ($F_{HK}[\rho]$) here is:

$$F_{HK}[\rho] = T_{KS}[\rho] + J[\rho] + E_{xc}[\rho] \tag{2.48}$$

And, as the potential is the partial derivative of the potential energy with respect to the density, the final form of $V_{KS}(r)$ is:

$$V_{KS}(r) = \sum_{i}^{N_{nuc}} \frac{Z_i}{|r - R_i|} + \int \frac{\rho(r_2)}{|r_1 - r_2|} dr_2 + \frac{\partial E_{xc}[\rho]}{\partial \rho(r)} \tag{2.49}$$

Unfortunately, $E_{xc}[\rho]$ is unknown and approximations are needed to perform the calculations. Similarly to HF, the SCF procedure starts with a trial set of orbitals that are updated iteratively until convergence is reached. However, as the exact functional is unknown, the final energy can be lower than the exact energy of the system, i.e., DFT is not variational in practice.[37]

The Exchange-Correlation Functional

As mentioned in the previous section, the exchange-correlation functional is still unknown and, thus, many strategies have been developed to find an accurate approximation.[37] In general, the exchange-correlation functional is divided into the exchange and the correlation:

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] \tag{2.50}$$

The most simple approaches are the Local Density Approximations (LDA), which are based on the uniform electron gas model. They consider the electron density locally and they assume that it is finite and constant. As a result, they are not suitable to study systems with non-uniform electron density such as molecules. In addition, they tend to overestimate electron correlation and underestimate the exchange. VWN is an example belonging to this rung and was proposed by Vosko, Wilk and Nusair.[43]

General Gradient Approximations (GGA) emerged as a way to improve the LDA performance. In the same way as LDA, they consider the electron density locally but, in addition, they also include the gradient of the electron density. In this way, they improve the description of chemical systems compared to LDA. A famous functional that belongs to this rung is BLYP, which combines the exchange proposed by Becke, and the correlation by Lee, Yang and Parr.[44, 45] However, they are still far from accurate as, for instance, they tend to underestimate reaction barriers. More precise results are usually obtained with meta-GGA functionals since they also consider the second derivative of the electron density. Truhlar and coworkers developed one of the most important meta-GGA functionals, which are known as the Minnesota functionals.[46]

Unlike pure density functionals, like LDA and GGA, hybrid functionals consist of a linear combination of HF exchange and the previous explicit exchange-correlation functionals. They usually provide accurate results and, hence, they are widely spread in the computational chemistry community. Most probably, B3LYP[47] is the most famous hybrid functional.

All these functionals struggle with the description of non-covalent medium range interactions, such as van der Waals energies. One way to correct this is including an additional dispersion energy term (E_{disp}) in the functional, which is usually calculated empirically.

$$E_{total} = E_{DFT} + E_{disp} \tag{2.51}$$

Alternatively, generalized random phase methods consider both occupied and non-ocuppied Kohn-Sham orbitals and, hence, the description of the dispersion improves. The OEP2 functional [48] belongs to this rung. Another option is the use of range-separated functionals, where the amount of HF exchange varies depending on the distance. They are important to describe charge-transfer. Two examples are CAM-B3LYP[49] and ωB97X[50].

Unfortunately, there is no systematic way of improving the functionals and, therefore, a proper benchmarking is always a must. Nevertheless, the functionals can be classified in the so called Jacob's ladder according to their expected accuracy. It ranges from the HF result to the exact solution and it follows the order: LDA, GGA, Meta-GGA, Hybrid functionals and Generalized random phase methods.[51]

Time-Dependent Density Functional Theory

Electronic excitations occur upon irradiation with a time-dependent potential. Therefore, excited states are no longer stationary states and, hence, the time-dependent Schrödinger equation (equation 2.1) needs to be solved. However, this is even more complicated than calculating the ground-state so more approximations are needed.

In this context, Runge and Gross proved that, for a given initial state, there is a one-to-one correspondence between the time-dependent external potential $v_{ext}(r,t)$ and the time-dependent density of the system $\rho(r,t)$. This is equivalent to the first Hohenberg-Kohn theorem and it means that $v_{ext}(r,t)$ is a unique functional of the $\rho(r,t)$ evolution in time.[52] Following the same strategy as in conventional DFT, we can define a fictitious system of non-interacting particles that fulfil the time-dependent Kohn-Sham equations:

$$i\frac{\partial \chi_i^{KS}(r,t)}{\delta t} = \left[-\frac{1}{2}\nabla^2 + v_{KS}[\rho](r,t) \right] \chi_i^{KS}(r,t) \tag{2.52}$$

In the same way as before, $v_{KS}[\rho](r,t)$ yields the temporal evolution of the density of both the non-interacting and real systems. Once again, the density can be calculated from the time-dependent Kohn-Sham orbitals $\chi_i^{KS}(r,t)$:

$$\rho(r,t) = \sum_{i=1}^{N_{elec}} |\chi_i^{KS}(r,t)|^2 \tag{2.53}$$

Likewise, the final form of $v_{KS}[\rho](r,t)$ resembles equation 2.49:

$$v_{KS}[\rho](r,t) = v_{ne}(r,t) + v_H(r,t) + v_{xc}(r,t) \tag{2.54}$$

Where $v_H(r,t)$ is the Hartree potential, which plays the same role as the classical Coulomb potential in equation 2.49:

$$v_H(r,t) = \int \frac{\rho(r_2,t)}{|r_1 - r_2|} dr_2 \tag{2.55}$$

All efforts in TDDFT focus on finding approximations to the exchange-correlation potential $v_{xc}(r,t)$, which is still unknown. According to perturbation theory, when the time-dependent potential is weak, we only need to consider the form of $v_{xc}(r,t)$ close to the initial state, which is taken to be the ground-state. This is the case of most spectroscopic experiments and, therefore, their time-dependent density can be written as:

$$\rho(r,t) = \rho_{GS}(r,t) + \delta\rho(r,t) \tag{2.56}$$

Hence, the exchange-correlation potential $v_{xc}(r,t)$:

$$v_{xc}[\rho_{GS} + \delta\rho](r,t) = v_{xc}[\rho_{GS}](r) + \int\int f_{xc}[\rho_{GS}](r,r',t-t')\delta\rho(r',t')dr'dt' \tag{2.57}$$

where f_{xc} is the exchange-correlation kernel and depends exclusively on the ground-state density according to:

$$f_{xc}[\rho_{GS}](r,r',t-t') = \left.\frac{\delta v_{xc}(r,t)}{\delta\rho(r',t')}\right|_{\rho=\rho_{GS}} \tag{2.58}$$

This approach is known as linear response theory and it is much simpler than dealing with the full exchange-correlation because, in this case, $v_{xc}(r,t)$ is a functional of just the ground state density.[52]

Another important concept in linear response is the point-wise susceptibility $S[\rho_{GS}]$.[52] It is defined as the impact that a small change in the external potential has on the ground state density. The non-interacting system also has its own S_{KS}, which differs from $S[\rho_{GS}]$, but, according to Kohn-Sham, both must lead to the same density response. Moreover, it is possible to relate the single-particle excitations of the non-interacting system with the S_{KS}:

$$S_{KS}(r,r',\omega) = 2\ \lim_{\eta\to 0^+} \sum_q \left[\frac{\xi_q(r)\xi_q^*(r')}{\omega-\omega_q+i\eta} - \frac{\xi_q^*(r)\xi_q(r')}{\omega+\omega_q-i\eta} \right] \tag{2.59}$$

Where $\xi_q(r) = \chi_i^*(r)\chi_a(r)$ and ω_q is the energy that represents the electronic transition from an occupied i to an unocuppied a Kohn-Sham orbital ($\omega_q = \varepsilon_a - \varepsilon_i$). Hence, when ω matches ω_q, the susceptibility function presents a peak that corresponds to an electronic excitation. However, the transitions in the real system differ from the ones in equation 2.59 unless the Hartree exchange-correlation effects are absent. In this sense, for frequency independent kernels, Casida showed that it is possible to calculate the transitions as an eigenvalue problem, where the eigenfunctions are the oscillator strengths, and the eigenvalues are the excitation energies. Most computational chemistry codes use this approach.[52]

2.1.5 Basis sets

The molecular orbitals in wave function methods and in DFT are expressed as a linear combination of basis functions:

$$|\psi_i(r_i)\rangle = \sum_{\mu=1}^{M_{basis}} c_{\mu i} \left|\phi_\mu(r_i)\right\rangle \tag{2.60}$$

where $c_{\mu i}$ are the coefficients that are optimized in the SCF procedure and $\left|\phi_\mu(r_i)\right\rangle$ are the basis functions. In the complete basis set limit, it is possible to recreate any given function. However, this limit is unreachable as it requires infinite functions. Therefore, in practice, basis sets can only yield approximate molecular orbitals that will be more or less accurate depending on the chosen basis functions.[37]

We can distinguish between Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO):

$$\phi^{STO}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}r^{n-1}e^{-\zeta r} \tag{2.61}$$

$$\phi^{GTO}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}r^{2n-2-l}e^{-\zeta r^2} \tag{2.62}$$

Where ζ is the nuclear effective charge; n, l and m are the quantum numbers; r, θ and φ are the polar coordinates; N is the normalization constant and $Y_{l,m}$ is the spherical harmonic function. STO orbitals reproduce hydrogenlike orbitals and, hence, they are very accurate. However, the calculation of the required integrals, which can only be solved numerically, is very computationally demanding. On the other hand, calculations with GTOs are much more efficient, making them the usual preferred choice.[37] In order to improve the accuracy of this kind of basis, a linear combination of gaussians is used to recreate STOs:

$$\phi^{CGTO} = \sum_{i}^{k} a_i \phi_i^{GTO} \tag{2.63}$$

Where ϕ_i^{GTO} are the primitive functions (PGTO) and a_i are the corresponding coefficients, which remain fixed during the calculations. ϕ^{CGTO} are known as the contracted GTOs and they allow a proper description of the molecular orbitals and the efficient calculation of the integrals. The larger the expansion in equation 2.63, the better the description of the STO.

The smallest basis that can be used, also known as the minimal basis set, contains only one basis function/orbital per electron. In order to improve their flexibility, and hence, their description, the number of orbitals can be multiplied by an integer number X. This leads to the XZ basis sets, including the Double Zeta (DZ) and the Triple Zeta (TZ), among others. Taking into account that core orbitals are not usually involved in chemical bonding, it is often convenient to include additional functions to valence orbitals. This results in the split valence basis VXZ.[53]

In some cases, VXZ basis sets are still not enough. When weak interactions, such as van der Waals or hydrogen bonds, are relevant, diffuse functions should be included. The latter consist of a small exponential term that allows the proper representation of charges that are far from the nuclei. Similarly, polarization is needed when functions with larger angular momentum are required. This usually happens in bond forming/breaking contexts.[53, 54]

In this master thesis, the Correlation Consistent-polarized Valence Double Zeta (cc-pVDZ)[55] basis set has been used. In Correlation Consistent basis sets, the exponents and the coefficients of the primitive functions were optimized including electron correlation in the calculations. Therefore, they tend to provide very accurate results.

2.1.6 Wave Function Analysis

The assignment of wave function character is crucial in the analysis of excited states calculations. However, it can be challenging in certain situations that require the processing of huge datasets or the interpretation of effects that arise in larger systems. Hence, in these cases, it is convenient to make use of methods that automatise the wave function analysis.

Fragment-based excited state analysis as implemented in TheoDORE [56] has been performed in this work. In this method, the system is divided into different units and the positions of the hole and the excited electron involved in the transition are analyzed to classify the states. For example, in Figure 2.2a), the excitation is local as both the hole and the electron belong to fragment 1. Alternatively, there is charge transfer in Figure 2.2b) since there is an electron transition between two different fragments. Moreover, multiple local excitations are also possible. This is the case of Figure 2.2c), where a dominant contribution in fragment 2 is coupled with two additional excitations in 1 and 3. Finally, Figure 2.2d) shows a linear combination of two opposite direction charge transfer states. Apart from the location of the electron-hole pair, the dynamic connection that the arrows represent is also relevant. This is what distinguishes Figures 2.2 c) and d).

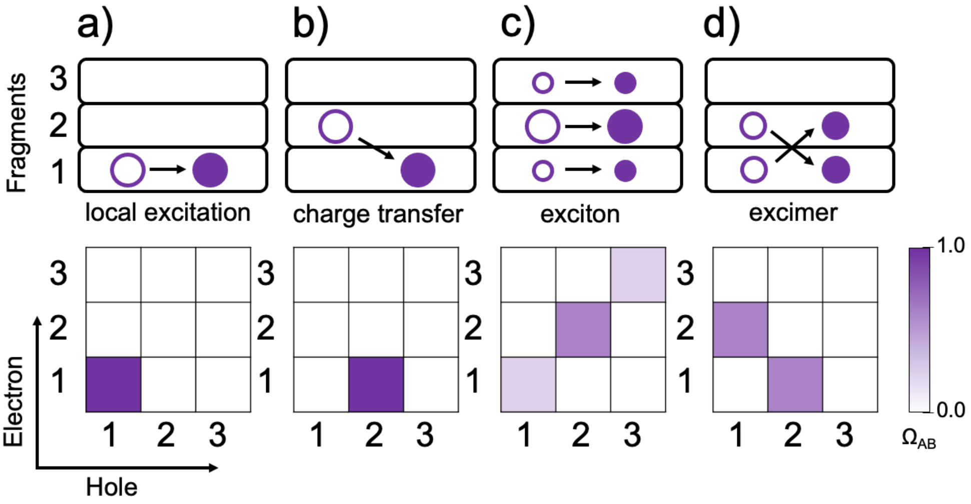

Fig. 2.2 Top pannel: representation of different types of excited states depending on the localisation of the hole and the electron, and their dynamic connection. Bottom pannel: electron-hole correlation plots for three fragment systems

In this context, the one-electron transition density matrix (1TDM) between the ground and excited states is considered to describe the electron-hole pair:

$$\gamma_{0I}(r_h, r_e) = n \int \cdots \int \Psi_0(r_h, r_2, \cdots, r_n)\Psi_I(r_e, r_2, \cdots, r_n) dr_2 \cdots dr_n \tag{2.64}$$

where Ψ_0 and Ψ_I are the ground state and excited wave functions respectively. r_h and r_e are the coordinates of the hole and the electron in Ψ_0 and Ψ_I. From this, the charge transfer numbers (Ω_{AB}) can be computed:

$$\Omega_{AB} = \int_A \int_B |\gamma_{0I}(r_h, r_e)|^2 dr_e dr_h \tag{2.65}$$

where the hole and the electron are located in fragments A and B respectively. All Ω_{AB} values can be collected into a nF x nF Ω matrix where nF is the number of fragments. Figure 2.2 shows a visual representation of the corresponding Ω matrices from the previous examples. A darker colour indicates a larger Ω_{AB}. In this sense, for Figure 2.2 a) $\Omega_{11} = 1$ and the rest are 0. Therefore, only the bottom-left corner is coloured. Similarly, for Figure 2.2 b), just $\Omega_{21} = 1$ is coloured as the excitation goes from 2 to 1. Furthermore, the local excitations in Figure 2.2 c) render a less intense non-vanishing diagonal. On the other hand, the charge-transfer states from Figure 2.2 d) are represented in the non diagonal terms. Finally, the initial and final positions, corresponding to the hole and the electron respectively, can be obtained from Ω_{AB}. In this way, the mean position of the excitation (POS) can be derived, which is sometimes useful to characterize the $n\pi^*$ or $\pi\pi^*$ nature of the involved transitions.

2.2 Molecular Mechanics

2.2.1 Molecular Dynamics

Large systems, like most biomolecules, present a huge number of stable conformations in their PES due to their great amount of degrees of freedom. Therefore, a proper sampling of the configuration space is necessary in order to get an accurate description of the system. Unfortunately, a full electronic structure calculation is not feasible and, hence, classical approximations are more appropriate as they allow the generation of configurations at a lower computational cost. According to the ergodic theorem[57], the average of all the configurations is equal to the average geometry of a single individual over time. In this sense, one of the most popular approaches to sample the configuration space is Molecular Dynamics (MD), which is based on the time propagation of the Hamilton's equations of motion (Equations 2.68 and 2.69).[58]

Equations of Motion

The Hamilton's equations of motion can be derived from the Newton's second law and from the kinetic energy expression. Starting from the Newton's second law:

$$F_i = \frac{dp_i}{dt} \tag{2.66}$$

where F_i is the force and p_i the momentum of atom i, and knowing that:

$$F_i = -\frac{dV}{dr_i} \tag{2.67}$$

where V is the potential energy of a conservative system. Then, equating 2.66 and 2.67:

$$\frac{dp_i}{dt} = -\frac{dV}{dr_i} \tag{2.68}$$

which is the first Hamilton equation. Regarding the second, it is obtained by deriving the kinetic energy (T_i) with respect to the momentum:

$$\frac{dT_i}{dp_i} = \frac{d}{dp_i}\left(\frac{p_i^2}{2m_i}\right) = \frac{2p_i}{2m_i} = \frac{m_i v_i}{m_i} = v_i = \frac{dr_i}{dt} \tag{2.69}$$

where m_i is the mass of the particle and v_i its velocity.[58] Hence, both Hamilton equations predict the evolution of the coordinates and the momentum of each atom with time.

