
# Computational study of the interaction of a molecular photoswitch with the human $Na_v$ 1.4 ion channel

**Master thesis of :**

Vito Federico Palmisano

**Supervisors :**

prof. Shirin Faraji

dr. Juan José Nogueira

dr. Alexander de Vries

---

**University of Groningen**

**Faculty of Science and Engineering**

**Master in Theoretical Chemistry and Computational Modelling**

July 18, 2020

## Photobiology

I am lying on my back beneath the tree,
dozing, looking up into the canopy,
thinking: what a wonder!—I can see!

But in the greenery above my face,
an even greater miracle is taking place:
Leaves catch photons from the sun
and molecules from air around.
Quanta and carbon atoms become bound.
Life, for them, has just begun.

The sun not only creates life, it also takes away
mostly by deranging DNA.
Damage can be, in part, undone
by enzymes using photons from the sun.

Summer nears its end, already 'cross the sky
southward aiming birds are flying by.
Other birds for travel choose the night
relying on the stars for guiding light.

Imprinted in their little heads are Gemini,
Orion, Dipper, other features of the sky.
There is room for clocks that measure
day and night,
Correct for movement of the sky
and tell the time for flight

Deep into oceans, into caves
the sun cannot directly send its waves.
But through intricacies of foodweb's maze,
oxygen from chloroplasts, luciferin, luciferase,
at times, in place,
where night and darkness seem to reign,
solar quanta emerge as photons
once again.

L.O. Björn 2002

# Abstract

A non-invasive and orthogonal force which can induce motion is light. The ability to deliver information wrapped in form of wavelength and intensity, makes it a perfect stimulus for turning "on" and "off" molecules. At the molecular level, light can induce reversible phototransformations of chemical species causing changes in geometry, flexibility and polarity. Those transformations can occur with an high spatial and temporal precision, requirements needed to interact in a biological environment avoiding alteration of the biomolecular activity. Ion channels are ideal targets for photoswitches since the transport of ions through these membrane proteins can be regulated with light. In this computation work, the interactions of a molecular photoswitch inside a human's brain ion channel were identified, as well as the vertical excitation energy of the photoswitch in the presence of an environment. An accelerated atomistic molecular dynamics method is employed to search for available binding pockets in a human's brain voltage gated ion channel ($Na_v1.4$) for p-diaminoazobenzene, a molecular photoswtich. An end state free energy method is used to characterize the binding free energy of the ligand and the pairwise decomposition energy of interaction of the ligand with surrounding residues. Three distinct binding pockets are identitifed and the major interactions of p-diaminoazobenzene are characterized as van der Walls interactions with hydrophobic side chain aminoacids, phenylalanine and tyrosine. The vertical excitation energy of p-diaminoazobenzene was computed with time dependent density functional theory at polarizable continuum model (PCM) and hybrid quantum mechanics/molecular mechanics (QM/MM) schemes and compared. The vertical excitation energy of p-diaminoazobenzene in the presence of phenylalanine in the QM region and the MM environment are blue shifted by 0.40 eV with B3LYP-D3 and 0.50 eV with CAM-B3LYP-D3 and $\omega$B97X-D in comparison to the PCM calculations. Further calculations are needed to have a more rigorous analysis of the effect of the environment on the photophysics of p-diaminoazobenzene.

# Contents

# 1 Introduction

The ability of humankind to interpret and predict the natural world is remarkable. Humanity began to realize about the microscopic world approximately at the end of the V century B.C. when Leucippus, teacher of Democritus, defined matter as a plurality of fundamental constituents, splitting nature into atoms and emptiness. Nowadays, we can not only observe those microscopical constituents but also predict their behaviour by theoretical predictions and in part, also decide the course of their actions.

## 1.1 Photopharmacology

Pharmacotherapy from the greek word "pharmakon" (drug, poison), is a discipline which deals with the use of drugs and their effect in our body. The use of effective and safe pharmacotherapy is a cornerstone of appropriate patient care for both acute and chronic medical conditions[1]. However, many drugs related issues are still present, which are a major concern to our society. Within those problems, poor drug selectivity is a large limitation, considering that after the drug intake, the substance interacts with processes embedded in complex metabolic pathways[2, 3]. Moreover, poor drug selectivity can lower the threshold level of toxicity narrowing the therapeutic window, leading to a decrease in admissible dosage[4]. Over 85% of therapeutic drugs in clinical research are discarded because of their limited therapeutic window, meaning that many interesting compounds, which could treat currently untreatable conditions are discarded due to insufficient selectivity[5]. Another major problem in pharmacotherapy is drug resistance[6]. It is the ability of cancers or pathogens to acquire protection over the ability of drugs. Particularly, the usage of antimicrobial agents (antibiotics) is under risk due to the emergence of many resistant bacterial strains, and it is believed to cause treatment failures in over 90% of patients with metastatic cancer[7]. This phenomenon can also be related to poor drug selectivity, since the drug should only infect bacteria inside the body of the patient and stop its activity outside of it. This is not the case for conventional drugs which release its activity during all the time that they are in the patients body and also after expulsion, causing a negative impact on the environment. A recent advance in the field, which aim to tackle the problems of efficacy and toxicity, is network pharmacology. This area deals with the core assumption which frames the drug discovery approach, the "one gene, one drug, one disease" approach, where more effective drugs result from highly selective ligands and undesirable toxic side effect are removed. Many effective drugs instead, act via modulation of multiple proteins and can effectively bind to two or more molecular targets. Thus, understanding the polypharmacology of a drug and its effect on biological networks would be essential to improve efficacy and also understand toxicity[8].

To tackle the problem from another prospective, light has been recently thought as a solution[9]. The capacity to have a high temporal resolution over the drug activity and a non-invasive spatial precision is a brilliant approach. Light shows a large degree of orthogonality towards most biochemical systems, since photons do not "contaminate" the system and have a low toxicity. Also it can be delivered with high precision and regulated by adjusting wavelength and intensity to en-

sure a safe usage[10]. Photopharmacology - or optopharmacology - has been developed in the last decade to create chemical compounds able to control the functions of biological molecules using chemical sensitive switches[11]. The ability to control the temporal resolution of pharmaceutical drugs could partially solve some of the effects given by drug resistance as shown in Figure 1.

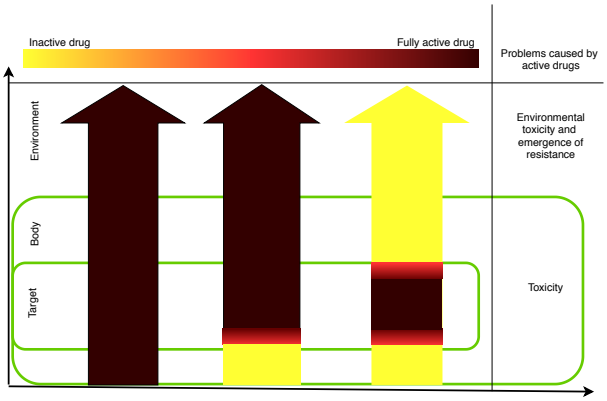


Figure 1: Drug activity for a conventional drug, a prodrug and a reversibly photocontrolled drug over time and space. The activity of a conventional drug is present during the all time in the patient's body, and after ejection, the active drug can have a negative impact on the environment with the evolution of drug-resistant pathogens. Prodrugs delivers the active substance at a later stage and photocontrolled drugs can avoid the negative effects by regulating their activity reversibly in space and time.

The principle of photopharmacology is the introduction of photoswitchable compounds inside the structure of the biomolecular target and reversibly control the drug by light irradiation, to potentially control *in vivo* different physiological conditions[12]. Very recently, a class of azobenzene derivatives has been used to manipulate the activity of voltage gated channels and regulate the transport of ions upon absorption of light[13, 14]. The main goal of this research is to investigate the interactions between p-diaminoazobenzene and the human channel $Na_v1.4$ to find suitable binding pockets and study the electronic excitation of the photoswitch upon absorption of light, by means of computational simulations. In the following sections, the main features of photoswitches and ion channels will be shortly introduced.

## 1.2 Photoswitches and Azobenzene

Molecular photoswitches are synthetic compounds that upon photon absorption, undergo a reversible change in their structure and properties. The process of vision, for example, is based on an elementary chemical step, a photoisomerization around a carbon-carbon double bond in the retinal chromophore, where the compound switches from a cis to a trans conformation upon irradiation. Although, molecular bistabiliy can be achieved with redox reactions, pH changes, temperature and chemical stimuli, photochemical switching has the highest spatial and temporal resolution.

There are a number of requirements that a chemical compound must meet to be introduced into a living system and be used as a photoswitch. The molecule should effectively absorb light at a

wavelength compatible with biological systems, longer than 340nm to avoid tissues damage. Secondly, the modification should lead to a substantial change in structure or properties to have a change in biomolecular activity. Finally, the modification should be stable at biological condition and non-toxic both before and after irradiation[15].

The most widely used photoswitch for biological environments is Azobenzene(AB), since it meets all the criterias established above[9]. The discovery of azo compounds dates back to the mid 1800s, widely used as a colouring agents in the dye industry. AB is present in nature in the trans conformation but in 1937, the evidence of a *cis*-AB conformation was observed after a lack of absorbance when the compound was exposed to light, and since then it became one of the best characterized photoswitches. AB is a diazene derivative in which the two hydrogens are replaced by two phenyl groups. Pure AB is 99% trans in the dark state but the molecule can either exist in the trans or cis conformation[16]. As shown in Figure 2, the trans $\rightarrow$ cis isomerization occurs photochemically with irradiation in the UV spectrum while the cis $\rightarrow$ trans isomerization occurs both thermally and upon UV irradiation[17].

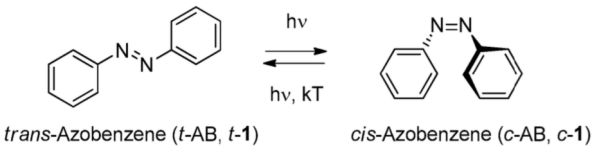


Figure 2: Isomerization reaction of Azobenzene. The trans $\rightarrow$ cis isomerization only occurs photochemically, the cis $\rightarrow$ trans occur both photochemically and thermally[10].

X-ray and computational data indicate that trans-AB has a $C_{2h}$ symmetry with a planar structure and a zero dipole moment[18, 19], while the cis-AB has a $C_2$ non planar conformation and a dipole moment of 3.0D[20, 21]. The *trans*-AB absorption spectrum has two well separated bands in the UV-vis region. The bright state (S2) consist of a $\pi \rightarrow \pi^*$ transition at $\lambda_{max} \sim 320$nm and a much weaker dark state (S1) arising from the symmetry forbidden $n \rightarrow \pi*$ absorption at $\lambda_{max} \sim 450$nm. The $\pi \rightarrow \pi^*$ transition of *cis*-AB absorbs at $\lambda_{max} \sim 270$nm while the $n \rightarrow \pi*$ absorbs also at $\lambda_{max} \sim 450$nm [22, 23].

The mechanism of photoisomerization is still under large debate and four mechanisms, namely, rotation, inversion, concerted inversion and inversion assisted rotation have been proposed as possible pathways, shown in Figure 3.

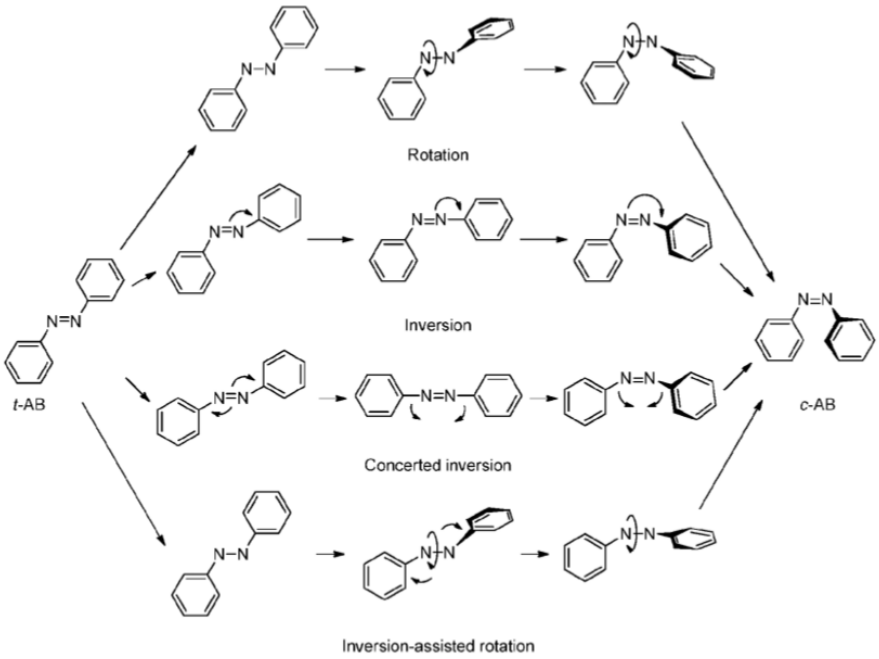


Figure 3: Isomerization mechanisms of trans-Azobenzene

The rotational pathway involve the cleavage of the N=N $\pi$ bond to allow free rotation changing the C–N=N–C dihedral angle while the N–N–C angle remains fixed at $\sim 120°$. In the inversion mechanism, one N=N–C angle increases to 180° while the dihedral C–=N–C remains fixed, resulting in a transition state with one sp hybridized azo-nitrogen atom. Concerted inversion has both N=N–C bond angles increasing to 180° generating a linear transition state, while in inversion-assisted rotation, large changes in the C–N=N–C and changes in N=N–C angles occur simultaneously. Multiple isomerization pathways have been invoked to explain experimental observations[17]. It was first thought that the excitation to $S_2$ would cleave the $\pi$ bond allowing rotational isomerization, but recent studies of ultrafast time-resolved fluorescence experiments have denied the existence of a rotational pathway[22]. Ultrafast, time resolved Raman[24] and fluorescence[25] indicated that a planar $S_1$ state is generated from $S_2$ relaxation, and excluded the occurrence of rotation in the $S_2$ state. The N=N stretching frequency in the $S_1$ state corresponds closely to the ground state ( 1428cm$^{-1}$ vs 1440cm$^{-1}$ ),indicating that the $\pi$ bond remains intact after the $S_2 \leftarrow S_0$ excitation. Several researches have found an $S_1$–$S_0$ conical intersection along the rotation pathway and it is generally agreed that, when excited to the $S_1$ state, relaxation occurs through the conical intersection along the midpoint of the rotation pathway[26, 27, 28].

There are several factors affecting the quantum yield of isomerization. Both *trans* $\rightarrow$ *cis* and *cis* $\rightarrow$ *trans* photoisomerizations occur with the highest quantum yields following $S_1 \leftarrow S_0$ excitation[29]. The $S_2$ state relaxes to the $S_1$ state with a quantum yield of approximately 1, but the $S_1$ state prepared by direct excitation has a longer lifetime than that obtained by the $S_1 \leftarrow S_2$ relaxation; the indirect excitation increases the vibrational energy of the $S_1$ state because additional deactivation pathways are open, decreasing the quantum yield[29]. The *trans* $\rightarrow$ *cis* isomerization quantum

yield following the $n \rightarrow \pi^*$ ($S_2$) excitation increases with increasing solvent polarity, while it does not affect photoisomerization quantum yield of $\pi \rightarrow \pi^*$[**?** ].