Integrators

The equations of motion need to be integrated numerically in order to obtain the temporal evolution of the system. Starting from an initial geometry, the equations are solved according to a specified time-step. Ideally, the chosen integration method should render an accurate approximation of the real trajectory of the system. Moreover, the energy of the system needs to be conserved, and the use of the largest possible time step is convenient to sample the configuration space faster.

One of the most popular integrators is the velocity Verlet algorithm[59], which obtains the corresponding coordinates (Equation 2.70) and velocities (Equation 2.71) for each time step.

$$r_i(t+\Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 \tag{2.70}$$

$$v_i(t+\Delta t) = v_i(t) + \frac{1}{2}(a_i(t) + a_i(t+\Delta t))\Delta t \tag{2.71}$$

In the first step, the forces at time t are calculated from the potential energy to get the acceleration according to Newton's second law:

$$a_i = \frac{F_i}{m_i} = -\frac{1}{m_i}\frac{\partial V}{\partial r_i} \tag{2.72}$$

Then, the positions at $t+\Delta t$ are computed from the velocity, position and acceleration at time t according to Equation 2.70. Next, the accelerations at $t+\Delta t$ are again obtained from equation 2.72 as they are required for equation 2.71. Finally, the new velocities at time $t+\Delta t$ are calculated from equation 2.71.

This kind of integration schemes are known as expansion based methods since they make use of Taylor series to propagate the coordinates and the velocities. The leap-frog integrator [60] also belongs to this kind. In this case, the positions and the velocities are asynchronous: the position is calculated for $t+\Delta t$, while the velocities are at $t+\frac{\Delta t}{2}$.

No matter the integrator, the time-step needs to be no longer than half the period of the fastest vibration. This is 1 fs in systems that involve bonds with H. There are several approaches that aim to increase the time step to make the sampling faster. In this sense, the SHAKE [61] and the RATTLE [62] algorithms freeze the X-H bonds to remove the fast vibration, making 2 fs time steps feasible. Furthermore, multiple time step integration methods can be used. The latter are based on the system decomposition in different time scales according to their corresponding vibration speed.[63]

Note that all integrators require an initial set of positions and velocities to start the propagation. However, most programs can compute an estimation of the initial velocities from a Maxwell-Boltzmann distribution and, hence, thermal equilibrium is assumed.

Force Fields

The integration of the equations of motion requires the evaluation of the potential energy gradient (Equation 2.72). In this sense, electronic structure calculations provide the most accurate estimation of the potential due to its strong physical basis. However, these computations are extremely expensive and are only feasible for small systems and short simulation times. On the other hand, Molecular Mechanics (MM) methods are faster and allow the calculation of bigger systems. But, as the physical basis of MM is much poorer, it requires a significant amount of initial parametrization. In the description of the system in MM, the electrons are treated together with the nuclei as “effective classical atoms” or Atom Types. Atom Types provide information about the bonding environment of the atoms and each one of them are described by a set of parameters. For instance, in this way, it is possible to distinguish sp^3 carbons from aromatic sp^2 carbons or carboxylic sp^2 carbons.[37]

As explained earlier, the forces to integrate the equations of motion are derived from the potential energy which, in MM, is calculated using Force Fields. A Force Field is an equation that relates the potential energy of the system with its internal coordinates.[37] They are normally split in bonded and non-bonded terms as follows:

$$V_{total} = \sum_{bonds} V_{R_B} + \sum_{angles} V_\theta + \sum_{torsions} V_\tau + \sum_{atoms} V_{vdW+rep} + \sum_{atoms} V_{elstat} \tag{2.73}$$

where the bond (V_{R_B}), angle (V_θ) and the torsion (V_τ) potentials are the bonded terms, while the vdW+repulsion ($V_{vdW+rep}$) and the electrostatic (V_{elstat}) potentials are the non-bonded terms. Conventional Force Fields are not expected to describe bond formation or rupture and, thus, bond distances (one bond separation between the interacting atoms) and angles (two bond separation) should be close to equilibrium. As a result, both V_{R_B} and V_θ interactions can be described by harmonic potentials since they provide fast and good approximations:

$$V_{R_B} = \frac{1}{2} k_b (R_B - R_{B0})^2 \tag{2.74}$$

$$V_\theta = \frac{1}{2} k_a (\theta - \theta_0)^2 \tag{2.75}$$

Here, R_{B0} and θ_0 are the parameters that represent the equilibrium bond lengths and angles, respectively. Similarly, k_b and k_a are the force constant parameters, which can be derived experimentally or by means of quantum chemical calculations. Note that, if a higher accuracy is required, more terms can be included in the Taylor expansions of equations 2.74 and 2.75. Regarding the torsional angle interactions V_τ (three bond separation), a Fourier-series expansion is more appropriate for its description as the potential is periodic with respect to the dihedral angle. This can be expressed as a linear combination of cosine functions:

$$V_\tau = \sum_i \frac{1}{2} k_{\tau,i} (1 + \cos(n_{\tau,i} \phi_\tau - \delta_{\tau i})) \tag{2.76}$$

where the parameter ϕ_τ is the torsional angle, $\delta_{\tau i}$ is the phase, $n_{\tau,i}$ describes the shape of the potential, and $k_{\tau,i}$ is the torsional energy barrier. In this context, the linear combination of torsional terms with different $n_{\tau,i}$ leads to potentials with minima of different depths. This is necessary to describe most systems.

The description of the non-bonded (four or more bond separation) vdW+repulsion interactions ($V_{vdW+rep}$) often employs model potentials such as Lennard-Jones (Equation 2.77), Morse or Hill-Buckingham. Finally, to account for electrostatics, Force Fields generally use Coulomb's Law and assign a fixed partial charge to each atom.

$$V_{LJ}^{ij} = 4\varepsilon_{depth} \left[\left(\frac{\sigma}{R_{ij}} \right)^{12}_{rep} - \left(\frac{\sigma}{R_{ij}} \right)^{6}_{vdW} \right] \tag{2.77}$$

where ε_{depth} is the depth of the potential, R_{ij} is the distance between the interacting particles i and j, and σ is the distance at which V_{LJ}^{ij} is 0. While the computational cost of bonded interactions increases linearly with the number of atoms (N_{atom}), non-bonded interactions scale as N_{atom}^2 since they require the computation of all pairs of atoms. Due to the fast decrease of the vdW+repulsion potential with the distance, a cutoff value of around 10 Å may be employed. In this way, only the atom pairs whose internuclear distance is lower than the cutoff are considered and, hence, the computational cost decreases. On the contrary, the use of cutoff distances is not indicated for the calculation of the electrostatic potential because of the R^{-1} long range decay. In these cases, methods like the Particle Mesh Ewald summation may improve the performance.[64]

Additionally, more terms may be included in the total potential V_{total} to describe it better. In all likelihood, the most relevant addition is polarization, whose inclusion is not possible when the charges are fixed on the atoms. For this, polarizable force fields have been developed. They usually include atomic multipole parameters to model permanent electrostatics and, also, atomic induced dipole parameters to represent polarization. In this way, they achieve a better physical description of the system, specially in environments where atomic charge delocalization is relevant.[65]

Finally, reference data is required to parametrise any Force Field. This has a strong empirical character and, therefore, each Force Field has a different purpose depending on the criteria to choose the parameters and how the Force Field was developed as a whole. As a consequence, one needs to choose an appropriate Force Field for the system, based on critical evaluation of the existing literature.

Periodic Boundary Conditions

The amount of atoms that can be simulated with MM methods (up to a few millions) is still relatively small compared to the usual number of atoms in macroscopic systems. As a result, the surface/bulk molecule ratio is very large, leading to excessive surface effects in the simulation. To avoid this, Periodic Boundary Conditions (PBC) can be employed. Within this model, the system is placed in an octahedral/orthorhombic box that is replicated along the three dimensions. Therefore, the equations of motion are integrated for only one box (primary cell) and all the other replicas (virtual cells) mimic its motion. Moreover, if one molecule leaves the primary cell, another one with the same momentum appears on the opposite side to conserve the number of atoms in the cell. In this way, PBC allow the simulation of bulk properties in MM calculations with a limited number of primary molecules.[53]

Thermostats and Barostats

Chemical processes usually take place at constant temperature and pressure (NPT ensemble). Therefore, it is convenient to use thermostats and barostats to keep these variables constant.[58] The instantaneous temperature is related to the velocity of the particles and, hence, the kinetic energy of the system:

$$T_{instantaneous} = \frac{2T}{k_B N_{DOF}} \tag{2.78}$$

where k_B is the Boltzmann constant and N_{DOF} is the number of degrees of freedom. In this sense, the simplest thermostat is velocity rescaling [66], which increases or decreases the velocities after a given number of time steps according to:

$$v_{new} = \lambda v_{old} \tag{2.79}$$

where λ is defined as:

$$\lambda = \sqrt{\frac{T}{T_{instantaneous}}} \tag{2.80}$$

In this way, the kinetic energy of the system equals the kinetic energy at the target temperature T. This algorithm is very easily implemented in MD codes but, if it is frequently used in the simulation, it suppresses the fluctuations in the kinetic energy, leading to biased trajectories. Alternatively, one can use other thermostats, such as the Langevin thermostat [67], which was used in this work. In this case, the velocities of the atoms are modified by means of a friction coefficient γ, and a random force F_{random} that keeps the temperature constant and counteracts the friction. As a result, the first Hamilton equation (2.68) is now:

$$\frac{dp_i}{dt} = F_i - \gamma v_i + F_{i,random} \tag{2.81}$$

where v_i is the velocity of atom i. Therefore, the friction is more noticeable in the fastest atoms.

During the MD simulations, the pressure of the system is obtained as:[68]

$$P = \frac{1}{3V}\left(3Nk_BT + \sum_{i\neq j} R_{i,j}F_{i,j}\right) \tag{2.82}$$

Where V is the volume of the periodic box, and $R_{i,j}$ and $F_{i,j}$ are the distances and forces between atoms i and j respectively. In order to keep a constant pressure, V can be scaled by a factor η_p, which is equivalent to λ in equation 2.79. This is defined in the Berendsen barostat [69] as:

$$\eta_p = \left(1 + \frac{\Delta t}{\tau_{rel}} \xi (P - P_0)\right)^{\frac{1}{3}} \tag{2.83}$$

where τ_{rel} is the relaxation time, ξ is the compressibility and P_0 is the target pressure.

Another popular barostat is the Monte Carlo barostat.[70] Within this method, the pressure is controlled stochastically through a trial volume change that may, or may not, be accepted according to a Metropolis algorithm.

2.2.2 Gaussian Accelerated Molecular Dynamics

Most biological processes usually take place within microseconds, milliseconds or even slower. Unfortunately, this is still too slow for standard MD, where simulations longer than 1 microsecond are very computationally expensive. As a result, high energy regions are not well sampled in MD simulations and, thus, enhanced sampling techniques are required to obtain information from these areas of the PES.[71]

Generally, a predefined reaction coordinate or collective variable in which to enhance the sampling is required in most methods. However, this is not convenient when there is no a priori knowledge of the system. In this sense, Gaussian Accelerated Molecular Dynamics (GaMD) [72] add a harmonic boost potential to smooth the PES and reduce the high energy barriers (Figure 2.3), allowing the sampling of multiple low-energy conformations without the need of specifying a reaction coordinate.

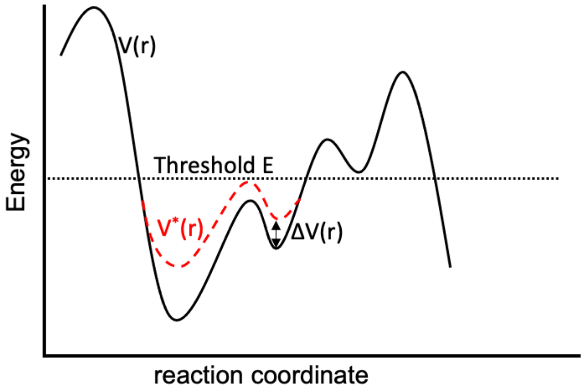

Fig. 2.3 Graphical representation of the boosted potential in GaMD. The potential remains the same above the threshold energy.

In a system whose potential energy V is lower than a threshold energy E, a boost harmonic potential (ΔV) is added, leading to a modified potential V^* according to:

$$V^* = V + \Delta V = V + \frac{1}{2}k(E - V^2), V < E \qquad (2.84)$$

where k is a force constant. On the other hand, if $V \geq E$ then the boost $\Delta V = 0$ and $V^* = V$.

Enhanced sampling techniques must not modify the relative order of the potentials. Therefore, if $V_1 < V_2$ then $V_1^* < V_2^*$. Replacing V_1^* and V_2^* by equation 2.84 gives:

$$E < \frac{1}{2}[V_1 + V_2] + \frac{1}{k} \qquad (2.85)$$

Although the relative order of V_1 and V_2 remains the same, the smooth potential leads to a lower potential difference: $V_2^* - V_1^* < V_2 - V_1$. Similarly, replacing with equation 2.84:

$$E > \frac{1}{2}[V_1 + V_2] \qquad (2.86)$$

Knowing that $V_{min} \leq V_1 < V_2 \leq V_{max}$ and combining both equations 2.85 and 2.86 then:

$$V_{max} \leq E \leq V_{max} + \frac{1}{k} \qquad (2.87)$$

where V_{max} and V_{min} are the maximum and minimum potential energies of the system. Equation 2.87 is only valid if $V_{max} \leq V_{min} + \frac{1}{k}$ and hence:

$$k \leq \frac{1}{V_{max} - V_{min}} \qquad (2.88)$$

It is useful to define k as $k = \frac{k_0}{V_{max} - V_{min}}$ where $0 < k_0 < 1$. In this context, a larger k_0 corresponds to a larger boost potential, which is more convenient to enhance the sampling. k_0 can be set to its upper bound, where $E = V_{max}$, or to its lower bound, where $E = V_{min} + \frac{1}{k}$. According to the former:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \text{x} \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right) \qquad (2.89)$$

where V_{avg} is the average potential energy. Regarding the k_0 lower bound:

$$k_0 = \min(1.0, k_0'') = \min\left(1.0, \left(1 - \frac{\sigma_0}{\sigma_V}\right) \text{x} \frac{V_{max} - V_{min}}{V_{avg} - V_{min}}\right) \qquad (2.90)$$

Finally, the standard deviation $\sigma_{\Delta V}$ of the boost potential ΔV (Equation 2.91) needs to be smaller than a predefined parameter σ_0 in order to get an accurate reweighted PES.

$$\sigma_{\Delta V} = k(E - V_{avg})\sigma_V \leq \sigma_0 \tag{2.91}$$

where σ_V is the standard deviation of the non-boosted potential V. The energetic reweighting allows to recover the original PES from the boosted potential by calculating the canonical ensemble distribution $p(A)$ as:

$$p(A_j) = p^*(A_j)\frac{\langle e^{\beta \Delta V(r)} \rangle_j}{\sum_{i=1}^{M} \langle p^*(A_i) e^{\beta \Delta V(r)} \rangle_i}, j = 1, \cdots, M \tag{2.92}$$

where M is the number of bins, $\beta = k_B T$ and $\langle e^{\beta \Delta V(r)} \rangle_j$ is the Boltzmann factor of the ensemble average of the boost potential ΔV for the frames that are in the jth bin. It is possible to approximate the Boltzmann factor by using a cumulant expansion. In GaMD simulations, ΔV follows a near gaussian distribution and, therefore, cumulant expansion to the second order is enough to obtain an accurate approximation. In this sense, a low anharmonicity of the ΔV distribution indicates an accurate reweighting of the PES.

GaMD algorithms take σ_0 as input and all other parameters (V_{max}, V_{min}, V_{avg} and σ_V) are calculated automatically following 5 steps. The first one is a conventional MD (cMD) calculation in which the system equilibrates and no statistics are collected. The second one is also cMD, but initial boost parameters are obtained to activate the GaMD algorithm. The third one corresponds to a GaMD equilibration in which the parameters are kept fixed. In the fourth, the parameters are updated until k_0 reaches a maximum value. Finally, the fifth step uses the previously converged parameters to run the GaMD production.

2.2.3 Binding Free Energy Calculations

The probability of a system to remain in a given state can be obtained from the free energy. Unfortunately, it is impossible to calculate the absolute free energy in large systems and, hence, obtaining relative free energies becomes more convenient:[71]

$$\Delta G_{BA} = -\beta^{-1} \ln\left(\frac{P_B}{P_A}\right) \tag{2.93}$$

where P_A and P_B are the probabilities of sates A and B. In this context, molecular docking [73] is a very popular technique in drug design that is based on the calculation of ligand binding free energies to find its binding pockets. It uses a stochastic algorithm that samples the space and a scoring function to calculate the binding energy of each of the configurations that it generates. Moreover, the receptor is usually taken to be rigid and, thus, it is very computationally efficient. However, the results are often far from accurate.