## 1.3 P-diaminoazobenzene, an Azobenzene Derivative

Substitutions among the aromatic rings of AB can mediate their optical properties and isomerization behaviour. The AB substitutions have been divided in three categories based on the relative energies gap between the $\pi \rightarrow \pi^*$ and the $n \rightarrow \pi^*$ electronic states : azobenzenes (AB), aminoazobenzenes (aAB) and pseudostilbenes (pAB), but only the aAB category will be discussed.

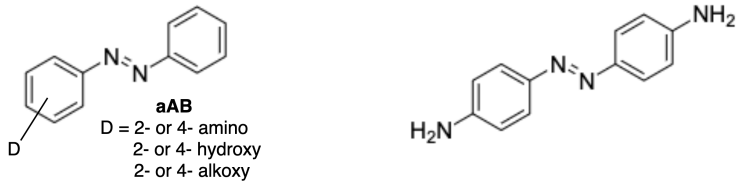


Figure 4: General structure of aABs and p-diaminoazobenzene. In the presence of electron donating substituents, the $\pi \rightarrow \pi^*$ transition is shifted to longer wavelegths leading to overlap between the two transitions.

The ABs with one or more amino or hydroxy substituents in the 2 or 4 position are classified as aABs. In this class, the presence of electron donating substituents cause a shift of the $\pi - \pi^*$ transitions to longer wavelengths, creating an overlap with the n $\rightarrow \pi^*$[16]. The aABs exhibit high quantum yields of isomerization when excited to the S1 state as in ABs and due to the smaller energy gap between the $S_1$ and $S_2$ states, it exhibit a higher quantum yield when excited to the S2 state. The lower energy $S_2$ state, relaxes to the $S_1$ state which possess less vibrational energy and does not access deactivation pathways, leading to an higher quantum yield of isomerization[30]. Experimental studies on para-disubstituted aABs indicate that the addition of substituents to the AB chromophore can affect the optical properties and the isomerization surfaces. The di-para amino substitution, highly affect the bond order of the compound and the isomeriazation barrier is reduced influencing the S2 state. Steady-state absorption spectroscopy have shown a remarkable change in the electronic excited states, where the dominant absorption transition ($S_2 \leftarrow S_0$) is redshifted by $\sim 7000\text{cm}^{-1}$ upon addition of p-amino substituents[31]. Theoretical studies with time dependent density functional theory (TD-DFT) at B3LYP/6-31G* level of theory have shown that singlet vertical excitation energies for the weak $S_1 \leftarrow S_0$ transition were similar for all ABs and derivatives. By molecular orbital (MO) inspection, it was shown that for both ABs and aABs the transition originates from the lone pair on the central nitrogens. The bright state transition, $S_2 \leftarrow S_0$ ($\pi \rightarrow \pi^*$), for AB has a maximum of 3.77 eV slightly lower than the steady-state absorption maximum of 3.96 eV. For p-diaminoazobenzene (aAB), the calculated energy was 3.26 eV , slightly higher than the steady-state absorption maximum of 3.15 eV, but confirming the experimental results that electron donating groups causes a red-shift in the $\pi \rightarrow \pi^*$ absorption[32]. For such reasons, p-diaminoazobenzene, an aAB derivative was selected for the research, since it exhibit a high quantum yield of isomerization when excited to both $S_1$ and $S_2$ states and absorbs

light at a wavelength compatible with biological systems where light can penetrate the tissues with no side effects.

## 1.4 Ion Channels and Structure of the Human's Brain $Na_v1.4$ Channel

Ion channels are transmembrane proteins capable of controlling the movement of ions and small molecules between the extracellular environment and the cytoplasm[33]. They are a collection of protein domains that together generate a water filled pore which allow the passage of ions through a cell membrane in response to chemical stimuli, temperature changes or mechanical forces. The gating mechanism control the opening and closure of the pore through conformational changes of the protein domains, combined with a sensor mechanism that detects and respond to stimuli. Voltage gated ion channels are a class of channels which respond to changes in membrane potential with conformational changes, which leads to opening and closing of selective pores. Voltage gated channels selective for sodium ($Na_v$), potassium ($K_v$) are the major contributors to axonal electrical transmission which field play a key role for many physiological processes including neural transmission, muscle contraction, cognition, regulation of blood pressure and cell proliferation[34].

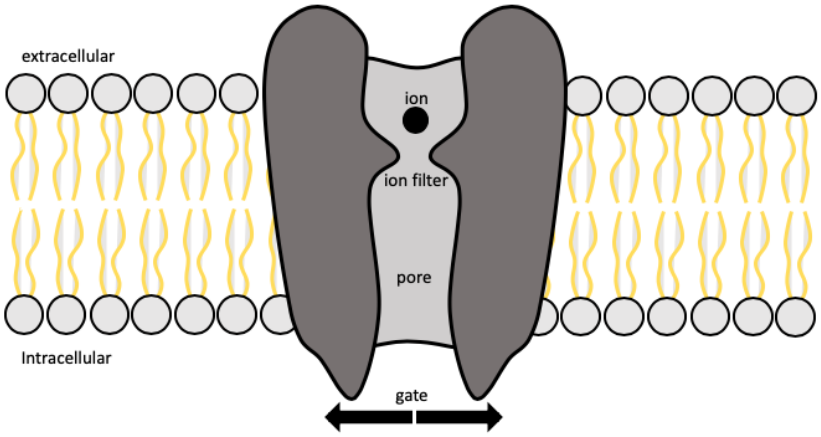


Figure 5: Graphical representation of an ion channel in a lipid membrane. The black dot represents an ion approaching the selectivity filter.

Sodium channels plays a central role in physiology. They transmit depolarizing impulses rapidly through cells and cell networks enabling coordination of higher processes from locomotion to cognition. In 2018, the crystal structure of a human's brain voltage gated $Na_v1.4$ channel was determined via cryo-electron microscopy (c-EM) at 3.2 Å resolution[35]. In "$Na_v1.4$" channel the subscript "$v$" indicates the physiological regulator (voltage), the first number indicates the gene subfamily and the following number indicates the order in which the gene was identified. The $Na_v1.4$ works as action potential initiator and transmitter in skeletal muscles and it is a target of local anesthetics used to treat myotonia[36].

The structure consists of an $\alpha$ subunit associated with auxiliaries $\beta$ which differ in number depending on the function. The $\alpha$ is organized in four homologous domains (I-IV), each containing six transmembrane $\alpha$ helices (S1-S6) and an additional pore loop located between the S5 and

S6 segments. In each of the four domains of the $\alpha$ subunit, the S1-S4 $\alpha$ helices constitute the voltage-sensing domain (VSDs) while the four S5 and S6 helices form the pore domain. The VSDs undergo a conformational alteration in response to membrane potential changes leading to the opening and closing of the pore domain. The S4 segment contains positively charged amino acid residues Arg/Lys which work as gating charges and move across the membrane to initiate channel activation in response to depolarization of the membrane[37]. At resting potential, the Arg/Lys residues are attracted to the inner cytoplasmatic side and the pore domain is closed and upon membrane depolarization, the outward movement of the Arg/Lys residues, results in pore opening. Figure 6 shows a 2D representation of the organization of sodium channel subunits.

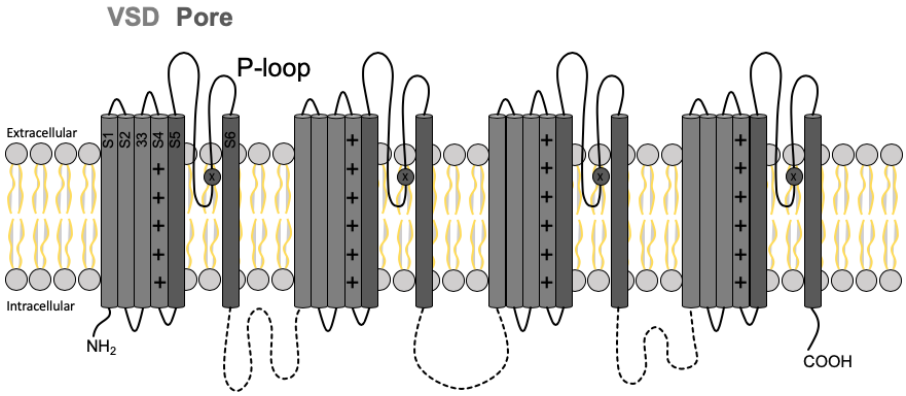


Figure 6: 2D representation of transmembrane organization of sodium channel subunits. S1-S4 represents the VSD domain while the S5-S6 represents the pore domain. In between the S5 and S6 subunits, the selectivity filter is shown, represented by a circle.

Located on the first turn of the P-loop, between the S5 and S6 subunits, the ion permeation filter is found, responsible for the selection and interactions of the passing ions. It is composed by $Asp^{406}/Glu^{761}/Lys^{1244}/Ala^{1536}$ (DEKA) and was resolved with electron microscopy reconstruction with a local resolution of 2.8 Å. The distance between the amine of K ($Lys^{1244}$ on subunit III and the opposing carboxyl oxygen of D ($Asp^{406}$) on subunit I is $\sim$ 3.5 Å which is the point of ion permeation. Figure 7 below shows a 3D representation of four $\alpha$ subunits of the $Na_v1.4$ channel.

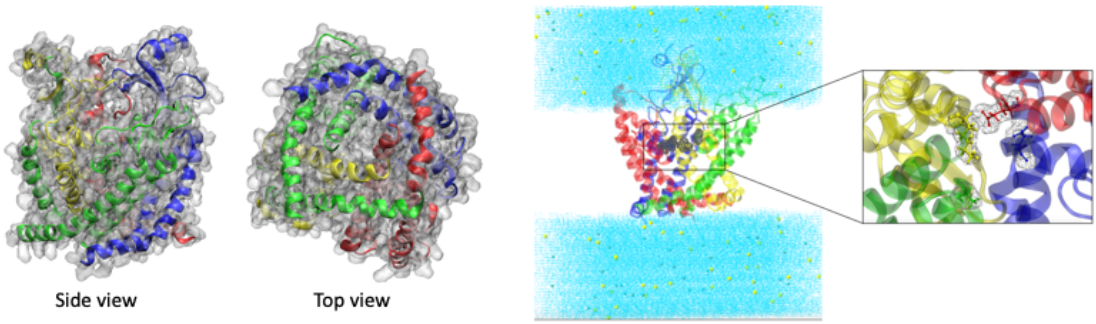


Figure 7: Left side : side and top view with New Cartoon of the four domains of $Na_v1.4$ ion channel. Right side : New Cartoon representation of the solvated $Na_v1.4$ showing the residues representing the DEKA filter.

## 1.5 Photoswitches in Ion Channels

More than 60 channelopathies have been identified as human deseases that are brought through mutation of ion channels[38]. Malfunction of those channels could lead to disastrous pathological consequences and, given this central role in human physiology, ion channels became an attractive target for drug discovery. Many strategies have been employed to optically control electrical signals and manipulate the conductivity of ion channels with high spatial and temporal precision. In the last decade, the field of photopharmacology has boomed, with the development of new methods that directly tag protein domains with light-sensitive proteins or synthetic chromophores. Between the chemical approaches, two main techniques have been developed for the control of receptors[39]. The first, caged compounds approach, contains a removable protecting group that upon irradiation is cleaved, resulting in a rapid release of the active compound. Those compounds are biologically inactive before photolysis, whereas after light stimulation they become active. In the second more recent approach, photoswitchable ligands are used, containing a photoisomerizing group that, unlike caged compounds have the advantage to reversibly alternate between an active and an inactive form with exposure to two different wavelengths of light.

A family of azobenzene's photosensitive ligands have been successfully used to manipulate the activity of voltage gated channels. P-diquaternary ammonium azobenzene (QAQ) has been used to resemble the activity of local anesthetics (lidocaine) in nociceptive neural receptors. The photoswitch reversibly suppresses neuronal excitability by regulation of voltage gated $Na^+$, $Ca^{2+}$ and $K^+$ channels. After it accumulates intracellularly, it blocks the voltage-gated channels in the trans form but not in the cis form. In *vivo* studies in rats have shown that QAQ enables reversible optical silencing of mouse nociceptive neurons, serving as a light-sensitive analgesic[13]. Another azobenzene derivative, acrylamide-azobenzene-quaternary ammonium, was found to target 4 voltage gated ion channels by acting on the same binding site of a pore blocking agent, tetraethylammonium[14]. The blocking agent binds intracellularly leading to a fast inactivation process[40].

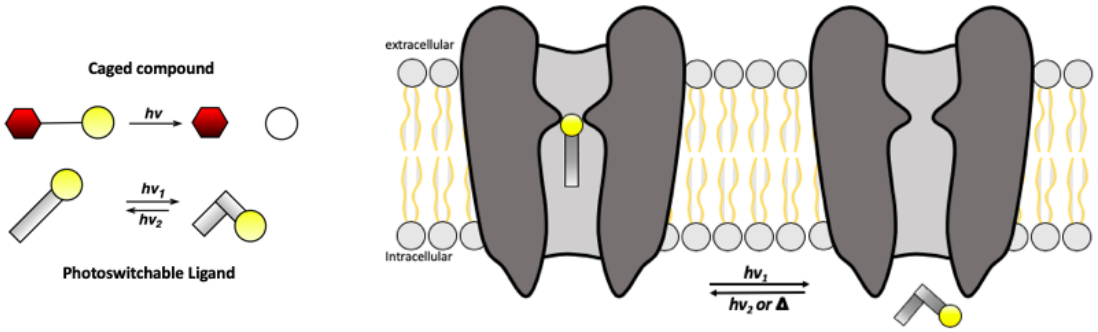


Figure 8: Left side : Representation of a caged compound activity and a photoswitchable ligand. Right side : photoswitchable blocker for voltage-gated potassium channels. The quaternary ammonium moiety is highlighted by an orange sphere. The blocker binds to the intracellular vestibule and blocks ion conduction in the trans configuration[14]

For the current study, a human's brain voltage gated sodium channel ($Na_v1.4$) will be under investigation.

# 2 Materials and Methods

## 2.1 Molecular Dynamics

Computation based on molecular models is playing an essential role in biology, biological chemistry and biophysics. The continuing growth of computing power has made it possible to analyze, compare and characterize large and complex data sets obtained from experiments on biomolecular systems. To begin a biomolecular modelling study of a particular system the spatial resolution, timescale, and number of degrees of freedom (DoF) must be examined. Those three factors determine the level of modelling at which a particular process can be described. In atomistic molecular dynamics, mainly atomic and molecular degrees of freedom are considered with the corresponding classical force field and a classical Newtonian dynamics to sample the degrees of freedom. In this branch of modelling, a system size can range up to $10^5$ - $10^6$ atoms, which is still a small quantity compared to Avogadro's number.

| **Methods** | **Degrees of freedom** | **Properties, Processes** | **Timescale** |
|---|---|---|---|
| quantum dynamics | atoms, nuclei, electrons | excited states, relaxation, reaction dynamics | picoseconds |
| quantum mechanics (ab initio, DFT, semiempirical, valence bond methods) | atoms, nuclei, electrons | ground and excited states, reaction mechanisms | no timescale |
| classical statistical mechanics (MD, MC, force fields) | atoms, solvent | ensembles, averages, system properties, protein-ligand binding | microseconds |
| statistical methods (database analysis) | groups of atoms, amino acid residues, bases | structural homology and similarity | no timescale |
| continuum methods (hydrodynamics and electrostatics) | electrical continuum, velocity continuum | rheological properties | supramolecular |
| kinetic equations | populations of species | population analysis, signal trasduction | macroscopic |

To treat such small systems, the simulation takes place in box surrounded by infinite replicas of itself, a periodic boundary condition (PBC). The presence of a boundary would have a large effect due to an unrealistic amount of atoms near the surface of the box, changing the properties of the system[41]. Figure 9 shows the four main components of molecular dynamics.

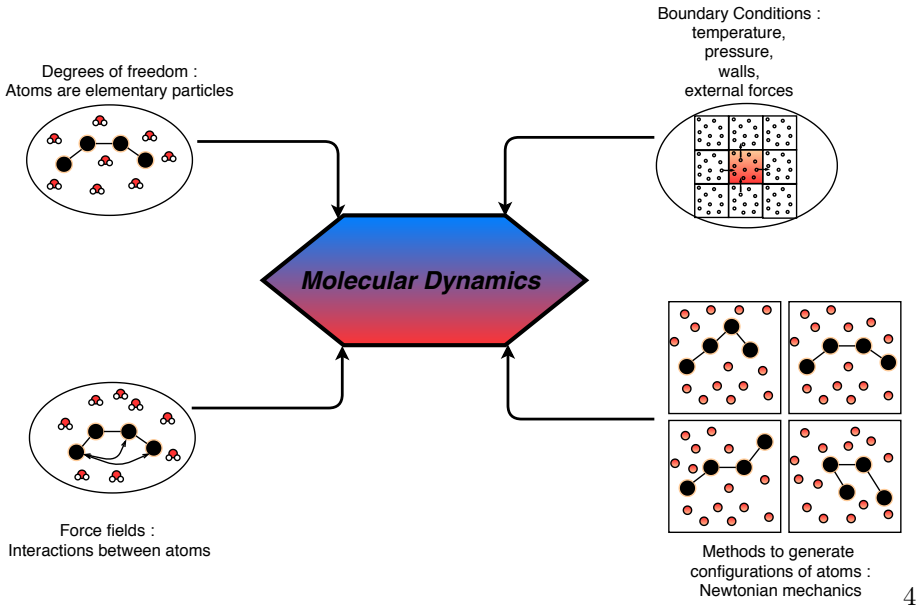


Figure 9: The four main components of molecular dynamics : degrees of freedom (upper left), boundary conditions (upper right), force fields (bottom left) and integrator (bottom right)

### 2.1.1 A Deterministic Approach

The system under investigation can be propagated by either deterministic or probabilistic approaches to generate a trajectory which describes the coordinates evolution in time. In the simulation, several snapshots/frames containing the coordinates of the system are obtained and averaged over the trajectory to compute estimates of the desired properties. Molecular dynamics compute the movements of atoms along time by integration of Newton's equations of motion,

$$\frac{d^2x_i(t)}{dt^2} = \frac{F_i(t)}{m_i} \tag{1}$$

with $F_i(t)$ being the force exerted on atom $i$ at time t, $r_i(t)$ the vector position of the atom $i$ at time $t$ and $m_i$ the mass of the atom[42]. The force $F_i$ can be written as,

$$F_i = -\frac{dV(x^N)}{dx_i} \tag{2}$$

which is the gradient of the potential $V(x^N)$. Since there is no analytical solution to equations of motion for more than three interacting bodies, the dynamics must be approximated numerically,

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2x(t)}{dt^2}\Delta t^2 + ... \tag{3}$$

accounting for the truncation of the Taylor expansion after the second term. The velocity-Verlet algorithm enable to find the trajectory of an object influenced by a force field by considering the expansion of coordinates forward and backward in time. It requires the atomic positions and accelerations at time $t$ and the positions from the prior step, $x(t - \Delta t)$, to determine the new

position at $t + \Delta t$[43].

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2x(t)}{dt^2}\Delta t^2 \tag{4}$$

$$x(t - \Delta t) = x(t) - \frac{dx(t)}{dt}\Delta t + \frac{1}{2}\frac{d^2x(t)}{dt^2}\Delta t^2 \tag{5}$$

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2}\Delta t^2 \tag{6}$$

where eq.(4) is a step forward in time and eq.(5) is a step backward in time. A popular algorithm based on the Verlet's equation is the leapfrog algorithm which uses the positions at time $t$ and the velocities at time $t - (\Delta t/2)$ to update both position and velocities via the calculated forces, $F(t)$ as show in eq.(7) and (8).

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt}\left(t + \frac{\Delta t}{2}\right)\Delta t \tag{7}$$

$$\frac{dx(t)}{dt}\left(t + \frac{\Delta t}{2}\right) = \frac{dx(t)}{dt}\left(t - \frac{\Delta t}{2}\right) + \frac{d^2x(t)}{dt^2}\Delta t \tag{8}$$

The computational cost of an integrator is crucial and it is beneficial to limit the number of force calculations during the simulation without deviating from the path of an analytical trajectory. Numerous algorithmic improvements can be applied to increase the efficiency of simulations. The SHAKE or constrained Verlet method, freezes the fastest modes of vibration by constraining the hydrogen atoms bonds to an equilibrium length[44]. This approximation does not affect the motion of large biomolecular systems[45]. One important note that classical MD using standard force fields uses harmonic potentials to describe bonding interactions and bond breaking and formations are not allowed, meaning that the topology of the chemical system will remain constant as a function of time[46].

### 2.1.2 The Force Field Problem

The motion along the degrees of freedom is governed by a potential energy. The computational effort is strongly dependent on the functional form of the Hamiltonian used to describe the energy. Classical molecular ynamics (MD) focuses on a classical Hamiltonian where the total energy of the system in a given configuration is split into a kinetic term *K(p)* and a potential energy term *V(x)*,

$$H(p, x) = K(p) + V(x) \tag{9}$$

The kinetic term has the form,

$$K(p) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} \tag{10}$$

with $p$ being the momentum of a particle and $N$, the total number of atoms in the system[47]. In classical MD, the potential energy term is generally referred to as "force field". A biomolecular force field consists of energy terms representing the covalent interaction between atoms such as bond-stretching, bond-angle bending, improper and proper dihedral-angle torsion and non bonded

interactions between atoms in different molecules and in a molecule separated by 3 or more covalent bonds,

$$V(x)^{total} = V^{bond} + V^{angle} + V^{dihedral} + V^{vwW} + V^{el} \tag{11}$$

A problem in the development of classical force fields is the extremely large variety of chemical compounds for which compatible force fields should be derived. By parameterizing the force fields on a physical basis and keeping them simple and local makes possible to transfer them from one compound to another[48]. The general amber force field has the following form,

$$V(x)^{total} = \sum_{bonds} K_b(b - b_{eq})^2 + \sum_{angles} K_\Theta(\Theta - \Theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + cos(n\phi - \gamma)] + 4\varepsilon_{i,j} \sum_{i<j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] + \frac{1}{4\pi\epsilon_0} \sum_{i<j} \left[ \frac{q_i q_j}{r_{ij}} \right] \tag{12}$$

The first three terms represent the bonded interactions, where $b_{eq}$ and $\Theta_{eq}$ are equilibrium bonds and angles; $K_b$, $K_q$, $V_n$ are force constants; n is the periodicity of the dihedral and $\gamma$ is phase angle for torsional angle parameters. The fourth terms represent the non bonded van der Walls interactions, where $r_{i,j}$ is the distance separating a pair of atoms $ij$, $\varepsilon_{i,j}$ is the depth of the potential well for the interaction of atoms $i$ and $j$, and $\sigma_{i,j}$ is the distance where the potential is exactly zero and the repulsion starts for the two atoms. Both $\varepsilon_{i,j}$ and $\sigma_{i,j}$ are specific parameters for pairs of atoms. The final term represents the electrostatic Coulomb interactions between a pair of atoms with point charges $q_i$ and $q_j$ at distance $r_{ij}$ and with $\epsilon_0$ being vacuum permittivity. The 12-6 Lennard-Jones and Coulombic terms are used to calculated the electrostatic and van der Walls nonbonded interactions, considering all pairs of atoms (i and j) in different molecules or in the same molecule but separated by a minimum of three bonds. One-four nonbonded interactions separated by three bonds are reduced by a scale factor, in order to use identical parameters for intra- and intermolecular interactions. Parameters for the bond lengths, angles and dihedral angles are often derived from quantum mechanical calculations or crystal structures. In AMBER, the charges to reproduce the electrostatic potential are obtained from quantum chemical calculations[49]. 11 Electrostatic interactions play a major role in biomolecular modelling, compared to covalent and van der Walls interactions. Even the interaction between two neutral molecules without dipole moments, but with a quadrupole moments is longer ranged than the van der Walls dispersion interaction, proportional to $r^{-6}$. Considering the electrostatic energy, to integrate the interaction $V^{el}(r)4\pi r^2$ from $r$ to infinity,

$$\int_0^\infty V^{el}(x)4\pi r^2 dr \tag{13}$$

the integral converges under the conditions,

$$V^{el}(x) \sim x^{-n} \qquad\qquad n > 3 \tag{14}$$

therefore the total electrostatic energy is strictly dependent on the boundary conditions surrounding the system[41]. Non bonded require a large computational power since the strength and

quantity of those interactions varies substantially as the simulation proceed, unlike the number of bonded interactions which remains unchanged during a MD simulation. To evaluate long-range electrostatic interactions, a technique which provides one way to efficiently handle long-range electrostatics is the lattice-sum method. In this method, the system is placed in a box and surrounded by an infinite replicas of itself to avoid the boundary problem. The basic idea of the particle-mesh-Ewald scheme is to replaced the direct summation of interaction energies between point charges into two summations,

$$E^{el} = E_{sr} + E_{lr} \tag{15}$$

$$E_{sr} = \sum_{i,j} \phi(x_j - x_i) \tag{16}$$

$$E_{lr} = \sum_{k} \Phi_{lr}(k)|\sigma(k)|^2 \tag{17}$$

where $\Phi_{lr}$(k) represent the Fourier transform of the potential and $\sigma$(k) the charge density. Since both summations converge quickly in their spaces, they can be truncated with a little loss of accuracy and a large improvement in computational efficency[50]. There is still room for large improvements in the parametrization of force fields.

### 2.1.3 The Thermodynamic Problem

While running a molecular dynamics simulation, in addition to the choices of algorithms to integrate the equation of motion and an appropriate force field, numerous additional factors must be considered. The boundary conditions refer to both the physical boundaries of the system but also to the way the state variables are controlled. The most common option to treat the edges of the simulated system is the PBC, which removes hard edges by virtually replicating the box in every direction as shown in Figure 10,

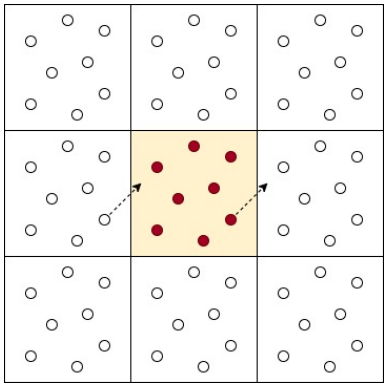

Figure 10: Representation of a PBC. The red particle travels through the wall on the right side and its replica (white particle) enters from the opposite side.

The particle exiting the box from one side, enters the from the opposite side. However, PBCs do not fully eliminate the effects of a finite box size and can still influence properties, such as

diffusion, bilayer modulation and lipid domain sized and dynamics[51]. As in real experiments, there is a choice under which thermodynamic conditions to perform the *in silico* experiment. In general, properties of the system such as pressure and heat capacity will depend upon the position and momenta of the N particles that compose the system. The instantaneous value of the property $A$ can then be written as $A(p^N(t)x^N(t))$. Over time, the instantaneous value of the property fluctuates as a result of interactions between particles, thus the experimental measurement is an average of $A$ over the time of the measurement, known as *time average*. At infinite time,

$$A_{ave} = \lim_{\tau \to +\infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), x^N(t)) dt \tag{18}$$

the average value of $A$ approaches the "true" average value of the property. For this reason, to calculate the values of the properties of the system, it seems to be necessary to simulate a dynamic behavior. The difficulty is that for "macroscopic" numbers ($\sim 10^{23}$ atoms) it is not feasible to calculate the trajectory. The development of statistical mechanics by Boltzmann and Gibbs aim to solve this problem by replacing the time average by an ensemble average :

$$\langle A_{ave} \rangle = \int \int dp^N dx^N A(p^N, x^N) \rho(p^N, x^N) \tag{19}$$

where the angle brackets $\langle\ \rangle$ indicate an ensemble average of the property $A$ over all replications of the ensemble generated by the simulation. In accordance to the *ergodic hypothesis*, the fundamental axiom of statistical mechanics, the ensemble average is equal to the time average at infinite time[52, 53].

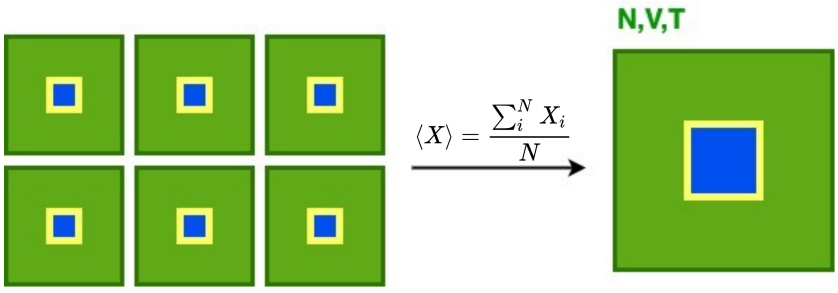


Figure 11: Representation of a canonical ensemble. The blue part is the simulated system, the yellow part are the rigid, impermeable, diathermic walls and the green part is the thermostatic bath. The macroscopic properties are obtained by averaging over all the microstates in the ensemble

Molecular dynamics is conventionally performed under conditions of constant number of particles (N), volume(V) and energy (E), i.e., microcanonical ensemble. Other ensembles are the isobaric-isoenthalpic ensemble (NPH) with fixed number of particles N, pressure P and enthalpy H; the isobaric-isothermal ensemble (NPT) with fixed number of particles N, pressure P and temperature T; the grand canonical ensemble ($\mu$VT) with fixed chemical potential $\mu$, volume V and temperature T; and finally the canonical ensemble NVT with fixed number of particles N, volume V and temperature T. However real life experiments are usually performed at constant temperature and volume (NVT), or constant pressure and temperature (NPT), hence those conditions are

simulated by the addition of external factors, thermostats and barostats.
From the equipartition theorem, the temperature is defined as,

$$\frac{3}{2}Nk_bT = \langle \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \rangle \tag{20}$$

which is a time averaged quantity[52]. In a single MD simulation frame, instead of a time average, the temperature is defined as instantaneous temperaure. In a canonical ensemble, the total temperature is constant, but in a microcanonical ensemble the instantaneous temperature will not always be equal to the target temperature but will undergo fluctuation in its value.