In order to improve the results, the Molecular Mechanics Generalized Born Surface Area (MMGBSA) [74] method may be employed. This method calculates the relative free energy between two solvated states from a given MD trajectory of the complex. In order to get rid of the fluctuations arising from the solvent-solvent interaction, that are usually an order of magnitude larger than the binding energy, the MMGBSA method replaces the explicit solvent by an implicit one that is calculated according to the Generalized Born implicit solvent model. In its most common approach, the binding free energy is estimated as:

$$\Delta G_{bind} = \langle \Delta G_{el} \rangle_C + \langle \Delta G_{vdW} \rangle_C + \Delta G_{pol} + \Delta G_{np} \tag{2.94}$$

where $\langle \Delta G_{el} \rangle_C$ and $\langle \Delta G_{vdW} \rangle_C$ are the ensemble average of the electrostatic and van der Waals free energies, which are calculated from the MM force field. ΔG_{pol} is the polar solvation term that includes the electrostatic interactions between the complex and the implicit solvent. And finally, ΔG_{np} accounts for the complex-solvent interactions that are not included in ΔG_{pol} (dispersion, repulsion, etc).

2.3 Hybrid Quantum/Classical Methods

2.3.1 Quantum Mechanics/Molecular Mechanics

The calculation of large systems and their environment is possible by means of Molecular Mechanics (MM). However, it leads to a significant loss in accuracy in the regions of interest as it cannot describe quantum effects such as bond breaking or photophysical processes. In order to deal with this problem, Quantum Mechanics/Molecular Mechanics (QM/MM) techniques divide the system in two regions: an inner region, that is treated quantumly (QM region), and an outer region, that is described at the MM level (MM region). The QM region contains the solute of interest, whereas the MM region calculates the environment.[75]

The main challenge in QM/MM is the modelling of the interaction between both regions. The most simple strategy is the subtractive scheme, in which the energy is calculated as:

$$E^{sub}_{QM/MM} = E_{QM}(I) + E_{MM}(I+O) - E_{MM}(I) \tag{2.95}$$

where the subscripts indicate the level of theory, and I and O refer to the inner and outer regions respectively. In this way, the complete system is calculated at the MM level, while the inner region is computed at both MM and QM. Then, the inner MM calculation is subtracted to avoid double counting. As a result, the description of the interaction between the QM and the MM regions is classical. Moreover, in principle, the inner region is not polarized by the outer region. However, most methods nowadays include polarization by including the outer charges in the QM Hamiltonian. Alternatively, the energy in the additive scheme is calculated as:

$$E^{ad}_{QM/MM} = E_{QM}(I) + E_{MM}(O) + E_{QM/MM}(I/O) \tag{2.96}$$

where $E_{QM/MM}(I/O)$ is the interaction energy between both regions and it is defined as:

$$E_{QM/MM}(I/O) = E^{bonded}_{MM}(I/O) + E^{vdW}_{MM}(I/O) + E^{el}(I/O) \tag{2.97}$$

The van der Waals ($E^{vdW}_{MM}(I/O)$) and bonded ($E^{bonded}_{MM}(I/O)$) terms are computed classically. Regarding the electrostatic term $E^{el}(I/O)$, there exist several schemes. The most straightforward way is the mechanical embedding. Within this approach, the electrostatic interaction is computed classically and, hence, the QM region is not polarized by the MM region. In order to include this polarisation, the electrostatic embedding scheme uses an effective hamiltonian that includes the charges of the MM region. In this way, the QM wave function is polarized by the point charges of the outer region. This model yields reasonable results but it still does not account for the polarisation of the MM region by the QM region. This is done by the polarizable embedding, in which the mutual polarization between both regions is calculated self-consistently. Therefore, it also requires a polarizable force field to describe the MM region. This makes the polarizable embedding scheme very computationally expensive and, thus, even though it is the most accurate approach, electrostatic embedding is more commonly used.[75]

No matter the partition scheme, the QM region may suffer from overpolarisation due to the lack of Pauli repulsion from the MM region. As a result, there is an artificial attraction between both regions that can be avoided by using screening terms. Furthermore, QM/MM methods generally struggle with the description of the I/O interface, specially when it involves covalent bonds. In these cases, the link atom scheme adds boundary atoms to each of the regions to prevent the presence of radicals in the QM region. Alternatively, in the localised-orbital scheme, the QM orbitals are frozen, keeping both electrons in the inner region.[75]

2.3.2 Polarizable Continuum Model

Similarly to QM/MM, the polarizable continuum model (PCM) approach also simulates the effect of the environment in a classical way. However, instead of including the solvent molecules explicitly, PCM introduces a dielectric continuum to reproduce its effect implicitly.[37]

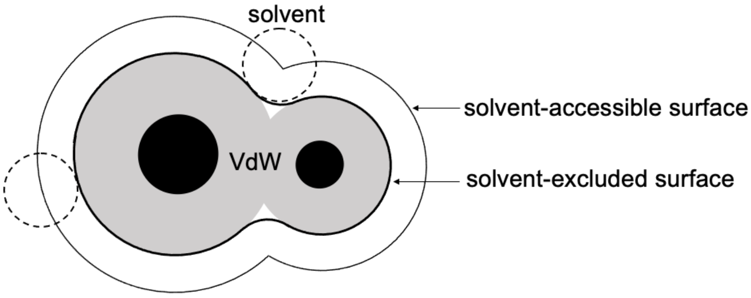

Fig. 2.4 Graphical representation of the cavity surfaces in PCM. The solvent rolls over the solvent-excluded surface to generate the solvent-accesible surface.

The solvation free energy is defined as:

$$\Delta G_{sol} = G_{sol} - G_{vacuum} = \Delta G_{ele} + \Delta G_{cav} + \Delta G_{dis} + \Delta G_{rep} \tag{2.98}$$

where ΔG_{cav} is the energy that results from the rearrangement of the molecules when a solute is introduced, and ΔG_{ele}, ΔG_{disp} and ΔG_{rep} are the electrostatic, dispersion and repulsion contributions. In this context, PCM defines a cavity around the solute, the solvent accesible surface (SAS), whose area is created by "rolling" a solvent molecule over the solvent-excluded surface generated by the van der Waals radii (Figure 2.4). Within this model, ΔG_{cav}, ΔG_{disp} and ΔG_{rep} are approximated empirically, as they are proportional to the SAS. Regarding the electrostatic interactions, they mainly contribute to the interaction potential (V_{int}) of the effective Hamiltonian:

$$G_{sol} = \langle \Psi_{sol} | \hat{H}_0 + \frac{1}{2} V_{int} | \Psi_{sol} \rangle \tag{2.99}$$

and

$$V_{int}(r) = \int \frac{\sigma(r_s)}{|r - r_s|} dr_s \tag{2.100}$$

where $\sigma(r_s)$ are the charges that are distributed along the SAS and generate the continuum, and $|r - r_s|$ accounts for the distance of the solute to the discrete values of $\sigma(r_s)$. Mutual polarization appears because $\sigma(r_s)$ depends also on V_{int}. As a result, both $\sigma(r_s)$ and V_{int} need to be calculated self-consistently according to the Self Consistent Reaction Field (SCRF), and the wave function is optimised according to the effect of the solvent. Finally, since the $\sigma(r_s)$ of the SAS are related to the continuum dielectric constant, different solvents can be modelled by changing this parameter.[37]

When modelling excited states, we can distinguish between two responses that result from the continuum-solute mutual polarization. The first one, which is known as the dynamical response, corresponds to the fast equilibration of the electrons of the solvent due to the changes in the electronic density of the solute upon excitation. The second one, the inertial response, refers to the slower equilibration of the nuclei of the solvent. The absorption spectrum of the molecule is calculated assuming a vertical excitation, i.e., the atoms in the solute remain in the same position as in the ground state and there is only dynamical response from the solvent (non-equilibrium solvation). In a first step, the energies of the excited states are calculated using the linear response approach, in which a correction based on the transition density is added to consider the dynamical response. Then, the energy can be corrected further using the sate specific approach, that is based on the electronic density difference between the ground state and an specific excited state. Conversely, optimization processes present larger relaxation times and, hence, the atomic nuclei of the solvent have time to equilibrate (inertial response). In these cases, equilibrium solvation calculations are indicated.[76]

Computational Details

3.1 Classical Simulation Details

The initial structures were generated with CHARMM-GUI[77], and the trajectories were obtained with AMBER 20[78] and Tinker-HP[79]. Moreover, preliminar binding pockets were found with AutoDock Vina[80]. All trajectories were analysed using cpptraj[81], including the clustering analysis, the root mean square deviation (RMSD), and the calculation of distances. VMD[82] and PyMOL[83] were used to visualize the trajectories and make the figures in this work. More specific details can be found in the following subsections.

3.1.1 Initial structures

The crystallized structure of the Na$_V$1.5 channel [24], PDB ID 7FBS, was taken from the Protein Data Bank (PDB).[84] Two models, the whole protein and a truncated pore model of the Na$_V$1.5 channel, were selected to run the simulations. The truncated pore model consists of the S5 and S6 α helixes of domains I-IV, whereas the whole channel model includes all α helixes (S1-S6) of domains I-IV. Both proteins were embedded inside a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer using the CHARMM-GUI membrane builder [77], and 4 disulfur bonds, between cysteines 281-336, 909-918, 1365-1386 and 1730-1744, were included as specified in the original pdb file. The structure was then solvated in aqueous 0.15M NaCl solution (physiological concentration) in a rectangular periodic box. The whole protein model includes 253 and 250 POPC molecules in the upper and lower layers, respectively, resulting in a total of 261464 atoms in the system: 67402 in the membrane, 18587 in the protein and 175475 in the aqueous solution. On the other hand, the truncated pore model only includes 100 and 105 POPC molecules in the bilayer, which are a total of 111566 atoms: 27470 in the membrane, 9684 in the protein and 74412 in the aqueous solution. Likewise, a solvated 0.15 M NaCl POPC lipid bilayer with 111 lipid per layer was built using the CHARMM-GUI membrane builder without embedding any protein.

The initial diprotonated *trans*-DAD geometry (DADH_2^{2+}) was built using GaussView [85], which was later optimized at the M06-2X[86]/cc-pVDZ[55] level of theory, and in implicit water solvation according to the implementation of IEFPCM[87] in Gaussian16[88]. The restrained electrostatic potential (RESP) charges were calculated at the Hartree-Fock/6-31G* level of theory with Gaussian16. For the non-polarizable force field simulations, the topology and the coordinate files were generated with Antechamber [89], parmchk2 and tleap as implemented in AmberTools20[78] to include the parameters of DAD on top of the initial files created by CHARMM-GUI. The Amber FF19SB [90], TIP3P[91], GAFF2[92] and Lipid21[93] force fields were used for the protein, water, ligand and lipids, respectively. For the polarizable force field simulations, diprotonated DAD was parametrised, using poltype2[94], to get its AMOEBA[95] parameters. Then, the topology and the coordinate files were created following the same procedure as in [96].

3.1.2 Docking

The ligand pdbqt file was created using Obabel[97] and then, the azo torsional angle was locked to study the binding pockets of *trans*-DADH_2^{2+} using AutoDock vina[80]. The receptor in the docking calculation was the truncated Na_V1.5 channel, whose pdbqt file was previously obtained with AutoDock. The location of the binding pockets were restrained to a box of 40Å x 40Å x 30Å centered on the pore of the channel. Finally, an exhaustiveness = 8 was specified for the calculation.

3.1.3 Molecular Dynamics Details

Non-polarizable Force Field Calculations

All MD non-polarizable calculations were performed using AMBER 20 [78]. In all cases, the multistep procedure recommended by CHARMM-GUI was used for initial relaxation. First, the system was minimized for 2500 steps using the steepest descent algorithm, followed by 2500 steps using the conjugate gradient method, which is more efficient close to the minimum. Then, the system was progressively heated from 0 K to 303.15 K in the NVT ensemble with a 1 fs time-step, and using the Langevin thermostat [67] with 2.0 ps^{-1} collision frequency. Finally, a 2 fs time-step production was performed in the NPT ensemble. Once again the temperature was controlled with the Langevin thermostat (1.0 ps^{-1}) at 303.15 K. On the other hand, the pressure was controlled with the Berendsen barostat [69] at 1.0 bar. In both the heating and the production, the electrostatic interactions were calculated using the particle-mesh Ewald method with a grid spacing of 1.0 Å, and the cutoff for the non-bonded interactions was set to 9 Å. Moreover, the bonds involving hydrogen atoms were constrained using SHAKE[61] and the Leap-frog integrator[60] was used. Finally, the dihedrals involved in the C-C-N=N-C-C of $DADH_2^{2+}$ were constrained according to [98] to keep $DADH_2^{2+}$ in the *trans* conformation. The length of the production depended on the simulated system and will be indicated in the results section.

Polarizable force field calculations

A POPC membrane was simulated with the AMOEBA force field[95] using the Tinker-HP [79] package. The simulation started from the equilibrated POPC membrane calculated with Amber. First, the system was minimized for 2000 steps using the steepest-descent algorithm. Then, it was heated in the NVT ensemble using the Bussi thermostat[99] at 303.15 K and the RESPA integrator[63] for 50 ps (2 fs time-step). Finally, a 60 ns (2 fs time-step) NPT production was run using the particle-mesh Ewald method, a non-bonded cutoff of 9 Å, the OPT4 polarization technique[100], the Bussi thermostat at 303.15K, the Monte Carlo barostat[70] to keep the pressure at 1 atm, and the RESPA integrator.

3.1.4 Gaussian Accelerated Molecular Dynamics details

Starting from the most favourable docking pose, a 100 ns conventional MD (cMD) simulation was performed. Then, three trial dual-GaMD simulations, where both the total and dihedral potentials are boosted, were calculated to determine the optimal σ_0 values (3, 4 or 5). A total of 1000000, 5000000, 2500000 and 20000000 steps were calculated for the first (cMD), second (preliminary collection of statistics), third (GaMD equilibration with fixed parameters) and fourth (GaMD parameters update) GaMD phases, respectively. Next, with the selected parameters, 4 independent 1 μs GaMD production GaMD trajectories (fifth GaMD phase), differing on their initial velocities, were simulated starting from the last frame of the previous 100 ns cMD production. In all cases, the same general specifications as in the non-polarizable MD calculations were used. Finally, the energy reweighting was performed using PyReweighting-2D.[101] To do this, the RMSD of the ligand and the distance between the N in the amine located on the side on the amide and the centre of mass of the DEKA aminoacids, were selected as the reaction coordinates.

3.1.5 Molecular Mechanics Generalized Born Surface Area details

After identifying the binding pockets from the GaMD simulation, 100 ns of cMD were calculated to explore their stability and compute their relative free energy with MMGBSA.

The analysis was based on 1000 cMD frames. Furthermore, an energy decomposition analysis was performed by taking into account the residues within 7Å from the ligand that are present in at least 2% of the frames.

3.2 Quantum Mechanics calculation details

3.2.1 Dynamic Absorption spectra

The script main_qminputs.py from MoBioTools [102] was used to extract 100 snapshots from the cMD simulations that were performed for the MMGBSA analysis of the binding pockets, and to generate the Gaussian16[88] QM/MM inputs to compute the absorption spectrum. The QM region included *trans*-DADH$_2^{2+}$, and the first 5 excited states were computed at the M06-2X/cc-pVDZ level of theory. The MM region consisted of the point charges of the rest of the system and, hence, the QM/MM calculation followed an electrostatic embedding scheme. The analysis of the excited states was performed using TheoDORE.[56] The same procedure was repeated for a 100 ns cMD of DADH$_2^{2+}$ in water. From this simulation, single point calculations of 100 snapshots were calculated using QM/MM, and then, the explicit water molecules were removed from the frames to perform QM/PCM calculations in the same configurations. To plot the spectra, a convolution of gaussians (σ=0.3) was performed. The height of the gaussians corresponds to the oscillator strength of the excitations.

3.2.2 Static Absorption spectra

Single point TDDFT calculations of deprotonated *trans*-DAD (DAD), DADH$_2^{2+}$ and azobenzene in both gas phase and solvent (non-equilibrium IEFPCM), water for the DAD species and dichoromethane (DCM) for azobenzene, including the first 5 excited states, were performed at the B3LYP[47]/cc-pVDZ, CAM-B3LYP[49]/cc-pVDZ, ωB97XD[50]/cc-pVDZ and M06-2X[86]/cc-pVDZ levels of theory in Gaussian16. In both gas phase and solvent, the geometries were previously optimized with their corresponding levels of theory (minima in their PES but different geometries). In addition, the same calculations were repeated for DAD and DADH$_2^{2+}$ with geometries optimized at gas M06-2X/cc-pVDZ (common geometry). Then, the SA(5)-CASSCF absorption spectrum was computed in *trans*-DAD, *trans*-DADH$_2^{2+}$ with a 14 electrons, 13 orbitals active space$\pi\pi^*$ (14,13 AS), and azobenzene (12,11 AS) in gas phase at their gas M06-2X/cc-pVDZ geometries using OpenMolcas.[103] The CASPT2 energies were then calculated with an imaginary shift of 0.3 and with and without IPEA shift. Finally, the DAD and DADH$_2^{2+}$ geometries were optimized at the SA(3)-CASSCF level of theory with an 8,7 AS, and the CASPT2 energies were computed under the same conditions as before. The molecular orbitals were visualized using gmolden.[104] Figure 3.1 shows the orbitals in the AS.