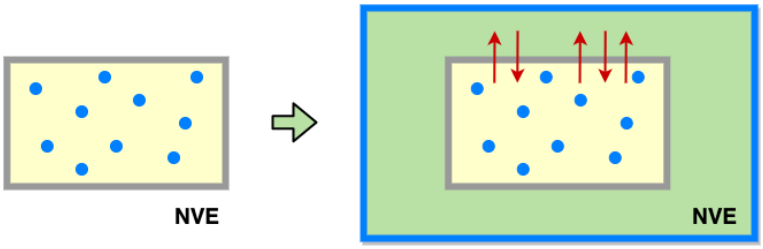


Figure 12: Representation of a thermal bath (green) surrounding the simulated system inside of the box with grey walls. A temperature can be imposed by bringing the system into thermal contact with a large heat bath, at fixed temperature T. The velocities are scaled to match the bath temperature

The thermostat algorithms controls those fluctuations by altering the Newtonian's equations of motion which have constant energy (microcanonical ensemble)[48]. It can either be deterministic or stochastic and also either global or local depending on whether it is coupled to the dynamics of the full system or a small subset[46]. For example, the Langevin thermostat enrich the microcanonical equations of motion with Brownian dynamics, including the viscosity and random collision effect of an implicit solvent[54]. It uses an equation of the form,

$$F = F_{interaction} + F_{friction} + F_{random} \tag{21}$$

where $F_{interaction}$ is the standard interactions calculated during the simulation, $F_{friction}$ is the damping used to tune the "viscosity" of the implicit bath and $F_{random}$ which gives random collisions with solvent molecules.
Barostat algorithms, similarly control the pressure by coupling the system to a weakly interacting pressure bath. In MD simulations, the pressure is calculated via the virial equation of state which allows to calculate non-ideal gasses properties,

$$\frac{PV_m}{RT} = 1 + \frac{B_{2v}(T)}{V_m} + \frac{B_{3v}(T)}{V_m^2} + ... \tag{22}$$

where $V_m$ is the molar volume, $\frac{B_{2v}(T)}{V_m}$ and $\frac{B_{3v}(T)}{V_m^2}$ are the second and third virial coefficients which give a relation between the molar volume, pressure and temperature for a non-ideal gas. For an

ideal gas the expression is truncated at the first term,

$$\frac{PV_m}{RT} = 1 \tag{23}$$

but for non ideal gasses, the value can be between 0 and $\infty$. Any non zero value of the virial coefficients indicates the degree of interactions between particles which depend on the potential energy of interaction, $V(x)$. In a MD simulation, the pressure can be controlled by changing the volume of the simulation box. The Berendsen (weak coupling) barostat, for example, scales the volume of the box by a scaling factor which produces more realistic fluctuations in the pressure as it slowly approaches the target pressure[55]. Constant pressure can also be achieved stochastically performing Monte Carlo (MC) moves that adjust the system volume. The MC barostats are computationally beneficial since the *virial* is not required. This method well describes the distribution of volumes in the NPT ensemble[56].

### 2.1.4 The Search Problem

Biomolecular systems are characterized by a large number of degrees of freedom, as well as a wide range of time and spatial scales, from femtoseconds and tenths of nanometers to milliseconds and micrometers. Thus the phase space representing such a system is a very rough surface with a multitude of mountains and depths. This aspect makes the search for global minima rather difficult if not impossible[41]. In fact, the state of a biomolecular system is not described by a single global minimum but by a statistical ensemble of configurations in which the weight of the configuration is ruled by the Boltzmann factor,

$$\rho(r) = exp(-V(x)/k_bT). \tag{24}$$

As mentioned above, in a molecular dynamics simulation, the ensemble average $\langle \mathrm{A} \rangle$ is calculated by the time average

$$A_{ave} = \lim_{\tau \to +\infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), x^N(t))dt \tag{25}$$

assuming that,

$$A_{ave} = \langle A \rangle \tag{26}$$

Most systems cannot be proved to have ergodic behaviour, but if during a simulation, all the configurations for which $V(r)$ is low are visited, eq.(26) will hold[57]. The problem arises when regions of high probability are separated by large energy barriers and it will be unlikely that all significant configurations will be sampled. This is known as *problem of quasi-nonergodicity*. With the inability to visit all important conformations, different methods have been developed to speed up the sampling of the phase space without loosing the properties of the system[58]. These techniques can be classified in, i) deforming the potential energy surface to increase of surmounting free energy barriers; ii) extending the dimensionality to circumvent energy barriers; iii) perturbing the forces iiii) multi-copy approaches;

Within the strategy of direct deformation of the potential energy surface (PES), an approach

is "accelerated molecualar dynamics" (aMD) in which energy wells are elevated to enable more frequent barrier crossing. This allows a search for the PES minima in an optimized computation time and reproducing the canonical probability distribution after reweighting the statistic of the aMD method. For the purpose of the binding pockets search in the $Na_v1.4$ channel, an aMD method, Gaussian accelerated molecular dynamics (GaMD) will be used and compared to the sampling of classical dynamics to investigate the sampling efficiency of both methods.

### 2.1.5 Gaussian Accelerated Molecular Dynamics

An accelerated molecular dynamics approach modifies the potential energy surface, *V(x)*, by adding a bias potential, $\Delta V(x)$, to the true potential such that the surface near the minima is raised and the region near a barrier or a saddle point is left unaffected as shown in figure 13.

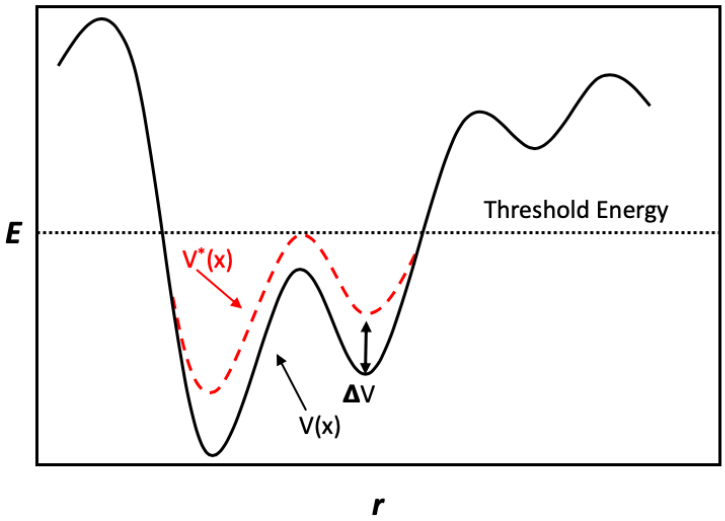


Figure 13: Graphical representation of a general boosted potential. The potential is boosted below an energy threshold (E) and keeps its original value above it.

The boost potential, $\Delta V(\mathrm{x})$, decreases the energy barriers and accelerate transitions between different low-energy states, sampling conformations which are not accessible with standard molecular dynamics. Because aMD does not require any predefined collective variables, it can sample molecular conformations without a priori knowledge or constraints[59]. However aMD is known to suffer from large energetic noise during reweighting rising problems to recover the classical free energy landscape[60]. While aMD boost potential can be sufficiently low for proper reweighting of simulations on small systems such as alanine dipeptide and fast folding proteins, it is on the order of hundreds of kcal/mol in simulations of normal proteins and it does not allow for a precise reweighting to perform free energy calculations. In a recent study, it was shown that when the boost potential follows near-Gaussian distribution, there is an improvement over the aMD reweighting and a reduction in energetic noise[61]. GaMD has been then developed to achieve both unconstrained enhanced sampling and a proper energy reweighting[62]. The method enhances the conformational sampling by adding harmonic functions to create a boost potential. Considering a system with N atoms at positions $\vec{x} = (\vec{x_1}, ..., \vec{x_N})$, when the system's potential $\mathrm{V}(\vec{x})$

is lower than a threshold energy (E), a boost potential is added as,

$$\Delta V(\vec{x}) = \frac{1}{2}k(E - V(\vec{x}))^2, \qquad V(\vec{x}) < E \tag{27}$$

where k is the harmonic force constant. The modified potential instead is given by,

$$V^*(\vec{x}) = V(\vec{x}) + \frac{1}{2}k(E - V(\vec{x}))^2, \qquad V(\vec{x}) < E \tag{28}$$

and if the potential is above the threshold energy, $V(\vec{x}) > E$, the boost potential is set to zero giving $V^*(\vec{x}) = V(\vec{x})$. There are two criteria that must be fulfilled; first, given two potential values $V_1(\vec{x})$ and $V_2(\vec{x})$ found on the original surface, if $V_1(\vec{x}) < V_2(\vec{x})$, $\Delta V$ should be a monotonic function that does not change the order of the values of the potential, therefore, $V_1^*(\vec{x}) < V_2^*(\vec{x}$. Then by replacing $V^*(\vec{x})$ with eq.(28) and isolating E, we can obtain

$$E < \frac{1}{2}[V_1(\vec{x}) + V_2(\vec{x})] + \frac{1}{k} \tag{29}$$

Secondly, if $V_1(\vec{x}) < V_2(\vec{x})$, the potential difference observed on the smoothened PES should be smaller than that of the original surface, so $V_2^*(\vec{x})$ - $V_1^*(\vec{x}) < V_2(\vec{x})$ - $V_1(\vec{x})$, and replacing again $V^*(\vec{x})$ with eq.(28), we can obtain

$$E > \frac{1}{2}[V_1(\vec{x}) + V_2(\vec{x})] \tag{30}$$

With $V_{min} \leq V_1(\vec{x}) < V_2(\vec{x}) \leq V_{max}$, the threshold energy E is set in the following range

$$V_{max} \leq E \leq V_{min} + \frac{1}{k} \tag{31}$$

where $V_{min}$ and $V_{max}$ are the system minimum and maxinum potential energies. To ensure the validity of eq.(31),

$$k \leq \frac{1}{V_{max} - V_{min}} \qquad k = k_0(\frac{1}{V_{max} - V_{min}}). \tag{32}$$

then $0 < k_0 \leq 1$. The magnitude of the applied boost potential is determined by $k_0$, where with an larger $k_0$ the boost is higher providing enhanced sampling by decreasing energy barriers[60]. Figure 15 shows the structure of the GaMD algorithm.

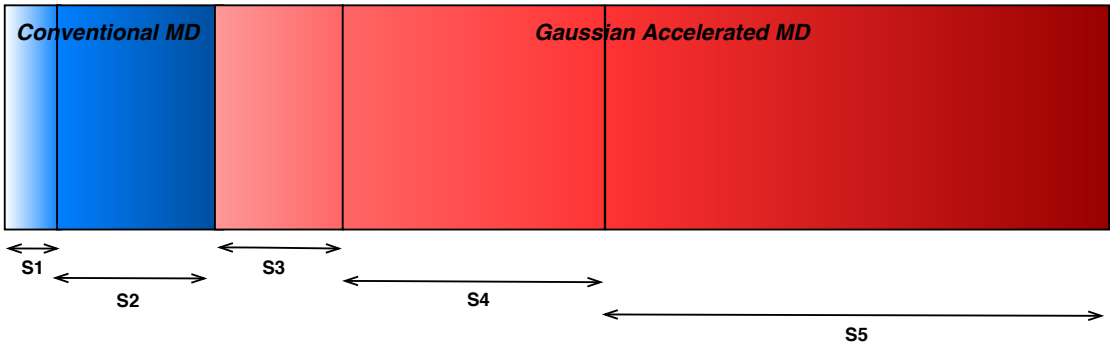


Figure 14: Graphical representation of the GaMD algorithm. The algorithm is divided in 5 stages. The first two colored in blue, S1 and S2 correspond to classical MD while from S3-S5 to GaMD

The algorithm is partitioned in five stages (S1-S2-S3-S4-S5) as shown in Figure 14. The S1 corresponds to a conventional MD (cMD) preparatory stage used to equilibriate the system and where no statistics of the potential are collected. In S2, the cMD potential statistics are collected, ($V_{max}$,$V_{min}$,$V_{avg}$ and $\sigma V$) and can be used to "activate" the GaMD algorithm. Following this cMD part, the GaMD part starts at S3, where a GaMD pre-equilibiriation using the potential statistics is performed, and the boost is applied but the boost parameters are kept fixed. In S4, the GaMD equilibriation starts and the boost parameters are updated until $k_0$ reaches its maximum value, meaning that the boost potential is maximum. The S5 is a GaMD production run, where the boost parameters are kept fixed from the final update of S4 and the boosted PES statistics are collected to reweight the energies at the end of the simulation[63].

### 2.1.6 The Free Energy Problem

In thermodynamics, the free energy is the amount of internal energy of the system which can be used to do work, and determines the direction of the thermodynamic process as well as the probability that the system will remain in a given state[64]. In biomolecular studies, the free energy determines all molecular processes such as protein folding, ligand-receptor binding, chemical reaction, etc., thus, its accurate description is essential to explain those processes. Experimental methods for thermodynamic properties are often expensive and time consuming, so theoretical calculations are becoming more and more important in many fields such as drug design[65]. The free energy $G$ in the canonical ensemble (NVT), is given by

$$G = -k_b T ln Q \tag{33}$$

with $k_b$ being the Boltzmann's constant, $T$ the temperature and $Q$ the partition function of the molecular system. In a thermodynamic equilibrium the partition function can be expressed as,

$$Q = \frac{1}{h^{3N} N!} \int \int e^{-k_b T H(p,x)} dp dx \tag{34}$$

where $N$ is the number of particles in the system, $h$ is the Planck constant and the factor $N!$ is for indistinguishable particles. The Hamiltonian $H(p, r)$ gives the total energy of the system in

a given set of momenta and coordinates, i.e., a configuration[47]. The absolute free energy can only be calculated for limited number of cases, primarily for small systems governed by a very simple Hamiltonian. For larger systems with strong interactions between particles, an analytical solution of the partition function is not achievable. An analytical solution could be obtained by omitting the interactions between particles as in ideal gasses or with the use of symmetry as in ideal crystals, but all the biological events occur in the liquid phase where it is rather difficult to obtain a reference state[57]. This is why in this area it is more convenient to use relative free energy between two states rather than the absolute free energy. A general statistical mechanical expression for the relative free energy is written as,

$$\Delta F_{BA} = F_B - F_A = -\beta^{-1} ln \frac{Q_B}{Q_A} = -\beta^{-1} ln \frac{\rho_B}{\rho_A} \tag{35}$$

where $\rho_A$ and $\rho_B$ correspond to different probabilities for the system to be in state A or B. In drug design, for example, the binding free energy is often used to quantify the binding strength of a ligand-receptor interaction[66]. The most widely used approach in protein-ligand binding related to drug design is molecular docking[67]. This technique generates a number of possible conformations/orientations (poses) of a ligand within the binding sites. It uses an deterministic or stochastic algorithm to sample the phase space and a scoring function to assess the binding energy of each pose. It is computationally efficient since the receptor is usually taken as a rigid body, i.e., large reduction of the number of DoF, but the scoring function are usually not of high accuracy and fail in comparing ligands with similar binding affinity (∼2kcal/mol). On the opposite side of the scale of accuracy, there are alchemical free energy methods also called pathway methods[68, 69]. Alchemy is an ancient branch of natural philosophy practiced through Europe, Asia and Africa where alchemists attempted to transmutate metals into gold or creating the elixir of immortality. Those methods acquired the name since are based on a coupling parameter $\lambda$ which leads, for example, a non-physical transformation between a compound and a related chemical species.

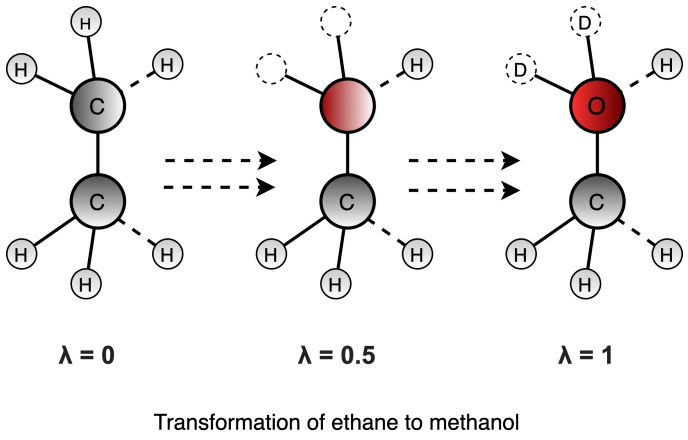


Figure 15: Alchemical transformation starting from ethane at $\lambda$=0 to methanol at $\lambda$=1

At each step of $\lambda$, an MD simulation is performed to adequately sample the conformations and provide overlap with the conformations available in the adjacent $\lambda$ values[70]. This requires a

huge computational cost which is not suitable if many different drugs needs to be compared. In between the two classes of methods, the end state method are located[71]. Those approaches are also based on sampling but only of the solvated complex (end-state) trajectory. The most popular end state method is the Molecular Mechanics Generalized Born Surface Area (MMGBSA), which has been widely used to rescore docking poses and predict binding affinities from MD ensemble simulations. In addition, it also allows to analyze contributions from individual residues by free energy decomposition analysis giving detailed energy contributions[72]. In MMGBSA the total binding free energy $\Delta G_{bind}$ is given by free energy difference of the complex (C), the ligand (L) and the receptor (R),

$$\Delta G_{bind} = G_C - G_R - G_L \tag{36}$$

or in an ensemble average form,

$$\Delta G_{bind} = \langle G_C \rangle - \langle G_R \rangle - \langle G_L \rangle \tag{37}$$

The averages should be estimated from three separate simulations, i.e., the simulation of the solvated complex, solvated ligand and solvated unbound receptor. This approach is denoted as three-average-MMGBSA(3A-MMGBSA) as in eq.(38)

$$\Delta G_{bind} = \langle G_C \rangle_C - \langle G_R \rangle_R - \langle G_L \rangle_L \tag{38}$$

But it is more commonly used to obtain the ensemble averages of the free receptor and ligand directly from the complex trajectory in a one-average-MMGBSA (1A-MMGBSA) fashion since it requires less computation time and it also improves precision[71], as in eq.(39)

$$\Delta G_{bind} = \langle G_C - G_R - G_L \rangle_C \tag{39}$$

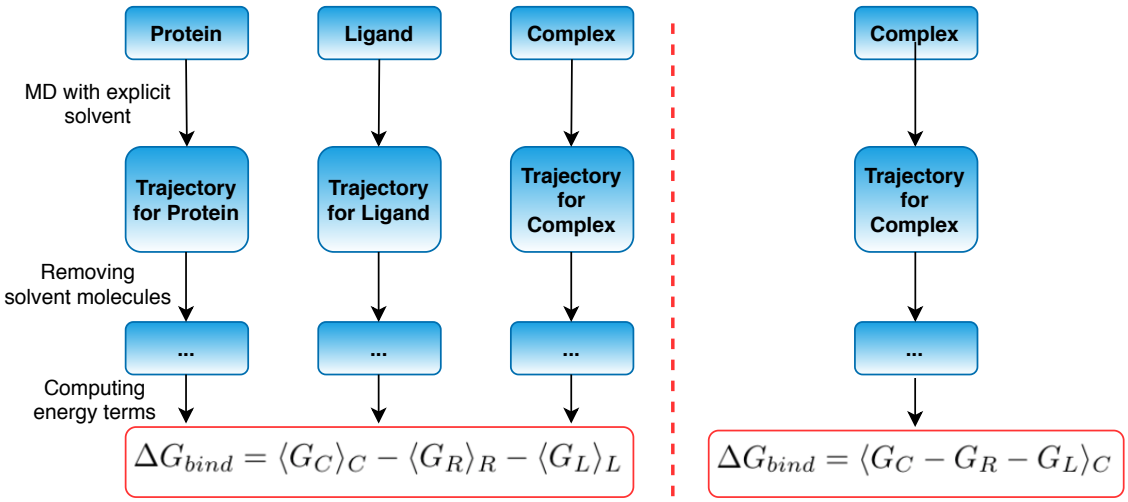


Figure 16: Graphical representation of 3A-MMGBSA (left) and 1A-MMGBSA (right). The 3A-MMGBSA requires an ensemble average obtained from three distinct simulations, i.e., solvated ligand, solvated receptor and solvated complex. The 1A-MMGBSA extracts the ensemble average for both solvated ligand, solvated receptor and solvated complex from one trajectory, i.e., solvated complex.

The free energy G of the complex,receptor and ligand is estimated by,

$$G = E_{bond} + E_{el} + E_{vdW} + G_{pol} + G_{np} - TS \tag{40}$$

where the first three terms are the standard MM force energy terms for bonded $E_{bond}$ (bond,angle,dihedral), $E_{el}$ electrostatic and $E_{vdW}$ van der Walls interactions. $G_{pol}$ is the polar contribution to the solvation free energy while $G_{np}$ is the non-polar contribution to the solvation free energy. The final term is the absolute temperature, $T$, multiplied by the entropy, $S$, which can be determined by a normal mode analysis of the vibrational frequencies on a set of conformational snapshots obtained from the MD simulations, but due to the heavy computational cost, this term is usually neglected[71]. It is assumed that the binding to the different pockets and the relative binding free energy of similar ligands will have similar entropy variation. In the 1A-MMGBSA, the $E_{bond}$ will be equal to zero in a one trajectory calculation,

$$E^{tot}_{bond} = E^{C}_{bond} - E^{R}_{bond} - E^{L}_{bond} = 0 \tag{41}$$

since the bonded free energy of the ligand and unbounded receptor summed with be equal to the complex's bonded free energy. Eq.(40) simplifies to

$$G = E_{el} + E_{vdW} + G_{pol} + G_{np} \tag{42}$$

and the total free energy as,

$$\Delta G_{bind} = \langle \Delta G_{el} \rangle_C + \langle \Delta G_{vdW} \rangle_C + \Delta G_{pol} + \Delta G_{np} \tag{43}$$

The electrostatic term, $E_{el}$ is calculated using Coulomb's law with atomic charges taken from the MM force field. It was shown that obtaining restrained electrostatic potential (RESP) charges from sophisticated quantum mechanical calculations does not substantially improve the results, but the ligand functional groups should be well parametrized by the force field to give relatively exact atomic charges[73]. The polar solvation term $G_{pol}$ represents the electostatic interactions between the solute and the continuum solvent and is calculated by solving analytically the Generalized Born implicit solvent model. Finally, the non-polar solvation energy $G_{np}$ accounts for cavitation, dispersion and repulsion energy between the solute and the solvent which are not included in the $G_{pol}$ term. This term is estimated with a linear relation to the solvent accessible surface area (SASA). The explicit solvent of the MD simulation is removed and replaced with the GBSA implicit solvent model to estimate the solvation energy.

## 2.2 Hybrid Quantum Mechanics/Molecular Mechanics and Electrostatic Embedding

Quantum mechanical events are manifested in a multitude of biological events such as electron rearrangements in biochemical reactions, coupled proton-electron transfer, charge transfer and photoexcitations, which are the base of fundamental biological processes. Most biological systems are too large to be described at any level of *ab initio* theory. The size ranges from $10^4$ to $10^5$ atoms and some biomolecular processes can take place on a timescale longer than seconds. A classical description of the nuclear degrees of freedom is required to capture events which occur at the $\mu$s or $ms$ scale. Electronic rearrangement are ignored in molecular mechanics force (MM) field but during a chemical reaction, a quantum mechanical (QM) description is needed for the system involved in the reaction. To overcome this limitation, methods have been developed to treat the system built upon the level of accuracy desired. Hybrid quantum mechanics/molecular mechanics (QM/MM) strategy treats the involved chemical reaction region at a quantum mechanical level and retains the classical force field for the surrounding part. In a condensed phase, most of chemical reactions have a local character making legitimate to divide the system into an inner QM region (I) and an outer MM region (O), as shown in Figure 17.

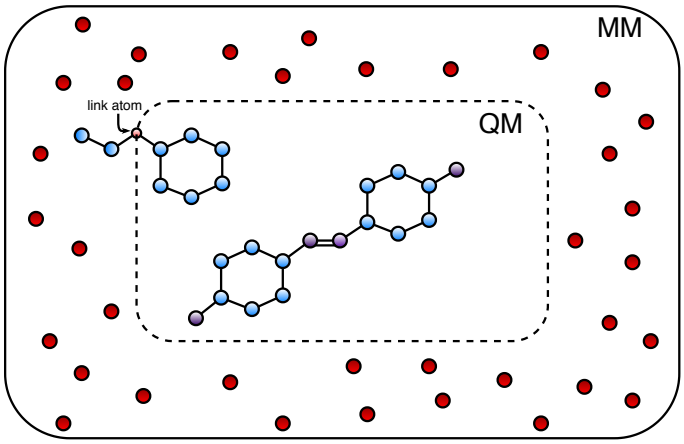


Figure 17: Graphical representation of a QM/MM calculation. The inner part (QM) incorporates the p-diaminoazobenzene and a residue. The outer part represents the MM region, linked to the QM region with the link atom approach. The red point charges of the outer MM region, enter the electronic Hamiltonian of the QM system in the electrostatic embedding approach.

The hybrid QM/MM potential energy contains three classes of interactions : interactions between atoms in the QM region, between atoms in the MM region and between all QM and MM atoms[74]. Two approaches have been developed to evaluate the QM/MM potential energy, subtractive and additive approach.
A subtractive scheme requires an MM calculation of the entire system, a QM calculation of the inner subsystem and a MM calculation of the inner subsystem,

$$V_{QM/MM} = V_O(MM + QM) + V_I(QM) - V_O(QM) \tag{44}$$

where the subscripts indicate the level of theory and the terms in brackets the regions under study. In this scheme, no explicit $QM/MM$ coupling terms are required which makes the implementation simple, but requires the force field parametrization of the inner region which can sometimes be difficult to obtain. A standard subtractive scheme includes a "mechanical coupling" between the $I$and $O$ regions in which the electrons of the QM system do not "feel" the classical electrostatic field of the MM environment.
In an additive approach, the potential energy of the whole system is a sum of the MM energy terms, the QM energy terms and QM/MM coupling terms as shown in the equation below,

$$V_{I/O} = V_I(QM) + V_O(MM) + V_{I-O}(QM + MM) \tag{45}$$

In contrast to the subtractive equation, the MM calculation is only performed on the outer subsystem and a coupling term $V_{I-O}(QM + MM)$ is employed to collect the interaction between the two subsystems. The QM calculation is now performed in the presence of the MM environment. In the "mechanical embedding", all interactions between the two regions are treated at the force field level. The electronic wavefunction is solved for the isolated QM region, avoiding polarization induced by the MM region. An improvement over the mechanical embedding, used to include po-

larization effects, is the electrostatic embedding scheme. In this approach, the MM point charges are incorporated as one-electron terms in the QM Hamiltonian, which will contain an additional term,

$$\hat{H}^{el}_{QM-MM} = -\sum_{i}^{N}\sum_{J}^{L}\frac{q_J}{\mid r_i - x_J \mid} + \sum_{a}^{M}\sum_{J}^{L}\frac{q_J Q_a}{\mid x_a - x_J \mid} \tag{46}$$

where $q_J$ are the MM charges located at $x_J$ and represent the one electron operator while $Q_a$ are the nuclear charges of the QM atoms at $x_a$ and $r_i$ indicates the electron positions. The indices $i$, $J$ and $a$ run over the N electrons, L point charges and M QM nuclei, respectively. The electron structure of the QM region can adapt to the charge distribution of the MM environment and be polarized by it[75]. A problem that arises by using MM atomic charges is the risk of over-polarization near the QM/MM boundary. The point charges in the MM side of the interface may attract or repel the electrons too strongly, which could lead to electron density "spilling" out into the MM region. This issue appear because the Pauli exchange interactions between QM and MM systems are only considered on the level of atomic pair potentials. In this absence of electronic wave functions in the MM part, the QM electrons do not experience any Pauli repulsion from the point charges of the MM region, which could lead to a movement of the electrons towards positively MM charges, being "spilled out" of the QM region.

When the QM and MM regions are connected by chemical bonds, a truncation must be performed to divide the calculation. The most widely used solution introduces a link atom at an appropriate position along the bond vector between the atoms connecting the two regions[75].

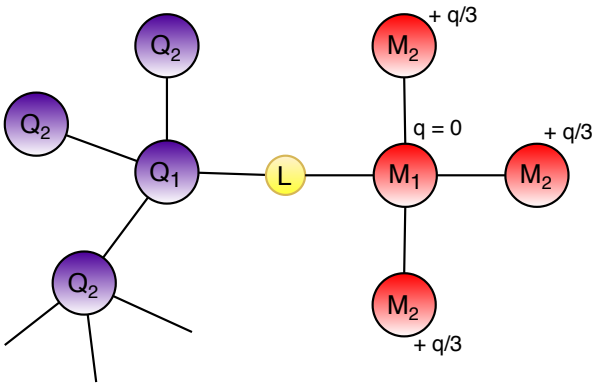


Figure 18: Graphical representation of the link atom approach. The red circles represent the MM region while the purple circles represent the QM region. The two systems are linked by a link atom, represented with a yellow circle. The $M_1$ atom distributes its partial charge in the MM region.

The link atom, is not part of the real system but is covalently bound to $Q_1$ to saturate its valency. The QM calculations are then performed on a system consisting of the QM region and the link atom while the $Q_1$–$M_1$ bond is described at the MM level with the $M_1$ charge set to 0, to alleviate the problem of overpolarization [75].