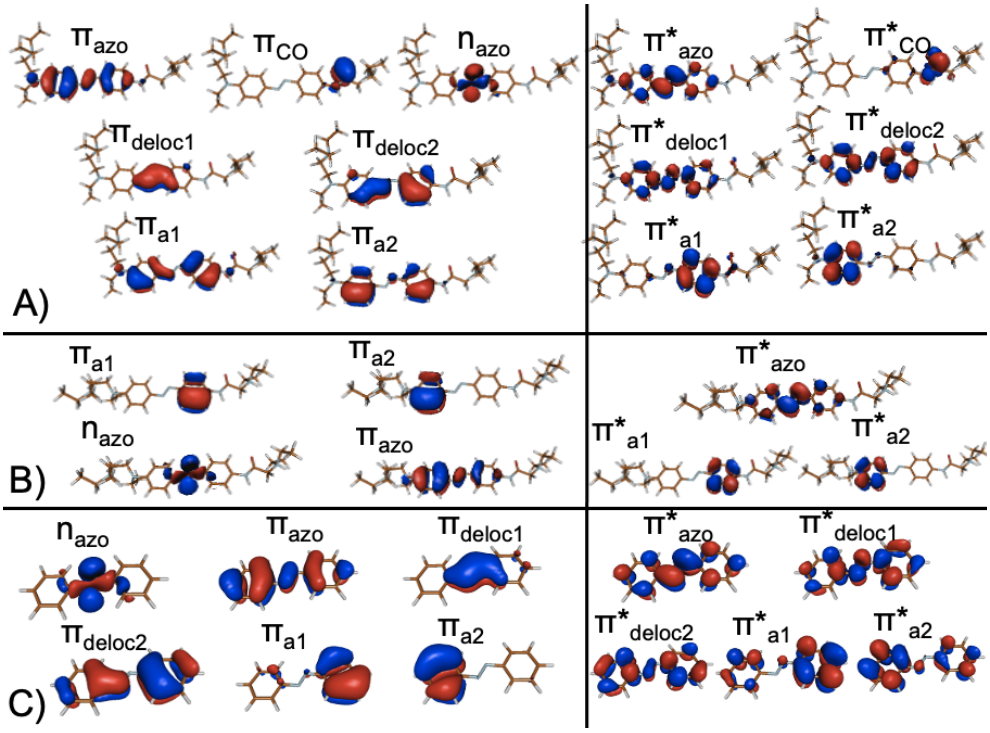
π_{azo}
π_{CO}
n_{azo}
π^*_{azo}
π^*_{CO}
π_{deloc1}
π_{deloc2}
π^*_{deloc1}
π^*_{deloc2}
π_{a1}
π_{a2}
π^*_{a1}
π^*_{a2}
A)
π_{a1}
π_{a2}
π^*_{azo}
n_{azo}
π_{azo}
π^*_{a1}
π^*_{a2}
B)
n_{azo}
π_{azo}
π_{deloc1}
π^*_{azo}
π^*_{deloc1}
π_{deloc2}
π_{a1}
π_{a2}
π^*_{deloc2}
π^*_{a1}
π^*_{a2}
C)

Fig. 3.1 DAD molecular orbitals (A) in the CASSCF active space (geometries at M06-2X) and (B) in the CASSCF optimization. (C) Active space in azobenzene. The orbitals in DAD and $DADH_2^{2+}$ are very similar orbitals so only one set is presented.

3.2.3 Photoisomerization PES

The *trans-cis* photoisomerization PES of azobenzene, DAD and $DADH_2^{2+}$ in gas phase, implicit water (DAD species) and DCM (azobenzene), were calculated with TDDFT (M06-2X/cc-pVDZ) in Gaussian16. First, a conventional geometry optimization of the bright excited state (S_2) was performed until a minimum point was found or a crossing point with the S_1 was reached. In the case of degeneracy with the S_1, the geometry optimization of the S_1 state was calculated from the S_2/S_1 structure. When the S_1 minimization lead to a crossing point with the S0, the *trans-cis* transition state (TS) structure was calculated in the ground state. Then, an Intrinsic Reaction Coordinate (IRC) calculation was performed to connect the TS with the *cis* and *trans* conformations. In the cases where the minimization of the excited states lead to a stationary point, SHARC [105] was used to optimize the S_2/S_1 or S_1/S_0 intersections. Equilibrium solvation was used in the PCM excited states optimizations.

Results

In this section, the Na$_V$1.5 binding pockets of DADH$_2^{2+}$ will be presented, followed by the absorption spectra of both DAD species and their photoisomerization pathways.

4.1 Binding pocket search

4.1.1 Na$_V$1.5 channel equilibration

The Na$_V$1.5 channel has 18587 atoms, requiring many computational resources even at the MD level of theory. A common strategy to improve the performance in ion channel studies is focusing on the pore (S5-S6 in Figure 1.8), and excluding the VSs (Voltage Sensing domains S1-S4 in Figure 1.8) from the calculation.[106] This approximation is only valid if the VSs do not substantially change the structure of the pore. In order to ensure that this is also the case in the Na$_V$1.5 channel, two 500 ns MD simulations of the protein embedded in a solvated POPC membrane were performed. In one of them, the protein only consists of the pore, whereas the other calculation also includes the VSs (whole channel simulation) as shown in Figure 4.1.

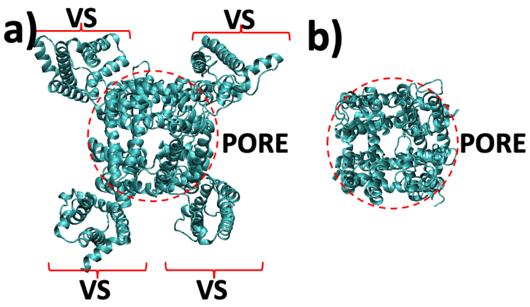

Fig. 4.1 a) Top view of the protein in the simulation containing the whole Na$_V$1.5 channel (pore + VSs). b) Top view of the protein in the simulation containing only the Na$_V$1.5 pore.

The energy, the temperature, the pressure and the density remain stable throughout both MD trajectories. What takes more time to converge is the RMSD (Figure 4.2), which measures the internal motion of the pore relative to its initial position in both simulations.

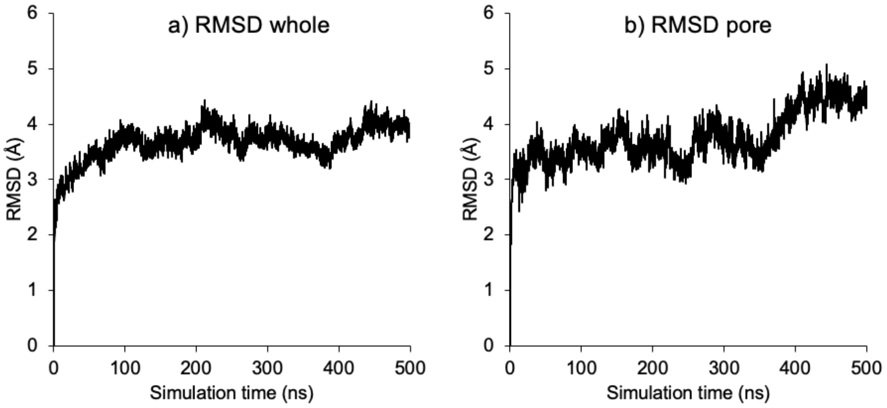

Fig. 4.2 a) RMSD of the pore in the simulation containing the whole $Na_V1.5$ channel (pore + VSs). b) RMSD of the pore in the simulation containing only the $Na_V1.5$ pore.

The RMSD in Figure 4.2b did not converge in the 500 ns simulation, maybe due to the absence of the VSs. In any case, ∼1.5Å oscillations are not usually relevant so a clustering analysis of the last 100 ns is performed to find representative structures of the protein in both simulations. The RMSD between the common atoms of the representative structures is 3.91 Å. It is of the same order of magnitude as the RMSD with respect to the representative structures (Figure 4.3), indicating that simulations containing only the pore are good enough.

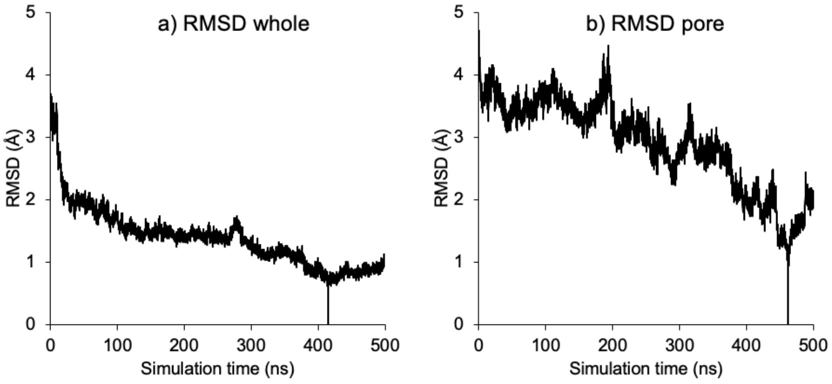

Fig. 4.3 a) RMSD of the pore in the simulation containing the whole $Na_V1.5$ channel. b) RMSD of the pore in the simulation containing only the $Na_V1.5$ pore. Both are relative to their representative structures (the vertical lines indicate their positions in the trajectories).

Finally a longer trajectory (1200 ns) of the channel containing only the pore is calculated to equilibrate the system further. According to Figure 4.4, a well equilibrated structure is achieved after 400 ns of simulation. In the same way as before, a clustering analysis of the converged frames is performed to find a representative structure of the pore. The resulting representative structure will be taken as the docking receptor and as the starting point in GaMD.

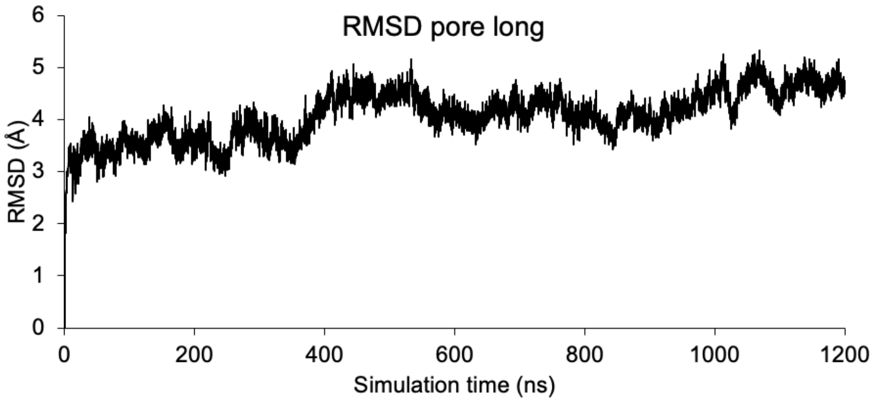

Fig. 4.4 RMSD of the pore in the 1200 ns simulation containing only the pore relative to the first frame.

4.1.2 Docking: preliminary binding pockets of DAD

Table 4.1 shows the affinities of the docking binding modes of $DADH_2^{2+}$ (occuring naturally at pH=7.4) in the $Na_V1.5$ representative structure, and the RMSD of the ligand in the different pockets relative to the most favourable pose. The resulting modes are shown in Figure 4.5.

Mode	**Affinity (kcal/mol)**	**RMSD (Å)**
1	-8.71	0
2	-8.47	1.99
3	-8.12	1.96
4	-7.87	6.89
5	-7.61	7.78
6	-7.54	9.07
7	-7.47	8.00
8	-7.28	8.29
9	-7.15	6.34

Table 4.1 Affinities and RMSD of the binding modes of $DADH_2^{2+}$ according to docking.

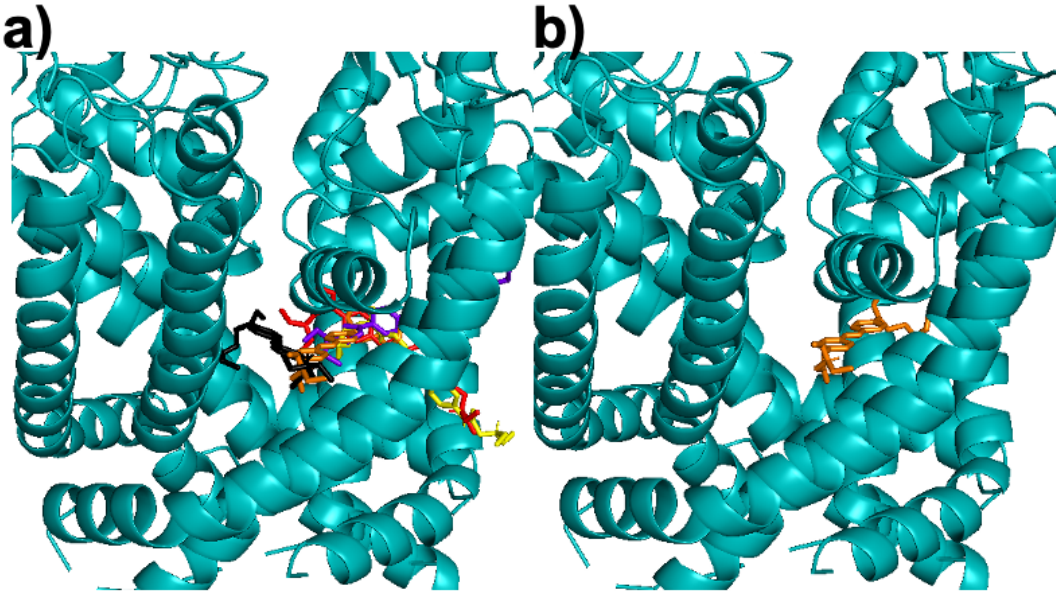

Fig. 4.5 a) Binding modes of $DADH_2^{2+}$ in the $Na_V1.5$ channel according to docking. Modes 1, 4, 5, 6 and 9 are orange, yellow, purple, red and black, respectively. Modes 2, 3, 7 and 8 were excluded because they resemble either 1, 5 or 6. b) Most favourable pose (mode 1) according to docking.

As it can be seen in Table 4.1, the RMSD of some modes are very similar, indicating that they are different poses belonging to the same binding pocket. This is the case of modes 1, 2 and 3; 5 and 7; and 6 and 8. Finally, mode 1 (Figure 4.5b) will correspond to the initial structure in the GaMD simulation. In this way, the enhanced sampling starts from a conformation that is potentially stable.

4.1.3 GaMD: binding pockets of DAD

As explained in the Theoretical Methods section, k_0 needs to be 1 in order to work with the largest possible boost potential (ΔV), which is the most convenient for the enhanced sampling. At the same time, the standard deviation of ΔV ($\sigma_{\Delta V}$) needs to be smaller than a predefined parameter σ_0 according to Equation 2.91. In order to ensure an accurate reweighting, $\sigma_{\Delta V}$, and hence, σ_0, should be as small as possible. Therefore, the strategy is to select the smallest σ_0 that allows $k_0 = 1$.

In this sense, Table 4.2 shows the $\sigma_{\Delta V}$ and k_0 values for both the potential and dihedral boosts of three trial GaMD simulations differing on σ_0. From these results, the most appropriate σ_{0P} and σ_{0D} for the production are 4 and 3, respectively. Note that trial simulations with $\sigma_0 < 3$ make no sense because in all cases $\sigma_{\Delta V} > 2$.

	Potential boost		**Dihedral boost**	
σ_0	$\sigma_{\Delta V}$	k_0	$\sigma_{\Delta V}$	k_0
3	3.45	0.61	2.63	1
4	2.51	1	2.54	1
5	2.49	1	2.54	1

Table 4.2 Selection of σ_{0P} and σ_{0D} for the GaMD production.

Four 1μs GaMD production simulations, differing on the initial velocities, were then performed starting from the most favourable docking pose (Figure 4.5b). Although specifying collective variables is not necessary for the GaMD production, they are required for the energy reweighting. By inspecting the resulting GaMD trajectories, the side of DADH_2^{2+} containing the amide seems to get closer to the DEKA filter in some of the configurations. Hence, a convenient reaction coordinate is the distance between the DEKA centre of mass and the amine next to the amide (Figure 4.6B). Moreover, measuring the RMSD of DADH_2^{2+} would also help in the identification of potential binding pockets, since configurations belonging to the same cluster have similar RMSD values. This results in the reweighted energy map in Figure 4.6A, where the 4μs are combined. Furthermore, the anharmonicity of the ΔV distribution is 0.0184 (near gaussian distribution), indicating that the reweighting is sufficiently accurate.