## 2.3 Quantum Chemistry

### 2.3.1 The Schrödinger equation

The ultimate goal of most quantum chemical calculations is to find the approximate solution of the time-independent, non relativistic Schrödinger equation,

$$\hat{H}\Psi_i(\vec{r_1},\vec{r_2},...,\vec{r_N},\vec{x_1},\vec{x_2},...,\vec{x_M}) = E_i\Psi_i(\vec{r_1},\vec{r_2},...,\vec{r_N},\vec{x_1},\vec{x_2},...,\vec{x_M}) \tag{47}$$

where $\hat{H}$ is the Hamilton operator for a molecular system consisting of $M$ nuclei and $N$ electrons in the absence of magnetic or electric fields. It can be written as,

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_AZ_B}{x_{AB}} \tag{48}$$

which is a differential operator representing the total energy. $A$ and $B$ run over the $M$ nuclei while $i$ and $j$ denote the electrons. The first two terms describe the kinetic energy of the electrons and nuclei and the other three terms define the potential constructed by the attractive electrostatic interactions between electron and nuclei and repulsive potential due to the electron-electron and nucleus-nucleus interactions, respectively. The $r_{ij}$ and $R_{AB}$ are the distance between the electrons $i$ and $j$, $r_{ij} = |\vec{r_i} - \vec{r_j}|$ and in eq.(48), $\Psi_i(\vec{r_1},\vec{r_2},...,\vec{r_N},\vec{x_1},\vec{x_2},...,\vec{x_M})$ represents the wavefunction of the i'th state of the system which depends on the $3N$ spatial coordinates $\vec{r_i}$ and the N spin coordinates of the electrons and the $3M$ spatial coordinates of the nuclei, $\vec{x_i}$. Finally $E_i$ is the eigenvalue representing the energy of the state described by $\Psi_i$. All the equations adopt the system of atomic units, where the mass of an electron, $m_e$, the modulus of its charge, $|e|$, the Planck's constant $h$ divided by $2\pi$ and the permittivity of the vacuum , $4\pi\epsilon_0$ are all set to one[76].

### 2.3.2 Born Oppenheimer Approximation and Variational Principle

The Schrödinger equation can be simplified by assuming that the motion of the nuclei and the electrons can be treated separately. The lightest of the nuclei, ($^1$H), has a mass of 1800 times the mass of an electron, thus the nuclear motion is much slower than the electrons' one. As an approximation, the electrons can be considered to move in the field of fixed nuclei, the second term of eq.(48) can be neglected and the repulsion between the nuclei can be considered to be constant. The Hamiltonian reduces to the electronic Hamiltonian,

$$\hat{H}_{el} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee} \tag{49}$$

and the electronic wavefunction will depend explicitly on the electronic coordinates and parametrically on the nuclear coordinates,

$$\Psi_{el} = \Psi_{el}(\vec{r_i};\vec{x_n}) \tag{50}$$

The time independent Schrödinger equation is an eigenvalue equation,

$$\hat{H}_{el} \left|\Psi\right\rangle = E \left|\Psi\right\rangle \tag{51}$$

where the interest goes in finding approximate solutions to it, since it cannot be solved exactly for systems with more than one electron. Given the Hamilton operator , it exist a set of infinite exact solutions to the Schrödinger equation,

$$\hat{H} \left|\Psi_n\right\rangle = E_n \left|\Psi_n\right\rangle \qquad\qquad n = 0, 1, ... \tag{52}$$

with

$$E_0 \leq E_1 \leq E_2 ... \leq E_i \leq ... \tag{53}$$

Since $\hat{H}$ is an Hermitian operator, the eigenvalues $E_i$ are real and the corresponding eigenfuctions are orthonormal, so by multiplying eq.(52) by its complex conjugate we find,

$$\left\langle\Psi_m\right| \hat{H} \left|\Psi_n\right\rangle = E_n \left\langle\Psi_m|\Psi_n\right\rangle \tag{54}$$

Assuming that the eigenfunction of $\hat{H}$ form a complete set of functions, the wavefunction $\Psi$ can be written as a linear combination of functions forming the set,

$$\left|\Psi_n\right\rangle = \sum_n \left|\Phi_n\right\rangle c_n \tag{55}$$

The expectation value of $|\Psi\rangle$ will be a function of the parameters $c_n$. The variational principle states that given a normalized wavefunction $|\Psi\rangle$ that satisfies the appropriate boundary conditions, the expectation value of the operator must be an upper bound of the exact ground state energy. The "quality" of the wavefunction can be measured by its energy, because the lower the energy, the more precise is the wavefunction describing the system. With the variational method, the parameters of the wavefunction can be varied until the expectation value reaches a stabilized minimum, which will correspond to the estimate of the ground state wavefunction[76].

### 2.3.3 Density Functional Theory

Electron Density The wavefunction $\Psi$ itself it not an observable. A physical interpretation can be associated with the square of the wavefunction,

$$|\Psi_i(\vec{r_1}, \vec{r_2}, ..., \vec{r_N})|^2 d\vec{r_1} d\vec{r_2} ... d\vec{r_N} \tag{56}$$

representing the probability that electrons $1, 2, ..., N$ can be found simultaneously in the volume elements $d\vec{r_1}d\vec{r_2}...d\vec{r_N}$. The probability interpretation leads to the electron density $\rho(\vec{r})$ defined as

$$\rho(\vec{r}) = N \int ... \int |\Psi_i(\vec{r_1}, \vec{r_2}, ..., \vec{r_N})|^2 ds_1 d\vec{r_2}...d\vec{r_N}, \tag{57}$$

which determines the chance of finding any electrons within the volume element $\mathrm{d}\vec{r_1}$ with arbitrary spin. The other N-1 electrons have arbitrary positions and spin in the state represented by $\Psi$. Since electrons are indistinguishable, the probability of finding any electron in that position is $N$ times the probability of finding one precise electron. The electron density, $\rho(\vec{r})$, has three spatial variables which integrates to the total number of electrons and vanishes at infinity. An important feature is that at any position of an atom, $\rho(\vec{r})$ displays a maximum with an infinite value, due to the attractive force of the positively charged nuclei; at these positions, the gradient of the density has a discontinuity and forms a *cusp*. The three properties of that define the density as a physical object able to describe the system are :

- The integral of the density defines the number of electrons.
- The cusps in the density define the position of the nuclei.
- The heights of the cusps define the corresponding nuclear charges.

The premise of density functional theory (DFT) is that all the motions and pair correlations in a many electron systems are contained in the *total electron density*[77]. The Hohenberg-Kohn Theorems The theoretical foundation has been established by Hohenberg, Kohn and Sham in 1964-1965. The first Hohenberg-Kohn (HK) theorem provides a proof that the electron density uniquely determines the Hamilton operator and, thus, also the properties of the system. "The external potential $V_{ext}$ is (to within a constant) a unique functional of $\rho(\vec{r})$; since, in turn $V_{ext}$ fixes $\hat{H}$ we see that the full many particle ground state is a unique functional of $\rho(\vec{r})$". With the proof, $\rho_0$ is a property of the system which contains the informations about $\{ N, Z_A, R_A \}$ as,

$$\rho_0 \Rightarrow \{N, Z_A, R_A\} \Rightarrow \hat{H} \Rightarrow \Psi_0 \Rightarrow E_0. \tag{58}$$

Since the complete ground state energy is a functional of the ground state electron density, also its individual terms will be functionals of $\rho_0$,

$$E_0[\rho_0] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] \tag{59}$$

where $T[\rho_0]$ represents the kinetic energy operator of electrons, $E_{ee}[\rho_0]$ is the electron-electron repulsion operator and $E_{Ne}[\rho_0]$ is the nucleus-electron attraction operator. It is convenient to collect the system independent parts into the *Hohenberg-Kohn functional* $\mathrm{F}_{HK}[\rho_0]$,

$$F_{HK}[\rho] = T[\rho_0] + E_{ee}[\rho_0] = \langle \Psi | \hat{T} + \hat{V_{ee}} | \Psi \rangle \tag{60}$$

$$E_0[\rho_0] = \int \rho_0(\vec{r}) V_{Ne} d\vec{r} + F_{HK}[\rho_0] \tag{61}$$

The Hohenberg-Kohn functional is the holy grail of density functional theory. If it were known exactly, the Schrödinger equation would be solved exactly and since it's a universal functional, it applies equally to the hydrogen atom as to a large molecule. The explicit form lies in the universe somewhere, but at least from the $E_{ee}$ term, the classical Coulomb part $J[\rho]$ is well known,

$$E_{ee}[\rho] = \frac{1}{2}\int\int \frac{\rho(\vec{r_1})\rho(\vec{r_2})}{r_{12}} + E_{ncl}[\rho] = J[\rho] + E_{ncl}[\rho] \tag{62}$$

and the $E_{ncl}[\rho]$ is the non-classical contribution to the electron-electron interaction containing all the effect of self interaction correction, exchange and Coulomb correlations.
But how can we be sure that a density is the ground state density desired? The second HK theorem states that $F_{HK}[\rho]$ delivers the lowest energy if and only if the given density is the true ground state density, $\rho_0$. This means that for any trial density $\rho'(\vec{r})$ associated with some external potential $V'_{ext}$, the energy obtained represents an upper bound to the true ground state energy $E_0$,

$$E_0 \leq E[\rho'] = T[\rho'] + E_{Ne}[\rho'] + E_{ee}[\rho'] \tag{63}$$

making use of the variational principle established in the previous section[77]. The Konh Sham approach The clever idea of Kohn and Sham was to realize that since it is not possible to accurately determine the kinetic energy through an explicit functional, it can partially determined in an approximated manner. They suggested to obtain the exact kinetic energy of a non-interacting reference system with the same density of a real interacting one. In this approach, one assumes a fictitious system of non interacting electrons with density $\rho^{ni}(\vec{r})$ that is the same as the density $\rho(\vec{r})$ of the real system of interacting electrons.

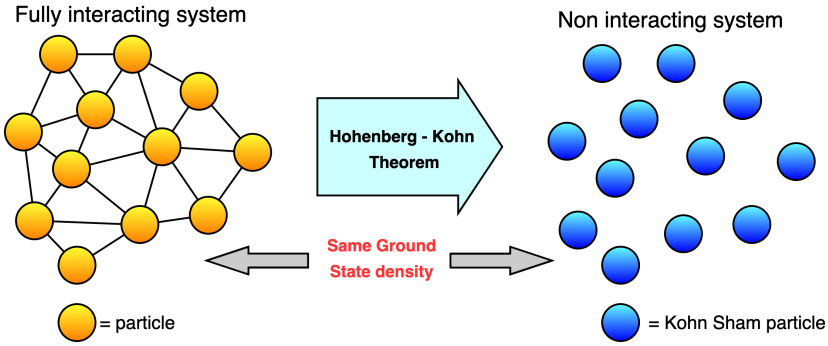


Figure 19: Graphical representation of a fully interacting system and a non interacting system. Both systems share the same electron density.

The non-interacting kinetic energy is not equal to the true kinetic energy even if the systems share the same density but they accounted for that introducing a separation of the functional $F[\rho]$,

$$F[\rho(\vec{r})] = T_S[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})] \tag{64}$$

where $T_S$ represents the non interacting kinetic energy,

$$T_S = -\frac{1}{2}\sum_i^N \left\langle \phi_i \middle| \nabla^2 \middle| \phi_i \right\rangle \tag{65}$$

where $\phi_i$ are orthonormal spin orbitals and $E_{XC}$ is the *exchange-correlation* energy defined as,

$$E_{XC}[\rho\vec{r}] \equiv (T[\rho] - T_S[\rho]) + (E_{ee}[\rho] - J[\rho]) = T_C[\rho] + E_{ncl}[\rho]) \tag{66}$$

The rest of the true kinetic energy $T_C$ is added to the non-classical electrostatic contributions, hence the $E_{XC}$ contains everything that is unknown such as the non-classical effects of self interaction correction, exchange, correlation and a portion of the true kinetic energy. So the energy for the real interacting system is given by

$$E[\rho(\vec{r})] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \tag{67}$$

where the only term with no explicit form is the $E_{XC}$. To find a solution, such that it can provide a Slater determinant characterized by the same density of the real system, the variational principle is applied to the energy of the real interacting system,

$$E[\rho(\vec{r})] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \tag{68}$$

where the only term with no explicit form is the $E_{XC}$. The orbitals $|\phi_i\rangle$ are calculated as eigenfunctions of the Kohn-Sham operator $\hat{h}^{KS}$,

$$\hat{h}^{KS}\,|\phi_i\rangle = (T_S[\rho(\vec{r})] + V_{Ne}[\rho(\vec{r})] + J[\rho(\vec{r})] + V_{XC}[\rho(\vec{r})])\,|\phi_i\rangle = E_i\,|\phi_i\rangle \tag{69}$$

The term $V_{XC}$, which is the potential due to the exchange-correlation energy $E_{XC}$ is defined as the functional derivative of $E_{XC}$ with respect to $\rho$,

$$V_{XC} \equiv \frac{\delta E_{XC}}{\delta \rho} \tag{70}$$

If the exact forms of $E_{XC}$ and $V_{XC}$ were known, the Kohn-Sham approach would give the exact energy. Unlike in the Hartree-Fock model, where the approximation is introduced from the start assuming the wavefunction as a single Slater determinant, the Kohn-Sham approach is in principle exact; the approximation comes in when the explicit form of the unknown functional for the exchange-correlation energy must be decided. This is why, the central goal of modern density functional theory is to find better and better approximations to these quantities[77]. Exchange-Correlation Functionals In DFT there is no systematic way towards improving approximate functionals, which represents the major drawback associated with this approach. In the last 30 years, a hierarchy of density functional approximations for $E_{XC}$ has been established from the simplest

to the most sophisticated[78]. We can write $E_{XC}$ as,

$$E_{XC}[\rho] = \int \rho(\vec{r}) V_{XC}(\rho(\vec{r})) d\vec{r} \tag{71}$$

where the integrated $V_{xc}(\rho(\vec{r}))$ is the exchange correlation energy density. The hierarchy is ordered by increasing the complexity to construct $\epsilon_{xc}(\rho(\vec{r}))$. At the lowest level, the local density is treated as a slowly varying function and $\epsilon_{xc}$ will be the energy per particle of a uniform electron gas of density $\rho(\vec{r})$[79]. In a uniform electron gas, an infinite number of electrons travel through a space of infinite volume in which there is uniform and continuous distribution of positive charge and the density maintains a constant value everywhere. This approximation is surprisingly accurate for the determination of molecular properties such as equilibrium structures, harmonic frequencies or charge moments but very poor in the description of atomization energy[80]. The errors were greatly reduced considering a non uniform electron gas as in the *Generalized Gradient Approximation* (GGA)[81]. It uses not only the information about the density $\rho(\vec{r})$ at a particular point $\vec{r}$ but also the information about the gradient of the charge density, $\nabla\rho(\vec{r})$, to account for the non-homogeneity of the true electron density. The $E_{XC}$ in GGA can now be written as,

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \tag{72}$$

where $\rho_\alpha$ and $\rho_\beta$ are the spin densities with $\rho_\alpha + \rho_\beta = \rho(\vec{r})$ and the $E_{XC}^{GGA}$ is usually split into its exchange and correlation contributions,

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \tag{73}$$

where approximations for the two terms are made individually. GGA functionals are frequently called *non-local*, but mathematically, the value of the functional at a point $\vec{r}$ depends only on the informations about the density $\rho(\vec{r})$ and its gradient $\nabla\rho(\vec{r})$; the *non-local* arises from the motivation that these functionals go beyond the *local* density approximation.