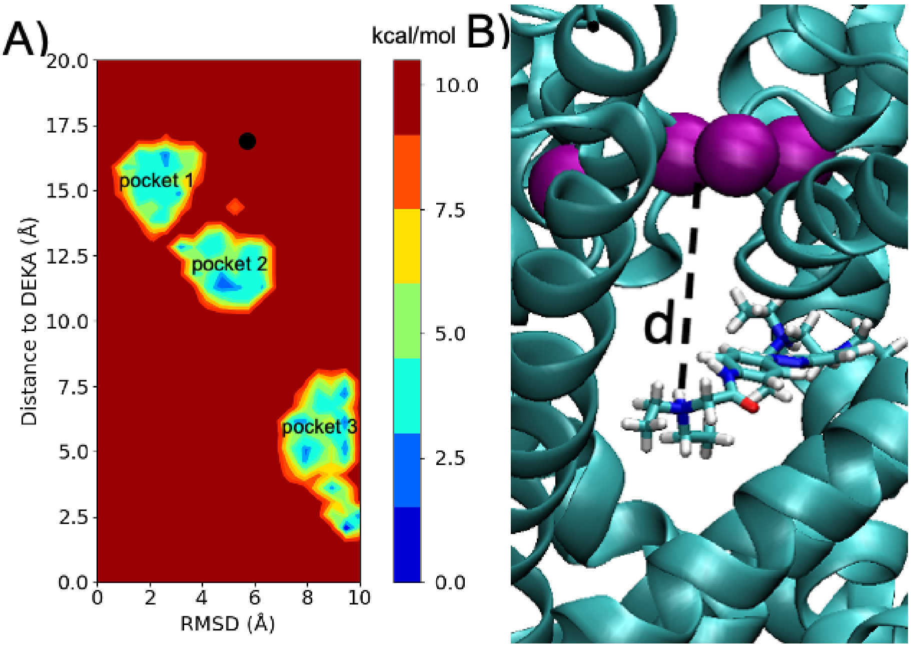

Fig. 4.6 A) Reweighted 2D energy map resulting from the four 1μs GaMD production simulations. The black dot indicates the position of the most favourable docking pose. B) Distance to DEKA reaction coordinate. The purple balls represent the DEKA aminoacids.

Based on Figure 4.6A, the most favourable docking pose is close to pocket 1, but it is not an actual binding site. Therefore, docking can only serve as an initial estimation. Moreover, Figure 4.6A suggests that there might be more than one minima in pocket 3. In order to clear this out, a clustering analysis, requesting 4 representative structures of the GaMD production, is performed. Reordering the snapshots in the trajectory according to their corresponding cluster results in the following RMSD graph:

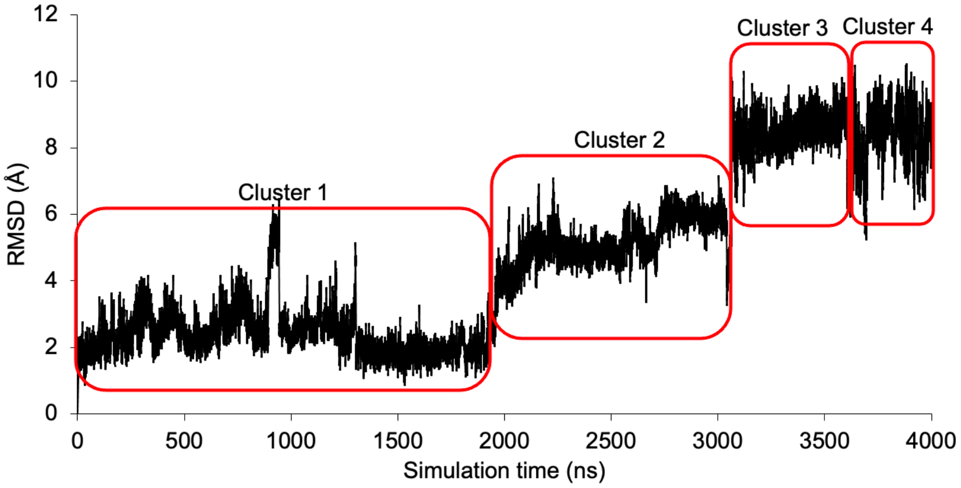

Fig. 4.7 RMSD of the GaMD trajectory reordered according to the identified clusters.

Clusters 1 and 2 correspond to binding pockets 1 and 2 in Figure 4.7, respectively. Likewise, clusters 3 and 4 correspond to pocket 3. Figure 4.8 shows the representative structures. The resemblance in the RMSD and in the structures of clusters 3 and 4 indicate that they are actually different conformations belonging to the same binding pose: pocket 3.

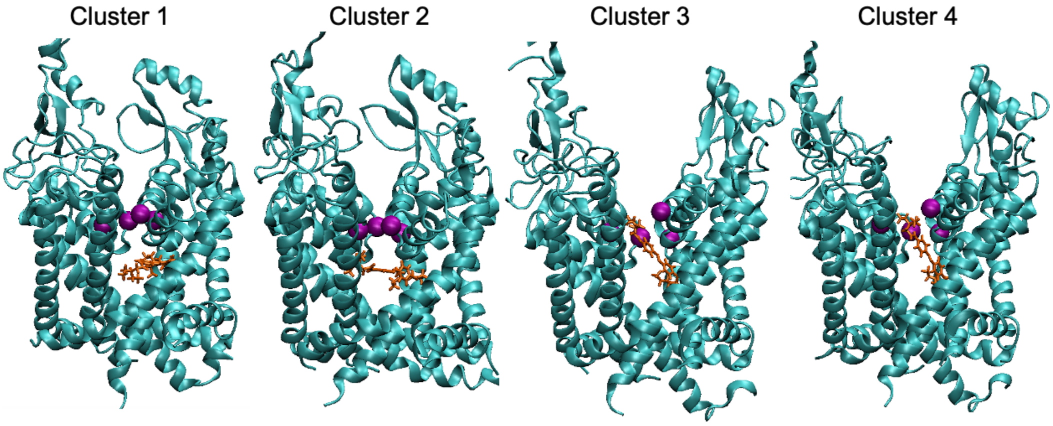

Fig. 4.8 Clusters 1, 2, 3 and 4 resulting from the clustering analysis of the GaMD trajectory. The purple balls indicate the position of the DEKA aminoacids and $DADH_2^{2+}$ is in orange.

Starting from the representative structures of clusters 1, 2 and 4, 100 ns conventional MD simulations were performed to determine the relative stability of the binding pockets. $DADH_2^{2+}$ in pocket 2 is not stable, indicating that it is not an actual binding pose. Conversely, $DADH_2^{2+}$ remains in pockets 1 and 3 during the 100 ns of their corresponding simulations, so they seem to be stable. MMGBSA analysis of 1000 frames of the trajectories estimate a free energy of -56.95 kcal/mol and -36.56 kcal/mol for pockets 1 and 3, respectively. Therefore, pocket 1 is the most stable binding pose of $DADH_2^{2+}$ in the $Na_V1.5$ channel. However, pocket 3 is also an interesting mode due to its closeness to the DEKA filter, as it may block ion conduction. In this sense, Table 4.3 and Figure 4.9 show the residues that interact the most with $DADH_2^{2+}$ according to the MMGBSA decomposition analysis of both pockets.

	Residue	ΔG_{vdW}	ΔG_{el}	ΔG_{pol}	ΔG_{nonpol}	ΔG_{tot}
Pocket 1	ASN297	-2.01	-6.57	3.80	-1.29	-6.07
	PHE460	-2.54	-0.39	-0.67	-1.82	-5.42
Pocket 3	THR412	-2.44	-3.97	3.41	-1.63	-4.63
	PHE413	-2.45	-1.42	1.74	-1.74	-3.87
	SER545	-1.39	-4.04	2.86	-1.14	-3.71

Table 4.3 Total energies (kcal/mol) of the most interacting residues in pockets 1 and 3. Electrostatic, vdW, polarization and non-polarization energy contributions are also shown.

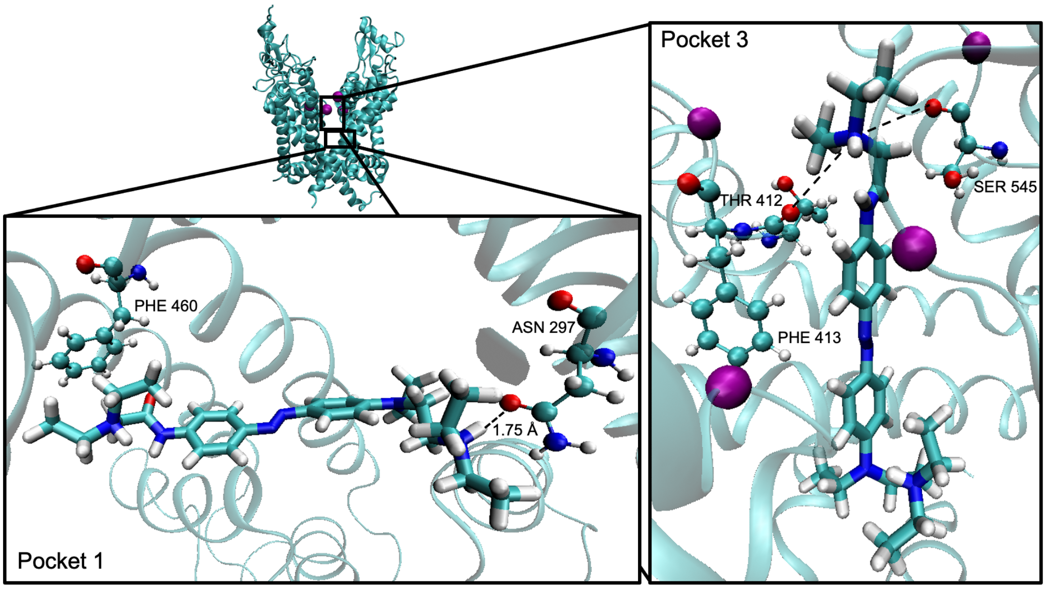

Fig. 4.9 Residues that interact the most with $DADH_2^{2+}$ in pockets 1 and 3 according to the MMGBSA analysis. The purple dots indicate the position of the DEKA aminoacids. Dashed lines indicate electrostatic interactions between a proton of $DADH_2^{2+}$ and an oxygen of an aminoacid.

The major contributions to stabilisation in both pockets are of electrostatic nature (ΔG_{el}). The protonated amino groups in $DADH_2^{2+}$ interact with the partial negatively charged O in ASN297 in pocket 1, and in THR412 and SER545 in pocket 3. This indicates that the protonation state of DAD is crucial in the stabilisation of the binding pockets. Therefore, the most favourable poses of deprotonated DAD may differ from the binding modes described here. It is particularly important to take this into account because, as it was mentioned in the introduction, only deprotonated DAD diffuses rapidly across the cell membrane and enters in the ion channel. However, DAD will most likely protonate again inside the pore of $Na_V1.5$, which is why the present work focuses on $DADH_2^{2+}$. Nevertheless, it would be interesting to repeat the same study with deprotonated DAD.

Further stabilisation in the binding pockets is achieved thanks to the van der Waals interactions (ΔG_{vdW}) arising from the benzene rings in PHE460 and PHE413 in pockets 1 and 3, respectively. Moreover, minor contributions from aminoacids that are not present in Table 4.3 also help in the stabilisation of the binding pockets. Regarding the polar (ΔG_{pol}) and non-polar (ΔG_{nonpol}) solvation contributions, it seems that ΔG_{pol} is more unfavourable when the ΔG_{el} contribution is higher. On the other hand, ΔG_{nonpol} is always of the same order of magnitude. Finally, although $DADH_2^{2+}$ in pocket 3 is very close to the DEKA aminoacids, they do not contribute strongly to its stabilisation. Nonetheless, the presence of $DADH_2^{2+}$ may still hinder the interaction between DEKA and the Na^+ ions.

4.2 Absorption spectrum of DAD

Table 4.4 shows the $n\pi^*$ and $\pi\pi^*$excitation energies of deprotonated (DAD) and diprotonated ($DADH_2^{2+}$) in gas phase and in implicit water, using different DFT functionals. In all cases, the involved orbitals are very similar so only one set is shown in Figure 4.10 as an example.

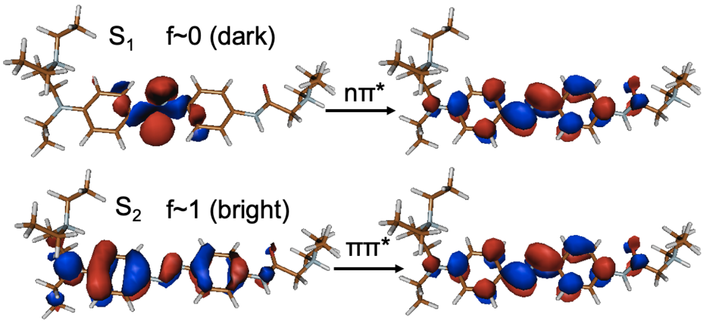

Fig. 4.10 Molecular orbitals involved in the S_1 $n\pi^*$ (up) and the S_2 $\pi\pi^*$ (down) excitations in $DADH_2^{2+}$ in water at the M06-2X/cc-pVDZ level of theory.

M06-2X is a hybrid meta-GGA functional, while B3LYP, CAM-B3LYP and ωB97XD are hybrid GGA. In addition, CAM-B3LYP and ωB97XD include long-range corrections, which is probably why they provide almost the same results in all the considered scenarios. Regarding M06-2X, it estimates similar $\pi\pi^*$ energies to CAM-B3LYP and ωB97XD, whereas $n\pi^*$ is closer to B3LYP. Upon comparison with the experimental $\pi\pi^*$ excitation, it seems that B3LYP provides the most accurate results. However, it is well known [49] that B3LYP fails in the calculation of electronic excited states. Moreover, a B3LYP calculation including vibrational motion, which would be more realistic than the single point calculations at the optimized geometry, would probably lead to energies lower than the experimental value. On the other hand, based on the experimental value, M06-2X, CAM-B3LYP and ωB97XD seem to be equivalent. Similar results are obtained when all the vertical excitations are calculated from a gas-phase M06-2X optimized geometry (See Appendix A).

	DAD				**$DADH_2^{2+}$**			
	Gas		Water (IEFPCM)		Gas		Water (IEFPCM)	
	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$
	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$
B3LYP	2.64	2.99	2.69	2.72	2.48	3.36	2.64	2.93
CAM-B3LYP	2.87	3.51	2.91	3.26	2.77	3.79	2.86	3.46
ωB97XD	2.85	3.53	2.89	3.28	2.76	3.81	2.84	3.49
M06-2X	2.64	3.52	2.69	3.26	2.55	3.79	2.65	3.42
exp	-	-	-	-	-	-	-	2.73

Table 4.4 $n\pi^*$ and $\pi\pi^*$ vertical excitation energies (eV) of DAD and $DADH_2^{2+}$ in gas phase and in implicit water at the TD-B3LYP, TD-CAM-B3LYP, TD-ωB97XD and TD-M062X levels of theory, using the cc-pVDZ basis set. The starting geometries were optimized at the corresponding level of theory. f indicates the approximate oscillator strength of the states. The experimental value for the $\pi\pi^*$ excitation in water [21] is also specified.

No matter the functional, when both DAD and $DADH_2^{2+}$ are solvated in water, the $n\pi^*$ energy increases and the $\pi\pi^*$ energy decreases. This energy shifts can be attributed to solvatochromism in polar solvents (water). In this case, it seems that the dipole moment in the $n\pi^*$ state is smaller than the dipole of the ground state. Hence, there is a greater stabilization of the ground state by the polar solvent (blue-shift). On the contrary, the dipole moment in the $\pi\pi^*$ state seems to be larger than in the ground state, leading to a greater stabilization of the $\pi\pi^*$ state and the subsequent red-shift. As a result, the gap between both electronic states decreases in polar environments. This is convenient to avoid the electronic relaxation through pathways different from the *trans-cis* photoisomerization, as less excess energy is available to make them accessible upon excitation to the bright $\pi\pi^*$ state.

Conversely, the $n\pi^*$ energy decreases and the $\pi\pi^*$ energy increases upon protonation (larger S_1-S_2 gap). Considering that the main goals of the derivation of azobenzene are red-shifting the $\pi\pi^*$ energy and decreasing the S_1-S_2 gap[13], it seems that the effect of the functional groups is less noticeable in DADH_2^{2+}. In this sense, Table 4.5 shows the $n\pi^*$ and $\pi\pi^*$ azobenzene excitation energies, which essentially involve the same orbitals as DAD.

	Gas		DCM (IEFPCM)	
	$n\pi^*(f\sim0)$	$\pi\pi^*(f\sim1)$	$n\pi^*(f\sim0)$	$\pi\pi^*(f\sim1)$
B3LYP	2.54	3.76	2.56	3.60
CAM-B3LYP	2.74	4.18	2.75	4.03
WB97XD	2.72	4.18	2.74	4.03
M062X	2.51	4.27	2.52	4.12
Exp	-	-	2.5-3.1	3.8

Table 4.5 $n\pi^*$ and $\pi\pi^*$ vertical excitation energies (eV) of azobenzene in gas phase and in implicit DCM at the TD-B3LYP, TD-CAM-B3LYP, TD-ωB97XD and TD-M062X levels of theory, using the cc-pVDZ basis set. The starting geometries were optimized at the corresponding level of theory. The experimental values were obtained from [14].