Beyond the GGA functionals, we can find the *meta-GGA*(MGGA) functionals, which also make use of the laplacians $\nabla^2\rho_\alpha(\vec{r})$ and $\nabla^2\rho_\beta(\vec{r})$[82].

An appropriate strategy to achieve more accurate exchange-correlation energy is the mixing of some exact Hartree-Fock exchange into $E_{XC}$, suggesting *hybrid density functionals*[83] of the form

$$E_{xc}^{hybrid} = E_{xc}^{GGA} + a(E_x^{exact} - E_x^{GGA}). \tag{74}$$

A hyper-GGA (HGGA) uses the exact exchange,

$$\epsilon_x^{HGGA}(\vec{r}) = \epsilon_x^{exact}(\vec{r}), \tag{75}$$

and constructs $\epsilon_c^{HGGA}$(r) from the meta-GGA ingredients plus $\epsilon_x$(r) in form of,

$$\epsilon_{xc}^{HGGA} = \epsilon_c^{MGGA} + (\epsilon_x^{MGGA} - \epsilon_x^{exact})\left[1 - e^{-v\left(\frac{\epsilon_c^{MGGA}}{\epsilon_x^{MGGA}}\right)}\right] \tag{76}$$

where $v$ is a large positive number. In an exchange dominated region of space where $\epsilon_c^{MGGA}/\epsilon_x^{MGGA} < 1$, we find $\epsilon_{xc}^{HGGA} \approx \epsilon_x^{exact} + \epsilon_c^{MGGA}$; otherwise we find $\epsilon_x c^{HGGA} \approx \epsilon_{xc}^{MGGA}$[84]. The exact exchange has the ability to compensate for the self-interaction of local exchange functionals. Before going on, a difference between short range (SR) and long range (LR) correlation must be introduced. In DFT, the Coulomb correlation energy $E_c$ is generally implied to be the *dynamic* (SR) correlation which is responsible for accounting the instantaneous interaction between electrons with opposite spin at short distances. Another type of correlation is the *non-dynamic* (LR) correlation which can be explained as the decrease in Coulomb repulsion of electrons due to the higher probability of occurrence of one electron near its nucleus, as in the $H_2$ dissociation. Opposed to the exact exchange $E_x^{exact}$, the semilocal exchange functionals can effectively take into account non-dynamic correlation. Consequently, the choice of a combination of different parts of functionals is essential to balance between the elimination of self-interaction and the inclusion of non-dynamic correlation[79]. The hybrid functional based on Becke's three parameters scheme (B3[85]) became one of the most popular functionals based on semi-empirical data,

$$E_{xc}^{hybrid} = a_0 E_x^{exact} + (1 - a_0)E_x^{LDA} + a_x \Delta E_x^{B88} + E_c^{LDA} + a_c \Delta E_c^{PW91}, \tag{77}$$

$$a_0 = 0.2 \qquad a_x = 0.72 \qquad a_c = 0.81$$

where $\Delta E_x^{B88}$ is the gradient correction to the exchange energy of the B88 functional, and $\Delta E_c^{GGA}$ is the gradient correction to the correlation energy. The parameters $a_0$, $a_x$ and $a_c$ are obtained by fitting the thermodynamical data obtained by using eq.(77) to the the corresponding experimental data. The most popular hybrid functional is known as B3LYP, in which the correlation functional PW91 is replaced by the LYP functional, taking the value of the three parameters directly from Becke's original paper[86, 87, 88, 89]. Long Range Exchange and Dispersion Corrections In DFT, the hybrid functional B3LYP appears to offer a great contribution, by number of applications, to the field of theoretical chemistry. However it is not successful in describing : (i) the polarizability of long hydrocarbon chains, (ii) excitations using time dependent DFT and most importantly (iii) charge transfer excitations[90]. The reason lies in the description by B3LYP of the exact potential, which behaves as $-0.2r^{-1}$, instead of the exact value $-r^{-1}$. The long range exchange correction (LC) scheme, corrects this deficiency by splitting the electron repulsion $r_{12}^{-1}$ into a short and a large part using the standard error function (erf) as,

$$\frac{1}{r_{12}} = \frac{1 - erf(\mu r_{12})}{r_{12}} + \frac{erf(\mu r_{12})}{r_{12}} \tag{78}$$

where the first term accounts for the SR interaction and the second term for the LR interaction[91]. The key idea is that the DFT exchange interaction is included using the first term (SR), and the LR orbital-orbital exchange interaction is described with the Hartree Fock (HF) exchange in the second term. If $\mu$=0, the LC DFT calculation corresponds to the pure non-LC DFT and if $\mu = \infty$ corresponds to the standard HF calculation. It is possible to combine the qualities of the energetics of B3LYP with the LR improvements of the LC scheme. The Coulomb-attenuating method (CAM), introduces two extra parameters in the previous equation[90],

$$\frac{1}{r_{12}} = \frac{1 - [\alpha + \beta \cdot erf(\mu r_{12})]}{r_{12}} + \frac{\alpha + \beta \cdot erf(\mu r_{12})}{r_{12}}, \tag{79}$$

The parameter $\alpha$ allows to incorporate the HF exchange contribution over the whole range by a factor of $\alpha$, and the parameter $\beta$ allows to incorporate the DFT part over the whole range by a factor of $1 - (\alpha + \beta)$. B3LYP would have a CAM potential with with $a_0$=$\alpha$=0.2 and $\beta$=0 while the LC equation takes $\alpha$=0 and $\beta$=1.0. The optimal values found for CAM-B3LYP are $\alpha = 0.19$ and $\alpha$+$\beta$=0.65 and $\mu$=0.33. Optimizing LC and hybrid functionals with identical numbers of parameters in their GGA exchange anc correlation terms leads to better results for *all* properties using the LC form.

Another LC base functional is $\omega$B97X, which mixes a small fraction or SR-HF exchange with SR-DFT exchange to provide non-local correction to the short range exchange[92]. This functional has the form of,

$$E_{xc}^{97X} = E_x^{LR-HF} + c_X E_x^{SR-HF} + E_x^{SR-B97} + E_c^{B97} \tag{80}$$

which includes a small fraction of exact SR exchange. Those functionals can be further improved by addition of empirical parameters to improve dispersion interactions[93]. The idea is to add an empirical atom-atom dispersion potential to an existing density functional to obtain a correction of zero cost and complexity. Those functionals named DFT-D have the form,

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \tag{81}$$

where $E_{disp}$ is the empirical dispersion correction given as,

$$E_{disp} = \sum_{i}^{N-1} \sum_{j=i+1}^{N} \frac{C_6^{ij}}{x_{ij}^x} f_{damp}(x_{ij}) \tag{82}$$

with N is the number of atoms in the system, $C_6^{ij}$ is the dispersion coefficient for atoms $ij$ and $x_{ij}$ is an interatomic distance between atoms $i$ and $j$. The function $f_{damp}$ is a term which establish the condition of zero dispersion correction at short interatomic distance and corrects asymptotic pairwise van der Walls potentials, and has the form,

$$f_{damp}(R_{ij}) = \frac{1}{1 + a(x_{ij}/x_r)^{-12}} \tag{83}$$

which reduces to 1 at large $R_{ij}$ and vanish at small $x_{ij}$ and the parameter $a$, controls the strength

of dispersion corrections.

### 2.3.4 Time Dependent Density Functional Theory and Charge Transfer

In 1984 Erich Runge and Hardy Gross proved that for time-dependent (TD) systems evolving from a given initial wavefucntion, all TD properties can be extracted from the time-evolving density, naming it TD-DFT[94]. The TD-DFT can be applied to the time-dependent Kohn-Sham formalism by propagating the time dependent Kohn-Sham orbitals,

$$\hat{h}^{KS}[\rho(r,t)]\,|\phi_i(r,t)\rangle = i\frac{\delta\,|\phi_i(r,t)\rangle}{\delta t} \tag{84}$$

where $\hat{h}^{KS}[\rho(r,t)]$ can be expressed as,

$$\hat{h}^{KS}[\rho(r,t)] = T_S[\rho] + E_{Ne}[\rho] + J[\rho(r,t)] + V_{XC}[\rho(r,t)] \tag{85}$$

The exchange-correlation functional $V_{XC}[\rho(r,t)]$ depends on the time dependent density and its exact form is now known. The $V_{XC}[\rho(r,t)]$ is assumed to a function of an electron density $\rho(\vec{r})$ which only depends parametrically on time and become a different function for each value of $t$[95]. The theory was further developed by Casida in 1995 with the linear-response (LR) formalism which calculates the response of the density to a time-dependent perturbation and allows to efficiently determine the solution of the TD-DFT equations for molecules[96].
The description and understanding of excited state (ES) processes is very challenging for experimental methods and theoretical calculations can access properties difficult to investigate by experiments. The analysis or electronic transition is based on the Frank-Condon principle. Since the nuclear masses are much larger than the masses of electrons, an electronic transition occurs within a stationary framework. As a result, the nuclear locations remain unchanged during the actual transition and adjust when electrons have adopted their final distribution.

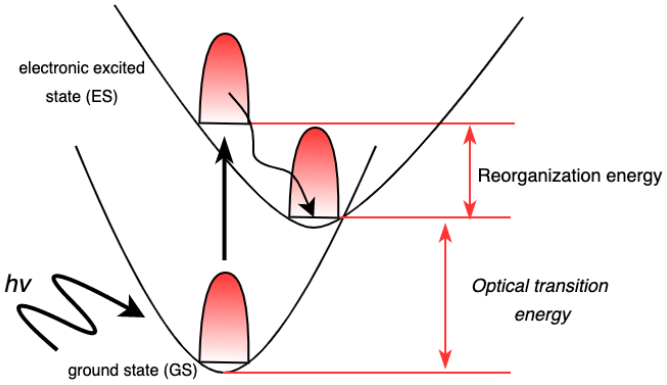


Figure 20: Vertical excitation of the electronic ground state resulting in an excited state. The excitation results in an unchanged nuclear disposition followed by a vibrational relaxation to the equilibrium state.

The modelling of ES properties starts with a vertical excitation. In such calculation, the geometry remains frozen, generally in the GS, and the transition energies to the first few low-lying

ES can be computed. The vertical transition energies are the difference between ES energies and GS energy at a frozen geometry,

$$E^{abs} = E^{ES}(x_{GS}) - E^{GS}(x_{GS}) \tag{86}$$

where $E^{ES}$ and $E^{GS}$ are the electronic energy of the ES and GS and $x_{GS}$ represents the nuclear geometry of the ground state. In a TD-DFT calculation for vertical excitation, the total number of ES state (roots) is computed including the multiplicity of the final states, i.e., singlet-singlet or singlet-triplet transitions. The choice of the functional is extremely important to obtain accurate results. The benchmark of the functionals can be split into two subgroups depending on the selected reference data. One could use as reference highly accurate wavefunction approaches (MR-CI, CASPT2), while the second approach relies on experimental data, where most published applications of TD-DFT rely on comparing $\lambda_{max}$ with $E^{abs}$ since in "real life", there is no vertical experimental data[97].

# 3 Computational Details

## 3.1 Atomistic Molecular Dynamics Simulation Details

All the modifications to the topology and trajectory files were performed using cpptraj, the main program in Amber for processing coordinate trajectories and data files and PyTraj, a python library which exposes cpptraj's functions to the python environment[98]. Visual Molecular Dynamics (VMD) was used for visual inspection of the trajectories, calculate the root mean squared deviation (RMSD) and build visual representations[99]. Python3.7 was used to produce plots and make statistical analysis using Matplotlib, Numpy, Pandas and Scipy.

### 3.1.1 Initial Structures

The c-EM crystallized structure of a truncated pore model of the $Na_v1.4$ channel was collected from the PDB, 6AGF as PDB ID[100][37]. The S5 and S6 $\alpha$ helices of domains I-IV were selected, the $\beta$ domain was removed, and after amidation and acetylation of the -N and -C end groups, the structure was placed inside a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer using CHARMM-GUI membrane builder. The structure was solvated in a rectangular periodic box with acqueous 0.15M NaCl to reach physiological salt concentration[101]. The residue indices selected are summarized in Table 1.

Table 1: Selected residues for the pore model

| **Domain** | **Chain ID** | **Residues** |
|---|---|---|
| I | A;B | 234 – 286; 336 – 451 |
| II | C | 683 – 805 |
| III | D | 1143 – 1298 |
| IV | E | 1464 – 1601 |

The p-diaminoazobenzene structure was build with a molecular editor, Avogadro, the geometry

of the molecule was optimized at MP2/6-31G* level of theory and the restrained electrostatic potential (RESP) charges were obtained with a single point energy of the optimized structure at the Hartee-Fock/6-31G* level of theory with Gaussian09 software[102, 103]. The required topology and coordinate Amber files were generated with a module of AmberTools 14, Antechamber[104]. A pdb file containing $Na_v1.4$ and p-diaminoazobenzene was build with VMD and the ligand was positioned at the center of mass (CoM) of the DEKA ring which also correspods to the CoM of the channel using a VMD extension, TkConsole. A python script, charmmlipid2amber.py, was used to convert the CHARMM-GUI PDB format to an Amber files format and all the molecular dynamics (MD) simulations were performed with the GPU version of the Amber Molecular Dynamics Package[105]. The solvated complex consists of 99499 atom in total, where the channel consists of 9328 atoms and the p-diaminoazobenzene of 28 atoms.

### 3.1.2 Molecular Dynamics Simulation Details

All MD simulations were performed using the GPU version of the AMBER 18 package [105]. The forces for the lipids and transmembrane protein were based on the CHARMM36 general force field and those of water with the TIP3P model[106, 107]. First, an energy minimization was carried out with the steepest descendent method for 2500 steps followed by conjugate gradient method for 2500 steps, to find a local energy minimum and avoid unreasonable atom's distance shift in a single time-step. Hydrogen bond lengths were kept fixed using the SHAKE algorithm[108]. Electrostatic interactions were calculated using the particle-mesh Ewald method with a grid spacing of 1.0 Å and the van-der-Walls cutoffs were set to 12 Å.
An NVT heating at 300K over 125 ps was performed twice with 1 fs timesteps, a positional restraint of 10 kcal/mol was applied to the entire protein and a positional restraint of 2.5 kcal/mol was applied to the POPC lipid bilayer. The heating steps, let the system "relax" and reach the desired temperature with no changes in volume. The positional restraints were releases and the desired density was achieved by running an equilibriation in an NPT ensemble, allowing the volume to change during the simulation. The system was equilibriated in an NPT ensemble with a Monte Carlo barostat and a semiisotropic pressure scaling, once for 125 ps with 1fs timesteps and three times for 500 ps with 2 fs timesteps. The production run was carried out in an NPT ensemble using Langevin dynamics at 300 K and a pressure of 1 atm, for 200ns.