As expected, the S_1-S_2 gap is larger in azobenzene, since, compared to DAD and DADH_2^{2+}, the $\pi\pi^*$ energy is significantly higher and the $n\pi^*$ is slightly lower. The resulting azobenzene - DADH_2^{2+} - DAD energy gap order can be related to the change in the geometry of the azo moiety (C-N=N-C). According to Table 4.6, larger N=N bonds and shorter C-N bonds favour the desired $\pi\pi^*$ red shifting and the S_1-S_2 energy decrease. Therefore, the protonation state of DAD affects the excitation energies due to the structural change that it produces. Finally, note that the considered functionals follow the same trend as in Table 4.4, and that the effect of the solvent is less noticeable in azobenzene because DCM is less polar than water.

	C-N dist (Å)	**N=N dist (Å)**	**N-C dist (Å)**
DAD	1.40519	1.26417	1.41339
DADH_2^{2+}	1.41831	1.24495	1.4232
Azobenzene	1.42488	1.24333	1.42487

Table 4.6 Bond distances in the Azo moiety of DAD, DADH_2^{2+} and Azobenzene at the M06-2X/cc-pVDZ level of theory in gas phase.

The azobenzene-DADH_2^{2+}-DAD energy trend also appears at the CASSCF and CASPT2 levels of theory (Table 4.7), supporting the TDDFT observations. However, these excitation energies are not always in agreement with the TDDFT estimations. As expected, the CASSCF energies are very large but, fortunately, the MS-CASPT2 corrections provide more reasonable results, indicating the importance of dynamic correlation in these kind of systems.

	DAD		**DADH$_2^{2+}$**		**Azobenzene**	
	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$
SA(5)-CASSCF	4.20	5.29	4.21	5.56	4.18	6.18
MS-CASPT2 IPEA	3.07	3.52	2.96	3.90	2.93	4.47
MS-CASPT2	2.74	2.84	2.62	3.20	2.59	3.88
exp	-	-	-	2.73	2.5-3.1	3.8

Table 4.7 $n\pi^*$ and $\pi\pi^*$ vertical excitation energies (eV) of DAD, DADH$_2^{2+}$ and azobenzene in gas phase at SA(5)-CASSCF and MS-CASPT2 with and without IPEA shift. The energies are optimized at M06-2X/cc-pVDZ. Experimental energies taken from [21] and [14].

Regarding the MS-CASPT2 energies, the IPEA shifted results provide higher values compared to the no IPEA calculations. This is the expected behaviour of the IPEA shift, an empirical correction that assumes an underestimation of the excitation energies based on the systematic errors that are found in the CASPT2 dissociation energies.[39] Upon comparison with the experimental results, it seems that MS-CASPT2 performs better without the IPEA shift. This makes sense considering that the shift is highly dependent on the amount of dynamic correlation, which is large in the systems that are studied in this work. However, the no IPEA MS-CASPT2 $\pi\pi^*$ energies are still higher than the experimental values, perhaps due to the lack of solvent in the CASSCF/CASPT2 calculations. In addition, it could be because the calculations are not from the CASSCF/CASPT2 optimized geometries. In this sense, Table 4.8 shows the SA(3)-CASSCF energies and their corresponding MS-CASPT2 corrections from the SA(3)-CASSCF optimized geometries in DAD, and the CASSCF and CASPT2 azobenzene energies from the minimized structure at the CASPT2 level in [15].

	DAD		**DADH$_2^{2+}$**		**Azobenzene**	
	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$
CASSCF	3.82	4.91	3.71	5.47	3.10	6.16
MS-CASPT2	2.21	3.45	2.04	3.98	2.86	4.18
exp	-	-	-	2.73	2.5-3.1	3.8

Table 4.8 $n\pi^*$ and $\pi\pi^*$ vertical excitation energies (eV) of DAD and DADH$_2^{2+}$ in gas phase at SA(3)-CASSCF and MS-CASPT2 without IPEA shift. The geometries are optimized at SA(3)-CASSCF. Experimental and azobenzene energies taken from [21], [14] and [15].

Surprisingly, the $\pi\pi^*$ energies get away from the experimental values but, at the same time, they get closer to the TDDFT estimations. In any case, these results are not conclusive because the CASSCF calculations use a different number of roots and a smaller active space compared to the calculations in Table 4.7 (unfortunately, the ongoing studies under the same conditions as in Table 4.7 could not be included in this report). Finally, it was not possible to determine the most proper functional among M06-2X, CAM-B3LYP and ωB97XD, as all of them perform similar compared to the experiment. Nevertheless, M06-2X was selected for the remaining calculations because its $n\pi^*$ energy is usually closer to CASPT2, although everything should be repeated with either CAM-B3LYP or ωB97XD. In fact, as they are long-range functionals, CAM-B3LYP and ωB97XD will be a very interesting choice for studying charge-transfer interactions in the protein environment in future works.

Up to now, vertical excitations from only the optimized ground state geometry were calculated. However, molecules are in constant motion and, hence, it is convenient to consider several configurations in order to obtain a more realistic absorption spectrum. In this sense, 100 snapshots were extracted from a 100 ns MD simulation of DADH$_2^{2+}$ in water. From these geometries, single point excitations were calculated in both implicit (IEFPCM) and explicit solvent (QM/MM) at the M06-2X/cc-pVDZ level of theory. The resulting spectra are shown in Figure 4.11a.

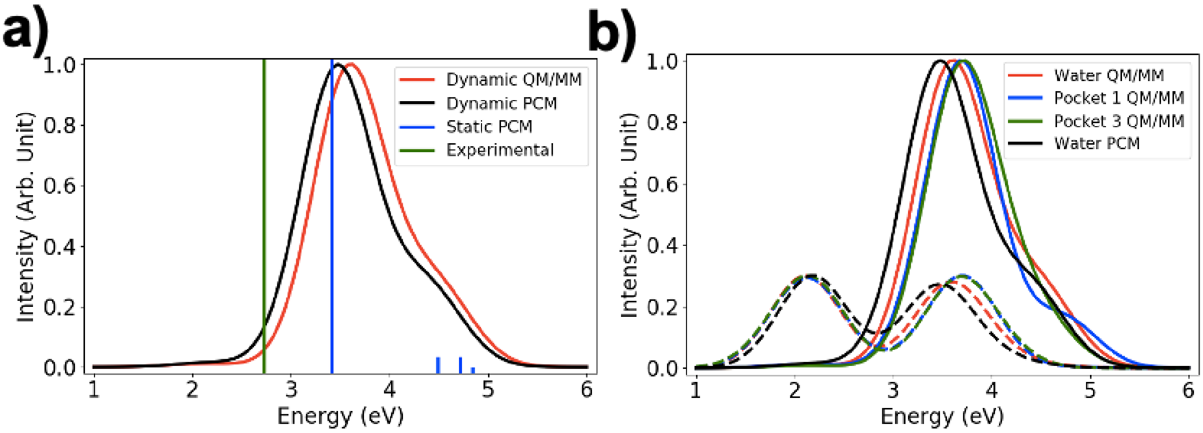

Fig. 4.11 a) Static (PCM) and dynamic (PCM and QM/MM) absorption spectra of DADH$_2^{2+}$ in water. b) Dynamic spectra of DADH$_2^{2+}$ in water (PCM and QM/MM) and in the Na$_V$1.5 pockets 1 and 3 (QM/MM). The dashed lines indicate the $n\pi^*$ and $\pi\pi^*$ density of states.

The static PCM result is the closest to the experimental value. This is probably because the minimum geometry in the $\pi\pi^*$ surface is very close to the Franck Condon region (FC), i.e., the optimized ground state structure. Therefore, when the configurations deviate from FC, the energies of both the ground and the $\pi\pi^*$ states increase. However, the energy change is higher in the $\pi\pi^*$ surface, resulting in higher excitation energies in the dynamic spectra. Upon comparison of QM/MM with PCM, it seems that PCM provides better results. Even though in PCM the solvent is included implicitly, DADH$_2^{2+}$ and water are mutually polarized self-consistently. In this way, PCM succeeds in providing more accurate descriptions of the general bulk-solute interactions. On the other hand, the electrostatic embedding in the performed QM/MM calculations does not consider the polarization of the water molecules by $DADH_2^{2+}$, leading to the most distant energies compared to the experimental results.

Similarly, to study the effect that the environment has on $DADH_2^{2+}$ in its $Na_V1.5$ binding pockets, 100 snapshots were taken from 100 ns MD simulations of $DADH_2^{2+}$ in the binding pockets. There are no significant differences between the resulting spectra in pockets 1 and 3 (Figure 4.11b). Moreover, both of them are slightly blue shifted compared to the spectrum in water. According to the density of states, the main reason of this deviation is the energy change in the $\pi\pi^*$ state, which is probably due to the lower polarity in the protein environment. In order to improve the calculations, the QM region should be bigger and, for example, include the aminoacids that interact the most in the binding pockets or some water molecules. The latter may also be convenient to get energies closer to the experimental value in Figure 4.11a.

Finally, it is important to note that 100 snapshots are sufficient to get converged spectra. In this sense, Figure 4.12 shows the average excitation of $DADH_2^{2+}$, weighted with the corresponding oscillator strengths, in the different environments for 10-100 equidistant snapshots. In all cases, the energy profiles flatten at around 70 snapshots.

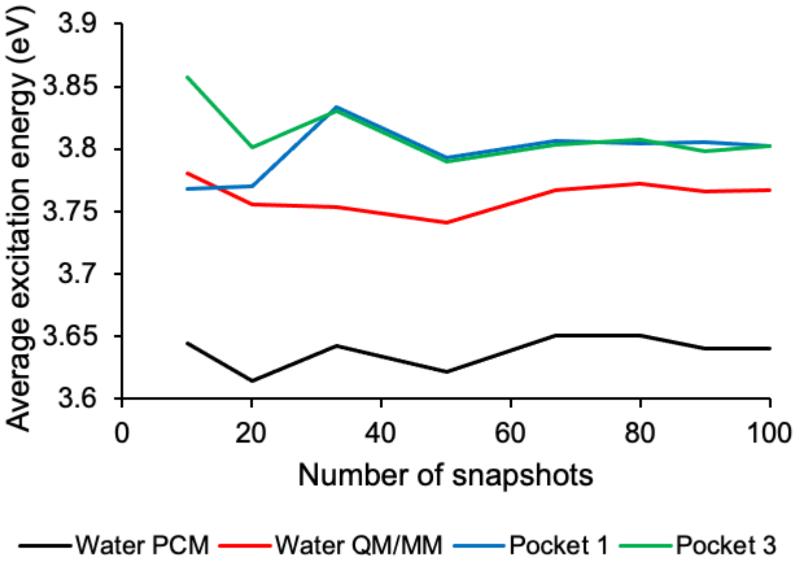

Fig. 4.12 Average excitation energy of $DADH_2^{2+}$ in the dynamic calculations for 10, 20, 33, 50, 67, 80, 90 and 100 snapshots.

4.3 DAD Photoisomerization

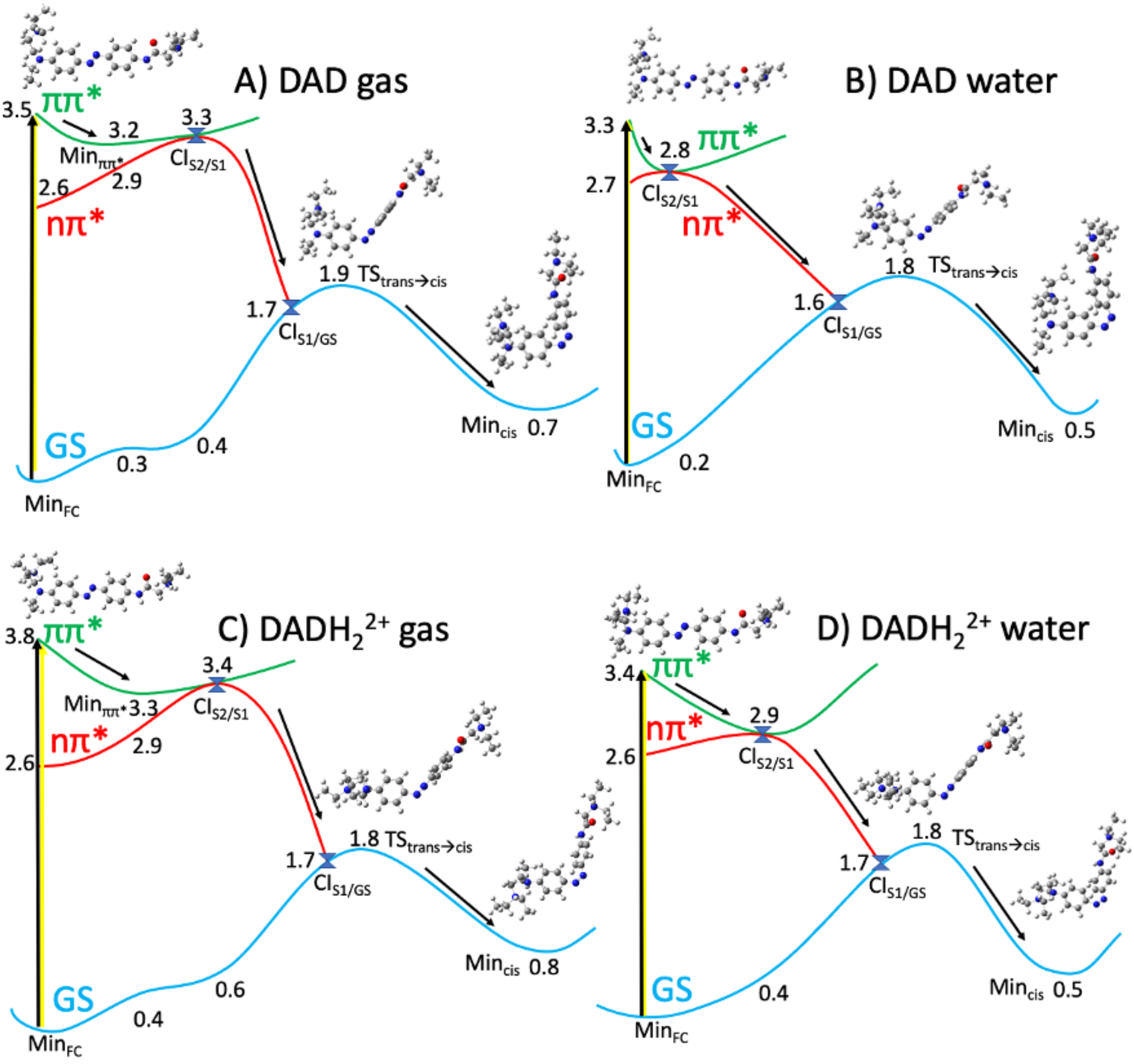
A) DAD gas
3.5
ππ*
3.2
3.3
Min$_{\pi\pi^*}$
Cl$_{S2/S1}$
2.6
2.9
nπ*
1.9 TS$_{trans\rightarrow cis}$
1.7
Cl$_{S1/GS}$
GS
0.3
0.4
Min$_{cis}$ 0.7
Min$_{FC}$
B) DAD water
3.3
2.8
ππ*
2.7
Cl$_{S2/S1}$
nπ*
1.8 TS$_{trans\rightarrow cis}$
1.6
Cl$_{S1/GS}$
GS
0.2
Min$_{FC}$
Min$_{cis}$ 0.5
C) DADH$_2^{2+}$ gas
3.8
ππ*
3.4
Min$_{\pi\pi^*}$ 3.3
Cl$_{S2/S1}$
nπ*
2.9
2.6
1.8 TS$_{trans\rightarrow cis}$
1.7
Cl$_{S1/GS}$
Min$_{cis}$ 0.8
GS
0.6
0.4
Min$_{FC}$
D) DADH$_2^{2+}$ water
3.4
ππ*
nπ*
2.9
Cl$_{S2/S1}$
2.6
1.8 TS$_{trans\rightarrow cis}$
1.7
Cl$_{S1/GS}$
GS
0.4
Min$_{cis}$ 0.5
Min$_{FC}$

Fig. 4.13 *trans*-*cis* photoisomerization pathway of DAD and DADH$_2^{2+}$ in gas phase and in water after excitation to the $\pi\pi^*$ (bright) excited state. Energies (in eV) calculated at M06-2X/cc-pVDZ level of theory relative to the FC ground state.

Figure 4.13 shows the *trans*-*cis* photoisomerization pathway of DAD and DADH$_2^{2+}$ in gas phase and in water, after excitation to the $\pi\pi^*$ (bright) excited state, at the M06-2X/cc-pVDZ level of theory. Similarly, the *trans*-*cis* photoisomerization pathway of azobenzene in gas phase and in DCM is presented in Figure 4.14. Upon excitation, the $\pi\pi^*$ state geometry relaxes to a minimum or directly to a crossing point with the $n\pi^*$ state. The minimum in the $\pi\pi^*$ PES appears in gas phase in DAD and $DADH_2^{2+}$, and in both azobenzene PES. This seems to be related to the red shift in the $\pi\pi^*$ excitations. As explained earlier, when solvating both DAD species in water, the $\pi\pi^*$ excitation energy decreases, and the $n\pi^*$ and $\pi\pi^*$ states get closer in energy. This changes the PES profile close to FC and, in this specific case, it results in the disappearance of the $\pi\pi^*$ minima in water. Compared to azobenzene in gas phase, the $\pi\pi^*$ minima in the DAD species are closer to the corresponding crossing points with the $n\pi^*$ state. Therefore, DAD and $DADH_2^{2+}$ will probably not stay long in this minimum before reaching the crossing point. On the other hand, there is a larger energy difference between the $\pi\pi^*$ minimum and the crossing point with $n\pi^*$ in azobenzene. As a result, the excited molecule may stay in the minimum long enough to deactivate by spontaneous emission, since its oscillator strength is large. Then, upon solvation, the $\pi\pi^*$ minimum gets closer to the crossing point, but it does not disappear as in the DAD species.

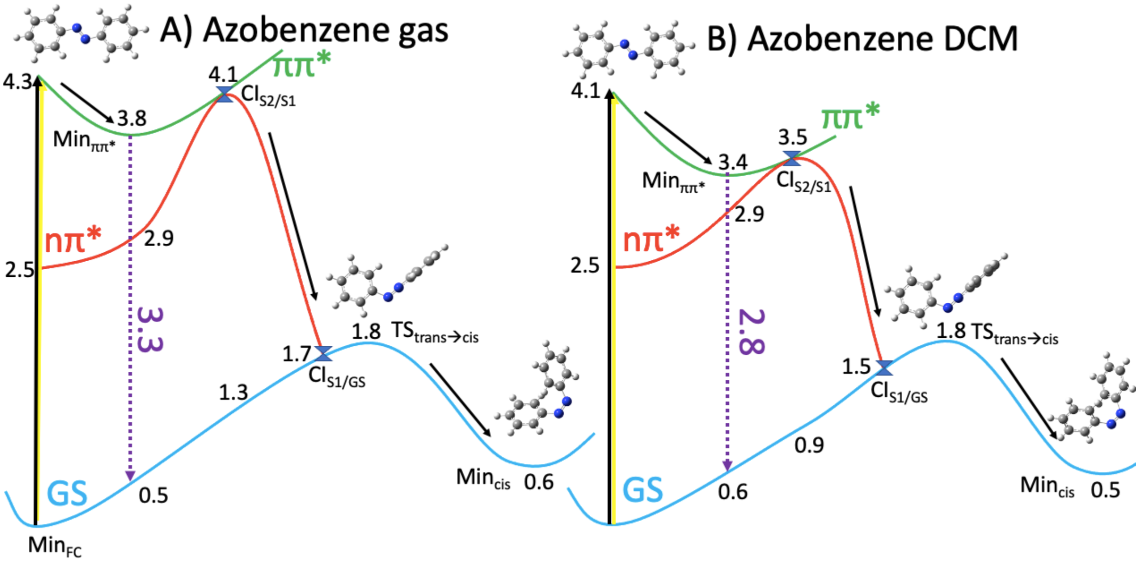

Fig. 4.14 *trans-cis* photoisomerization pathway of azobenzene in gas phase and in DCM after excitation to the $\pi\pi^*$ (bright) excited state. Energies (in eV) calculated at M06-2X/cc-pVDZ level of theory relative to the FC ground state.

When reaching the crossing points, the excited systems may jump to the $n\pi^*$ surface. From this point on, the PESs present no significant differences. In general terms, the system relaxes to a crossing point with the ground state, which is 0.1-0.2 eV lower in energy than the *trans-cis* TS in the ground state. Hence, if it crosses to the ground state and it reaches the TS, it may photoisomerize to the *cis*-DAD. Alternatively, it may also go back to the *trans*-DAD. The geometries also follow a similar trend throughout the photoisomerization pathways in all systems. The structures of the $\pi\pi^*$ minima and the crossing points with the $n\pi^*$ state are very close to FC. Then, when relaxing in the $n\pi^*$ surface, the C-N=N-C dihedral rotates and reaches almost 90° in the crossing point with the ground state. The dihedral is even closer to 90° in the TS, which continues evolving until the *cis* minimum.

All this suggests that the key point in the photoisomerization is the absence or presence of a minimum in the $\pi\pi^*$ surface, and how close it is to the crossing point with $n\pi^*$. This depends on the derivation of azobenzene and on solvation, since both of them favour the reduction in the energy gap between the $\pi\pi^*$ minimum and the $\pi\pi^*/n\pi^*$ crossing point. As a result, the solvated DAD species will probably present the best quantum yields among the considered systems, given that the absence of an emissive $\pi\pi^*$ minimum will not give the chance of relaxation through this alternative pathway. Following the same reasoning, the lowest quantum yield is expected in azobenzene in gas phase.

Note that there are no significant differences between the photoisomerizations of DAD and DADH_2^{2+} despite the observed energy shifts in the absorption spectra. In principle, the higher S_1-S_2 energy gap in DADH_2^{2+} could indicate a less efficient photoisomerization. However, it could actually be more favoured compared to DAD, since the crossing point between the $n\pi^*$ and the ground state is closer to the TS. Nevertheless, we are considering 0.1 eV energy differences, which could also be errors in the DFT calculation, specially in the excited/ground state intersection, where TDDFT often fails.

In any case, what is clear is that the environment strongly affects the DAD photoisomerization. Hence, a proper characterization of the interactions with the environment will be required in future works simulating the photoisomerization of DAD in the $\mathrm{Na}_V1.5$ binding pockets.

4.4 Outlook

On the light of these results, it is clear that more calculations are needed in order to understand the ion conduction blocking mechanism induced by DAD inside the $\mathrm{Na}_V1.5$ channel. Likewise, a more extensive characterisation of the static and dynamic aspects of the deactivation of DAD are also required for a deeper insight onto its photoisomerization mechanism. In this sense, it would be convenient to calculate the whole photoisomerization PES with more DFT functionals and at the CASPT2 level of theory. Then, upon comparison with the CASPT2 PES, a proper DFT functional could be chosen to simulate the photoisomerization of DAD, using QM/MM, in one of its binding pockets and including relevant residues in the QM region. In addition, we could also investigate the DAD photoisomerization by non-adiabatic dynamics to obtain properties such as quantum yields and reaction times. Finally, MD calculations could be performed to explore the effect that DAD has on ion conduction.

Moreover, it was noted that electrostatic interactions play an important role in the stabilisation of the binding pockets of $DADH_2^{2+}$. Hence, as mentioned earlier, it would be interesting to repeat the same study with deprotonated DAD to find its corresponding binding pockets, since they may differ from the ones of $DADH_2^{2+}$. Although $DADH_2^{2+}$ is more abundant at physiological pH, it exists in equilibrium with deprotonated DAD. Considering that only neutral species are capable of diffusing across cell membranes and entering ion channels, it is possible that the concentration of deprotonated DAD is higher than expected inside the pore of $Na_V1.5$. Therefore, different binding pockets that depend on the protonation state may result in different biological activities, which would be inconvenient.

Additionally, polarizable force fields would probably render a better description of the interactions. Hence, a next step of this project involves the use of the AMOEBA force field, as implemented in Tinker-HP, to find the binding pockets of DAD. Significant differences compared to non-polarizable force fields would probably mean that the latter are not good enough to study this system. In this sense, some preliminary results with AMOEBA have already been obtained. Up to now, 60 ns of a POPC membrane have been simulated, and some properties, shown in Figure 4.15, have been calculated and compared with a 1000 ns POPC non-polarizable calculation.

All the properties depicted in Figure 4.15 are similar no matter the kind of force field. Perhaps the most noticeable difference is the magnitude of the fluctuations in the area per lipid graph (Figure 4.15A), where it seems that polarization helps in keeping it constant. On the other hand, the differences in the electron density and the Deuterium order parameter (SCD) cannot be exclusively attributed to the force field yet, as it is possible that the polarizable calculation has not converged in 60 ns. In any case, it is still interesting to carry on with this study to analyze the effect of polarization on the interactions of DAD in the binding pockets. To this end, the input system was prepared for a GaMD simulation in Tinker and preliminary AMOEBA parameters were obtained using poltype2. Unfortunately, these calculations could not be run due to the lack of computational resources.

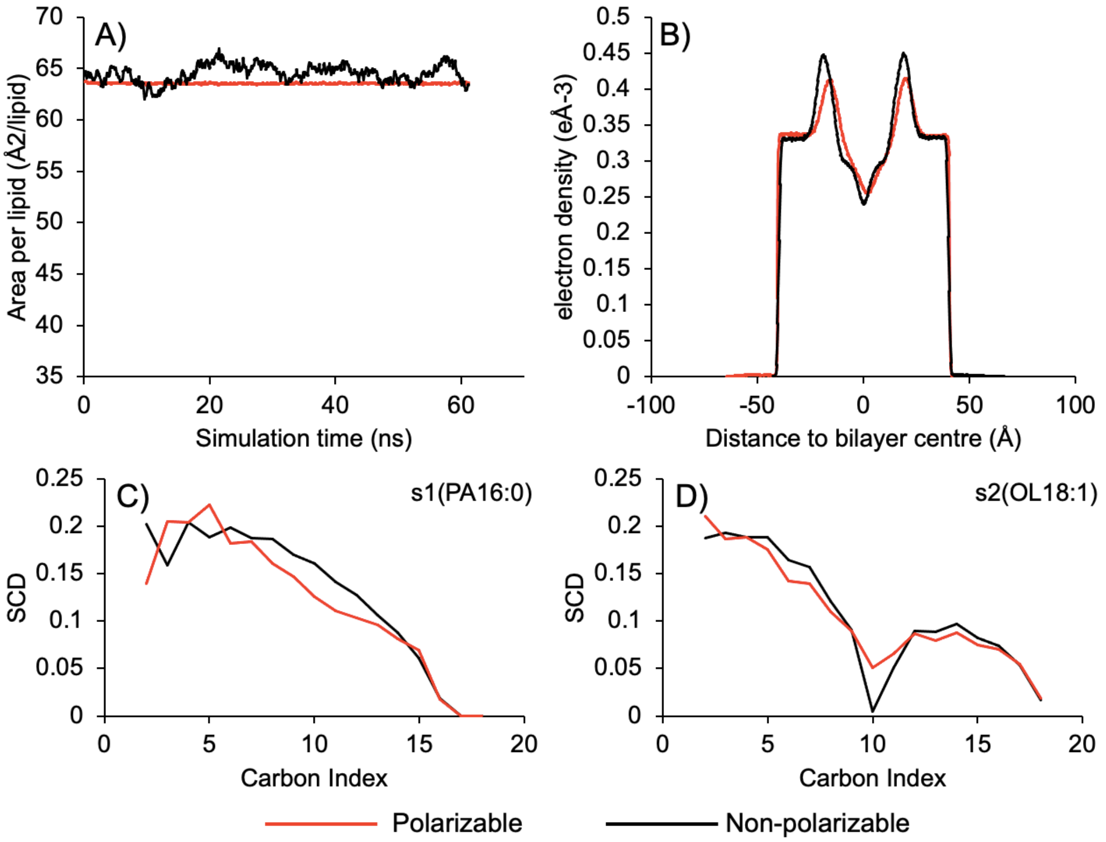

Fig. 4.15 Comparison of the area per lipid (A), the electron density across the membrane (B), the deuterium order parameter (SCD) of the palmitic (C) and the oleic acid chains (D) between a 60 ns POPC polarizable calculation and a 1000 ns non-polarizable one.

Conclusions

In this work, classical simulations, including docking, MD and GaMD, and static quantum and QM/MM calculations were performed to find the binding pockets of DAD in the $\mathrm{Na}_V1.5$ channel, compute its absorption spectrum in gas phase, solvent and in the most stable binding pockets, and explore the photoisomerization mechanism of DAD in gas phase and in solvent. The analysis of the resulting data lead to the following conclusions:

- Simulations containing only the pore of the $\mathrm{Na}_V1.5$ channel are enough to calculate its properties due to its resemblance to the pore structure in simulations containing the whole protein.

- Docking analysis only serves as an initial estimation of the binding poses of DADH_2^{2+}, given that they differ from the more accurate GaMD results.

- Out of the 3 binding pockets calculated with GaMD, only pockets 1 and 3 represent stable positions, where pocket 1 is the most favourable one. Furthermore, DADH_2^{2+} in pocket 3 is very close to the DEKA aminoacids and, thus, it may hinder ion conduction.

- The most important contributions to the stabilisation of the binding pockets are of electrostatic nature, indicating that deprotonated DAD may show different binding pockets from the ones obtained for DADH_2^{2+}. Moreover, it would be interesting to explore the effect of polarization by using polarizable force fields.

- Upon polar solvation, the $n\pi^*$ and $\pi\pi^*$ excitations in DAD get closer in energy due to the relative stabilization of the ground and excited states. This is convenient to avoid relaxation through pathways different from the *trans-cis* photoisomerization. The derivation of azobenzene has the same effect, which is related to the change in the geometry of the azo moiety. The opposite structural change occurs when protonating DAD. Hence, the $n\pi^*/\pi\pi^*$ energy gap increases.

- The inclusion of dynamical correlation is crucial in the excited states calculations in azobenzene and the DAD species. Moreover, CASPT2 energies are closer to the experimental results without the IPEA shift.

- The static PCM absorption spectrum renders lower energies than the dynamic spectra because the minimum in the $\pi\pi^*$ state is close to the FC region. Moreover, QM/PCM offers a better description than QM/MM due to the inclusion mutual polarization between the solvent and the solute by PCM. Finally, to improve the energies, the QM region should be increased.

- The presence of a minimum in the $\pi\pi^*$ state is key in the *trans-cis* photoisomerization. Solvation and the derivation of azobenzene allow a lower energy gap from the $\pi\pi^*$ minimum to the crossing point with $n\pi^*$, which favours the photoisomerization. This indicates that the interactions with the environment are crucial in the photoisomerization.

- After reaching the $n\pi^*$ state, the photoisomerization proceeds similarly in all the considered systems. The evolution of the geometries also follows the same trend, which is governed by the rotation of the C-N=N-C dihedral in the $n\pi^*$ state.

References

[1] R. Cassia-Moura. *Chapter 19 - The quest for ion channel memory using a planar BLM. Planar Lipid Bilayers (BLMs) and Their Applications*, volume 7 of *Membrane Science and Technology*. Elsevier, 2003.

[2] Juan J. Nogueira and Ben Corry. *Ion Channel Permeation and Selectivity. In the Oxford Handbook of Neuronal Ion Channels*. Oxford University Press, 2019.

[3] Harry Fozzard, Michael Sheets, and Dorothy Hanck. The sodium channel as a target for local anesthetic drugs. *Frontiers in Pharmacology*, 2, 2011.

[4] A. Scholz. Mechanisms of (local) anaesthetics on voltage-gated sodium and other ion channels. *British Journal of Anaesthesia*, 89(1):52–61, 07 2002.

[5] Sharan K. Bagal, Alan D. Brown, Peter J. Cox, Kiyoyuki Omoto, Robert M. Owen, David C. Pryde, Benjamin Sidders, Sarah E. Skerratt, Edward B. Stevens, R. Ian Storer, and Nigel A. Swain. Ion channels as therapeutic targets: A drug discovery perspective. *Journal of Medicinal Chemistry*, 56(3):593–624, 2013. PMID: 23121096.

[6] Katharina Hüll, Johannes Morstein, and Dirk Trauner. In vivo photopharmacology. *Chemical Reviews*, 118(21):10710–10747, 2018. PMID: 29985590.

[7] David J. Huggins, Woody Sherman, and Bruce Tidor. Rational approaches to improving selectivity in drug design. *Journal of Medicinal Chemistry*, 55, 2012.

[8] DB Longley and PG Johnston. Molecular mechanisms of drug resistance. *The Journal of Pathology*, 205(2):275–292, 2005.

[9] Jean Carlet, Peter Collignon, Don Goldmann, Herman Goossens, Inge C Gyssens, Stephan Harbarth, Vincent Jarlier, Stuart B Levy, Babacar N'Doye, Didier Pittet, Rosana Richtmann, Wing H Seto, Jos WM van der Meer, and Andreas Voss. Society's failure to protect a precious resource: antibiotics. *The Lancet*, 378(9788):369–371, 2011.

[10] Willem A. Velema, Wiktor Szymanski, and Ben L. Feringa. Photopharmacology: Beyond proof of principle. *Journal of the American Chemical Society*, 136(6):2178–2191, 2014. PMID: 24456115.

[11] Andrew A. Beharry and G. Andrew Woolley. Azobenzene photoswitches for biomolecules. *Chemical Society Reviews*, 40:4422–4437, 2011.

[12] Petr Klán and Jakob Wirz. *Photochemistry of Organic Compounds: From Concepts to Practice*. John Wiley & Sons Ltd., Chichester, UK, 1 edition, 2009.

[13] H. M. Dhammika Bandara and Shawn C. Burdette. Photoisomerization in different classes of azobenzene. *Chemical Society Reviews*, 41:1809–1825, 2012.

[14] Jianqiang Zhao, Yue Zhang, Lihua Gan, and Guanlei Wang. Experimental and dft study of uv–vis absorption spectra of azobenzene containing ester groups. *Computational and Theoretical Chemistry*, 1200:113244, 2021.

[15] Josep Casellas, Michael J. Bearpark, and Mar Reguero. Excited-state decay in the photoisomerisation of azobenzene: A new balance between mechanisms. *ChemPhysChem*, 17(19):3068–3079, 2016.

[16] Hermann Rau and Erik Lueddecke. On the rotation-inversion controversy on photoisomerization of azobenzenes. experimental proof of inversion. *Journal of the American Chemical Society*, 104(6):1616–1620, 1982.

[17] J. Griffiths. Ii. photochemistry of azobenzene and its derivatives. *Chemical Society Reviews*, 1:481–493, 1972.

[18] Sandra Monti, Giorgio Orlandi, and Paolo Palmieri. Features of the photochemically active state surfaces of azobenzene. *Chemical Physics*, 71(1):87–99, 1982.

[19] Shuai Yuan, Yusheng Dou, Weifeng Wu, Yun Hu, and Jianshe Zhao. Why does trans-azobenzene have a smaller isomerization yield for $\pi\pi^*$ excitation than for $n\pi^*$ excitation? *The Journal of Physical Chemistry A*, 112(51):13326–13334, 2008. PMID: 19053550.

[20] Le Yu, Chao Xu, Yibo Lei, Chaoyuan Zhu, and Zhenyi Wen. Trajectory-based nonadiabatic molecular dynamics without calculating nonadiabatic coupling in the avoided crossing case: trans - cis photoisomerization in azobenzene. *Physical Chemistry Chemical Physics*, 16:25883–25895, 2014.

[21] Laura Laprell, Ivan Tochitsky, Kuldeep Kaur, Michael B. Manookin, Marco Stein, David M. Barber, Christian Schön, Stylianos Michalakis, Martin Biel, Richard H. Kramer, Martin P. Sumser, Dirk Trauner, and Russell N. Van Gelder. Photopharmacological control of bipolar cells restores visual function in blind mice. *The Journal of Clinical Investigation*, 127(7):2598–2611, 6 2017.

[22] Manuel de Lera Ruiz and Richard L. Kraus. Voltage-gated sodium channels: Structure, function, pharmacology, and clinical indications. *Journal of Medicinal Chemistry*, 58(18):7093–7118, 2015. PMID: 25927480.

[23] Matthew E. Cross and Emma V. E. Plunkett. *Action potentials*. Cambridge University Press, 2 edition, 2014.

[24] Daohua Jiang, Richard Banh, Tamer M. Gamal El-Din, Lige Tonggu, Michael J. Lenaeus, Régis Pomès, Ning Zheng, and William A. Catterall. Open-state structure and pore gating mechanism of the cardiac sodium channel. *Cell*, 184(20):5151–5162.e11, 2021.

[25] Palmisano VF, Anguita-Ortiz N, Faraji S, and Nogueira JJ. Light-induced control of voltage gated ion channels. *ChemRxiv. Cambridge: Cambridge Open Engage*, 2022.

[26] William A. Catterall and Teresa M. Swanson. Structural basis for pharmacology of voltage-gated sodium and calcium channels. *Molecular Pharmacology*, 88(1):141–150, 2015.

[27] Mingyan Zhu and Huchen Zhou. Azobenzene-based small molecular photoswitches for protein modulation. *Organic & Biomolecular Chemistry*, 16:8434–8445, 2018.

[28] Mourot A, Fehrentz T, Le Feuvre Y, Smith CM, Herold C, Dalkara D, Nagy F, Trauner D, and Kramer RH. Rapid optical control of nociception with an ion-channel photoswitch. *Nature Methods*, 9:396–402, 2012.

[29] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28:1049–1070, 1926.

[30] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[31] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry*. Dover publications, 1996.

[32] B. O. Roos, P. R. Taylor, and P. E. Sigbahn. A complete active space scf method (casscf) using density matrix formulated super-ci approach. *Chemical Physics*, 48:157–173, 1980.

[33] B. O. Roos. The complete active space scf method in a fock-matrix-based super-ci formulation. *International Journal of Quantum Chemistry*, 18:175–189, 1980.

[34] K. Andersson, P.A. Malmqvist, B. O. Roos, A.J. Sadlej, and K. Wolinski. Second-order perturbation theory with a casscf reference function. *The Journal of Physical Chemistry*, 94:5483–5488, 1990.

[35] K. Andersson, P.A. Malmqvist, and B. O. Roos. Second-order perturbation theory with a complete active space self-consistent field reference function. *The Journal of Chemical Physics*, 96:1218–1226, 1992.

[36] I. N. Levine. *Quantum Chemistry*. Prentice Hall, 2000.

[37] Frank Jensen. *Introduction to Computational Chemistry*. Wiley, 2nd edition, 2007.

[38] B. O. Roos, K. Andersson, M. P. Fiilscher, L. Serrano-Andres, K. Pierloot, M. Merchan, and V. Molina. Applications of level shift corrected perturbation theory in electronic spectroscopy. *Journal of Molecular Structure: THEOCHEM*, 388:257–276, 1996.

[39] J. Patrick Zobel, Juan J. Nogueira, and Leticia González. The ipea dilemma in caspt2. *Chemical Science*, 8:1482–1499, 2017.

[40] J. Finley, A. Malmqvist, B. O. Roos, and L. Serrano-Andres. The multi-state caspt2 method. *Chemical Physics Letters*, 288:299–306, 1998.

[41] W. Koch and M.C Holthausen. *A chemist's guide to density functional theory*. Wiley-VCH, 2000.

[42] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864–B871, 1964.

[43] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58(8):1200–1211, 1980.

[44] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38:3098–3100, 1988.

[45] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37:785–789, 1988.

[46] Y. Zhao and D. G. Truhlar. A new local density functional for main group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics*, 125:194101, 2006.

[47] A. D. Becke. Density-functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98:5648, 1993.

[48] Tobias Benighaus, Robert A. DiStasio, Rohini C. Lochan, Jeng-Da Chai, and Martin Head-Gordon. Semiempirical double-hybrid density functional with improved description of long-range correlation. *The Journal of Physical Chemistry A*, 112(12):2702–2712, 2008.

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

[50] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics*, 10:6615–6620, 2008.

[51] J. P. Perdew and K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, pages 1–20, 2001.

[52] Miguel A.L. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, and E.K.U Gross. *Time-Dependent Density Functional Theory*, volume 706. Springer Berlin Heidelberg, 2006.

[53] Christopher J. Cramer. *Essentials of Computational Chemistry. Theories and Models.* Wiley, 2nd edition, 2004.

[54] James B. Foresman and AEleen Frisch. *Exploring Chemistry with Electronic Structure Methods*. Gaussian Inc., 5 edition, 2015.

[55] Jr. Dunning and T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, 90:1007–1023, 1989.

[56] F. Plasser. Theodore: A toolbox for a detailed and automated analysis of electronic excited state computations. *The Journal of Chemical Physics*, 152:084108, 2 2020.

[57] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Oxford University Press, 1987.

[58] Stewart A. Adcock and J. Andrew McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106:1589–1615, 5 2006.

[59] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159:98–103, 7 1967.

[60] R. Hockney. The potential calculation and some applications. *Methods in Computational Physics*, 9:136–211, 1970.

[61] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J.C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 3 1977.

[62] Hans C Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52:24–34, 10 1983.

[63] M. Tuckerman, H. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3):1990–2001, 1992.

[64] Efrem Braun, Justin Gilmer, Heather B. Mayes, David L. Mobley, Jacob I Monroe, Samarjeet Prasad, and Daniel M. Zuckerman. Best practices for foundations in molecular simulations. *Living Journal of Computational Molecular Science*, 11 2018.

[65] Zhifeng Jing, Chengwen Liu, Sara Y. Cheng, Rui Qi, Brandon D. Walker, Jean-Philip Piquemal, and Pengyu Ren. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annual Review of Biophysics*, 48(1):371–394, 2019.

[66] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 2007.

[67] Ruslan L. Davidchack, Richard Handel, and M. V. Tretyakov. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics*, 130:234101, 6 2009.

[68] T. W. Lion and R. J. Allen. Computing the local pressure in molecular dynamics simulations. *Journal of Physics Condensed Matter*, 24:284133, 2012.

[69] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81:3684–3690, 10 1984.

[70] Johan Aqvist, Petra Wennerström, Martin Nervall, Sinisa Bjelic, and Bjørn Brandsdal. Molecular dynamics simulations of water and biomolecules wit a monte carlo constant pressure algorithm. *Chemical Physics Letters*, 384:288–294, 2004.

[71] Clara D. Christ, Alan E. Mark, and Wilfred F. van Gunsteren. Basic ingredients of free energy calculations: A review. *Journal of Computational Chemistry*, 31(8):1569–1582, 2010.

[72] Yinglong Miao, Victoria A. Feher, and J. Andrew McCammon. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *Journal of Chemical Theory and Computation*, 11(8):3584–3595, 2015.

[73] Xuan-Yu Meng, Ming-Xing Song, Mihaly Mezei, and Meng Cui. Molecular docking: A powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7:146–57, 06 2011.

[74] Samuel Genheden and Ulf Ryde. The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5):449–461, 2015.

[75] Hans Martin Senn and Walter Thiel. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48:1198–1229, 2 2009.

[76] Marco Caricato, Benedetta Mennucci, Jacopo Tomasi, Francesca Ingrosso, Roberto Cammi, Stefano Corni, and Giovanni Scalmani. Formation and relaxation of excited states in solution: A new time dependent polarizable continuum model based on time dependent density functional theory. *The Journal of Chemical Physics*, 124:124520, 3 2006.

[77] Emilia L. Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C. Song, Eder M. Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M. Venable, Jeffery B. Klauda, and Wonpil Im. Charmm-gui membrane builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, 35(27):1997–2004, 2014.

[78] D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, M.K. Gilson G. Giambasu, H. Gohlke, R Harris A.W. Goetz, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, and P.A. Kollman. Amber 2020, 2020. University of California, San Francisco.

[79] Louis Lagardère, Luc-Henri Jolly, Filippo Lipparini, Félix Aviat, Benjamin Stamm, Zhifeng F. Jing, Matthew Harger, Hedieh Torabifard, G. Andrés Cisneros, Michael J. Schnieders, Nohad Gresh, Yvon Maday, Pengyu Y. Ren, Jay W. Ponder, and Jean-Philip Piquemal. Tinker-hp: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science*, 9:956–972, 2018.

[80] Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021.

[81] Daniel R. Roe and Thomas E. III Cheatham. Ptraj and cpptraj: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, 2013.

[82] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[83] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[84] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[85] Roy Dennington, Todd A. Keith, and John M. Millam. Gaussview Version 6, 2019. Semichem Inc. Shawnee Mission KS.

[86] Yan Zhao and Donald G. Truhlar. 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. *Theoretical Chemistry Accounts*, 120(1-3):215–241, 2008.

[87] J. Tomasi, B. Mennucci, and E. Cancès. The ief version of the pcm solvation method: an overview of a new method addressed to study molecular solutes at the qm ab initio level. *Journal of Molecular Structure: THEOCHEM*, 464:211–226, 5 1999.

[88] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian~16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[89] Junmei Wang, Wei Wang, Peter A. Kollman, and David A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247–260, 2006.

[90] Chuan Tian, Koushik Kasavajhala, Kellon A. A. Belfon, Lauren Raguette, He Huang, Angela N. Migues, John Bickel, Yuzhang Wang, Jorge Pincay, Qin Wu, and Carlos Simmerling. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, 16(1):528–552, 2020.

[91] Pekka Mark and Lennart Nilsson. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, 105(43):9954–9960, 2001.

[92] He X, Man VH, Yang W, Lee TS, and Wang J. A fast and high-quality charge model for the next generation general amber force field. *The Journal of Chemical Physics*, 11(153):114502, 2020.

[93] Callum J. Dickson, Ross C. Walker, and Ian R. Gould. Lipid21: Complex lipid membrane simulations with amber. *Journal of Chemical Theory and Computation*, 18(3):1726–1736, 2022.

[94] Brandon Walker, Chengwen Liu, Elizabeth Wait, and Pengyu Ren. Automation of amoeba polarizable force field for small molecules: Poltype 2. *Journal of Computational Chemistry*, 43(23):1530–1542, 2022.

[95] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren. Polarizable atomic multipole-based amoeba force field for proteins. *Journal of Chemical Theory and Computation*, 9(9):4046–4063, 2013.

[96] Zhifeng Jing, Joshua A. Rackers, Lawrence R. Pratt, Chengwen Liu, Susan B. Rempe, and Pengyu Ren. Thermodynamics of ion binding and occupancy in potassium channels. *Chem. Sci.*, 12:8920–8930, 2021.

[97] Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(33), 2011.

[98] Martin McCullagh, Ignacio Franco, Mark A. Ratner, and George C. Schatz. Dna-based optomechanical molecular motor. *Journal of the American Chemical Society*, 133(10):3452–3459, 2011.

[99] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 2007.

[100] K. Serkowski. Polarization techniques. volume 12 of *Methods in Experimental Physics*, pages 361–414. Academic Press, 1974.

[101] Yinglong Miao, William Sinko, Levi Pierce, Denis Bucher, Ross C. Walker, and J. Andrew McCammon. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*, 10(7):2677–2689, 2014.

[102] Gustavo Cárdenas, Jesús Lucia-Tamudo, Henar Mateo-delaFuente, Vito F. Palmisano, Nuria Anguita-Ortiz, Lorena Ruano, Álvaro Pérez-Barcia, Sergio Díaz-Tendero, Marcos Mandado, and Juan J. Nogueira. Mobiotools: A toolkit to setup quantum mechanics/molecular mechanics calculations. *Journal of Computational Chemistry*, 44(4):516–533, 2023.

[103] Ignacio Fdez. Galvan, Morgane Vacher, Ali Alavi, Celestino Angeli, Francesco Aquilante, Jochen Autschbach, Jie J. Bao, Sergey I. Bokarev, Nikolay A. Bogdanov, Rebecca K. Carlson, Liviu F. Chibotaru, Joel Creutzberg, Nike Dattani, Mickaël G. Delcey, Sijia S. Dong, Andreas Dreuw, Leon Freitag, Luis Manuel Frutos, Laura Gagliardi, Frederic Gendron, Angelo Giussani, Leticia González, Gilbert Grell, Meiyuan Guo, Chad E. Hoyer, Marcus Johansson, Sebastian Keller, Stefan Knecht, Goran Kovačević, Erik Källman, Giovanni Li Manni, Marcus Lundberg, Yingjin Ma, Sebastian Mai, João Pedro Malhado, Per Åke Malmqvist, Philipp Marquetand, Stefanie A. Mewes, Jesper Norell, Massimo Olivucci, Markus Oppel, Quan Manh Phung, Kristine Pierloot, Felix Plasser, Markus Reiher, Andrew M. Sand, Igor Schapiro, Prachi Sharma, Christopher J. Stein, Lasse Kragh Sørensen, Donald G. Truhlar, Mihkel Ugandi, Liviu Ungur, Alessio Valentini, Steven Vancoillie, Valera Veryazov, Oskar Weser, Tomasz A. Wesołowski, Per-Olof Widmark, Sebastian Wouters, Alexander Zech, J. Patrick Zobel, and Roland Lindh. Openmolcas: From source code to insight. *Journal of Chemical Theory and Computation*, 15:5925–5964, 11 2019.

[104] G. Schaftenaar, E. Vlieg, and Vriend. G. molden 2.0: quantum chemistry meets proteins. *Journal of Computer-Aided Molecular Design*, 31:789–800, 2017.

[105] Sebastian Mai, Philipp Marquetand, and Leticia Gonzalez. Nonadiabatic dynamics: The sharc approach. *WIREs Computational Molecular Science*, 8:1370, 2018.

[106] Martin B. Ulmschneider, Claire Bagneris, Emily C. McCusker, Paul G. DeCaen, Markus Delling, David E. Clapham, Jakob P. Ulmschneider, and B. A. Wallace. Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. *Proceedings of the National Academy of Sciences*, 110(16):6364–6369, 2013.

Benchmarking of excitations from a common M06-2X optimized geometry

	DAD				$\mathbf{DADH_2^{2+}}$			
	Gas		Water (IEFPCM)		Gas		Water (IEFPCM)	
	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$	$n\pi^*$	$\pi\pi^*$
	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$	$f \sim 0$	$f \sim 1$
B3LYP	2.71	3.06	2.77	2.81	2.60	3.31	2.73	3.09
CAM-B3LYP	2.89	3.50	2.95	3.29	2.80	3.76	2.90	3.53
ωB97XD	2.86	3.54	2.91	3.33	2.78	3.79	2.87	3.57
M06-2X	2.64	3.52	2.69	3.26	2.55	3.79	2.65	3.42

Table A.1 $n\pi^*$ and $\pi\pi^*$ vertical excitation energies (eV) of DAD and $DADH_2^{2+}$ in gas phase and in implicit water at the TD-B3LYP, TD-CAM-B3LYP, TD-ωB97XD and TD-M062X levels of theory, using the cc-pVDZ basis set. The starting geometries were optimized at the M06-2X level of theory in gas phase. f indicates the approximate oscillator strength of the states.