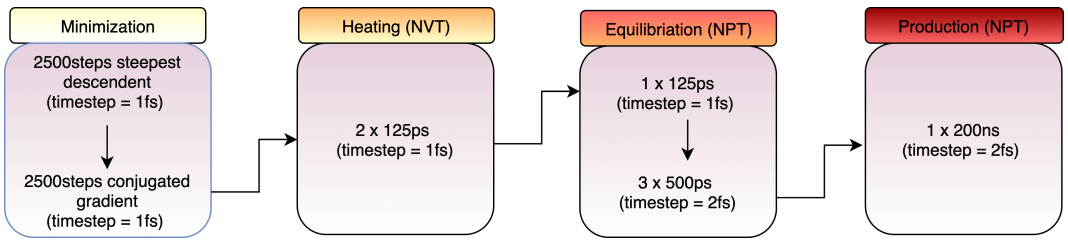


Figure 21: MD simulation steps.

The energies were printed every 2 ps while snapshots of the trajectory were recorded every 100 ps for further analysis.

### 3.1.3 Gaussian Accelerated Molecular Dynamics Simulation Details

GaMD has been implemented in both the serial and parallel versions on CPU and GPU, and it follows the same keys nomenclature of AMBER with an additional keyword to recognize the method. The final frame of the cMD simulation was extracted from the trajectory and employed as initial coordinate structure for the GaMD simulation. Four parallel GaMD production run were performed following the GaMD algorithm with the time lengths shown in Table 2.

Table 2: Time lengths of GaMD algorithm. The same procedure was applied to all four parallel simulations.

| GaMD x 4 | |
|---|---|
| **Stage 1** | 2ns |
| **Stage 2** | 8ns |
| **Stage 3** | 5ns |
| **Stage 4** | 35ns |
| **Stage 5** | 1000ns |

Since the system was already equilibriated in cMD for 200ns, state 1 and stage 2 were only performed to activate the algorithm and obtain the values of the cMD potential. Stage 4 was run to obtain the GaMD statistics, until the value of $k_0$ was stable to 1, to give maximum boost to the potential. The total production run was $4\mu$s divided between the four parallel simulation and snapshots were taken every 20ps for further analysis. A RSMD of p-diaminoazobenzene with respect to the aligned protein was calculated as a first indication of the ligand's position inside the channel. The potential energy was reweighted and the position of the ligand was further investigated obtaining two distance vector from the ligand's CoM to the DEKA's and intracellular gate (gate) CoMs ($d$) and the angle between the two vectors ($\Theta$). The free energy surface with respect to the two coordinates was computed to identify the binding pockets of the drug into the protein. Figure 22 below shows a graphical representation of the ligand inside the ion channel.

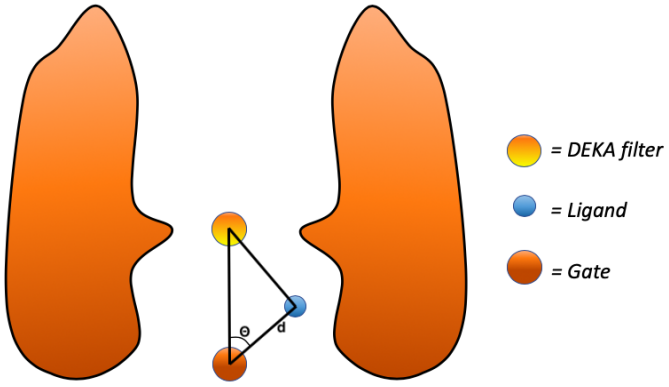


Figure 22: Graphical representation of the $\alpha$ domain of Na$v$1.4 ion channel. The light orange circle is the CoM of the DEKA filter filter, the darker orange is the CoM of the Gate and the blue circle is the CoM of the ligand.

Also the probability distribution of 4$\mu$s of GaMD was compared with 4$\mu$s of cMD with the same starting coordinate structure to compare the sampling of both methods.

### 3.1.4 Molecular Mechanics Generalized Born Surface Area Simulation Details

After identification of binding pockets from the 4$\mu$s of GaMD, 100ns of cMD were additionally run for each pocket to confirm that the binding was stable and to compute the binding free energy. The RMSD of p-diaminoazobenzene with respect to the aligned protein was calculated to control that the drug kept its position in the pocket and discard simulations with a change in binding pocket. If the ligand had a fluctuation larger than $\sim$ 2 Å in RMSD value, the simulation was discarded. An analysis was realized to choose the number of frames required to run MMGBSA. For the pairwise residue decomposition analysis, the residues within a radius of 5 Å from the ligand were selected, including all the residues present in at least 2% of the selected frames.

## 3.2 Quantum Chemistry Calculation Details

### 3.2.1 Polarizable Continuum Model Vertical Excitations

All simulations have been performed with the Q-Chem software and the results were visualized with IQmol molecular viewer[109]. The ground state geometry of Azobenzene and P-diaminoazobenzene were optimized at B3LYP/cc-pvdz level of theory. With the optimized geometries, the vertical transition energies to the valence excited states were computed at TD-DFT B3LYP-D3/cc-pvdz, CAM-B3LYP-D3/cc-pvdz and $\omega$B97X-D/cc-pvdz level of theory with the polarizable solvent model (PCM). These vertical values were straightforwardly compared to the experimental $\lambda_{max}$ which is a common but physically incorrect approximation.

### 3.2.2 Vertical Excitations in Protein Environment

The electrostatic embedding QM/MM scheme was employed for the calculations. The binding pocket with the lowest ligand's total binding free energy, pocket 4, was selected for the QM/MM

calculations. From the 100ns cMD trajectory, 20 equidistant frames were extracted (every 5ns) and based on the pairwise ligand-residue decomposition energy analysis, Phe571 was selected and included with p-diaminoazobenzene in the QM region. The default AMBER's link-atom approach was employed, which automatically adds a H atom at the QM/MM boundary, separating the residue Phe571 from the rest of the protein which is treated at the MM level. For each frame, the vertical transition energies to the valence excited states were computed with TD-DFT at B3LYP-D3/cc-pvdz CAM-B3LYP-D3/cc-pvdz and $\omega$B97X-D/cc-pvdz level of theory and the three functionals were compared.

# 4 Results and Discussion

## 4.1 Classical Molecular Dynamics Production - RMSD

The equilibriation of the structure of the ion channel was measured via RMSD, aligning it to its initial position. By aligning a structure, the translational motion is eliminated, and a RMSD gives an estimation of the internal motion of the protein relative to its initial position. Snapshots from the trajectory of the 200ns production run were taken every 0.1ns. Figure 23, shows the RMSD of the aligned channel relative to its initial position.

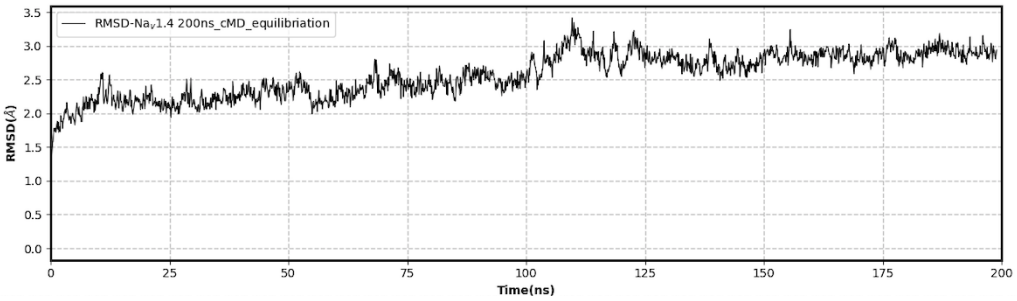


Figure 23: RMSD of aligned channel with respect to its initial position.

The protein equilibriates its internal motion at $\sim$ 125ns with a value oscillating between 2.5 Å and 3.0 Å from the initial structure. With an equilibriated structure, the accelerated dynamics can be performed to sample the phase space and find appropriate binding pockets for the ligand. The last snapshot was extracted from the trajectory and used as starting structure for the GaMD simulations.

## 4.2 Gaussian Accelerated Molecular Dynamics - Pocket Search

### 4.2.1 Free energy maps

A two dimensional energy profile was calculated by reweighting the four GaMD simulations. The reweighted energy profile allows to identify five low energy states corresponding to the five most visited ligand conformations, i.e., binding pockets inside the channel. Two arbitrary points in the channel were selected as reference. The CoM of the $\alpha$ carbons of the four residues composing the DEKA filter ($Asp^{406}$/$Glu^{761}$/$Lys^{1244}$/$Ala^{1536}$) and the CoM of the $\alpha$ carbons of four residues at the gate ($Val^{446}$/$Leu^{800}$/$Val^{1293}$/$Ile^{1596}$). The $d$ axis represents the vector distance between the CoM of the gate and the CoM of p-diaminoazobenzene and gives an indication of the depth of the ligand inside the channel. The $\Theta$ axis corresponds to the inner product between two vectors, both starting at the gate CoM and pointing respectively at the DEKA and ligand's CoM. The $\Theta$ axis gives an indication of the position of the ligand with respect to the geometrical center of the channel. The color scale from red to blue represents the free energy, directly proportional to the probability distribution of the ligand visiting a point the grid. Figure 24 gives a graphical representation of the channel and the ligand, gate and DEKA CoMs and the four energy maps relative to the four GaMD simulations.

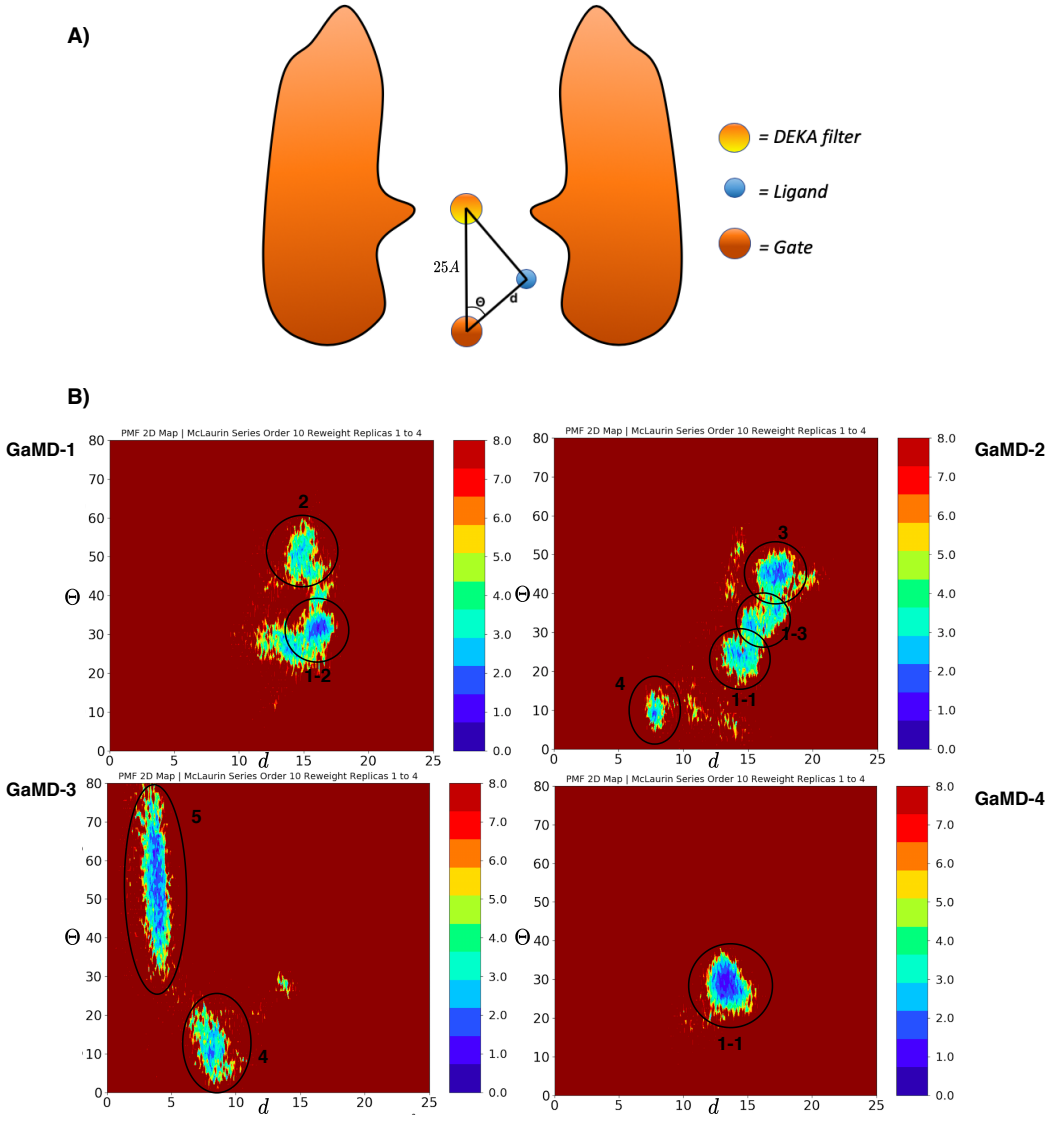


Figure 24: A) Graphical representation of the Na$v$1.4 channel. The yellow bead represent the CoM of the DEKA filter, the orange bead represent the CoM of the gate and the red bead represents the CoM of the ligand. The average distance between the DEKA and gate CoMs is $\sim$ 25 Å. B) 2D energy profile of the four GaMD simulations. The $d$-axis is the distance between the CoM of the $\Theta$ carbons of the gate and the CoM of p-diaminoazobenzene. The $\Theta$-axis corresponds to the angle between the two vectors, both starting at the gate CoM pointing at the DEKA ligand's CoMs.

Pocket 1, found at (15 Å-17Å, 30°-50°) shares three conformations since three different minima are found and shown as pocket 1-1,1-2 and 1-3. Pocket 2 is found at (15Å,50°-60°), similarly to pocket 3. Those two pockets share the same coordinates in the energy map but by visual inspection with VMD, the ligand was found to be in two different points in the channel, opposite to each other. The position's given by $d$ and $\Theta$ only give a 2D representation of the channel, thus, the same coordinates could correspond to two opposite points in the channel. Pocket 4 is found at (8Å,5°-15°) while pocket 5 is found at (5Å, 30°-70°). Since the ligand's CoM is adjacent to the gate ($d = 0$), the $\Theta$ axis will be more "sensible" to small changes in position and this is why it fluctuates between 30° and 70°. The pockets found via energy map were compared and identified

in the RMSD of the simulations.

### 4.2.2 Root Mean Square Deviation of the Photoswitch

Additionally to the energy maps, the RMSD of the ligand with respect to the aligned protein structure was calculated to determine its position in space with respect to the initial structure.

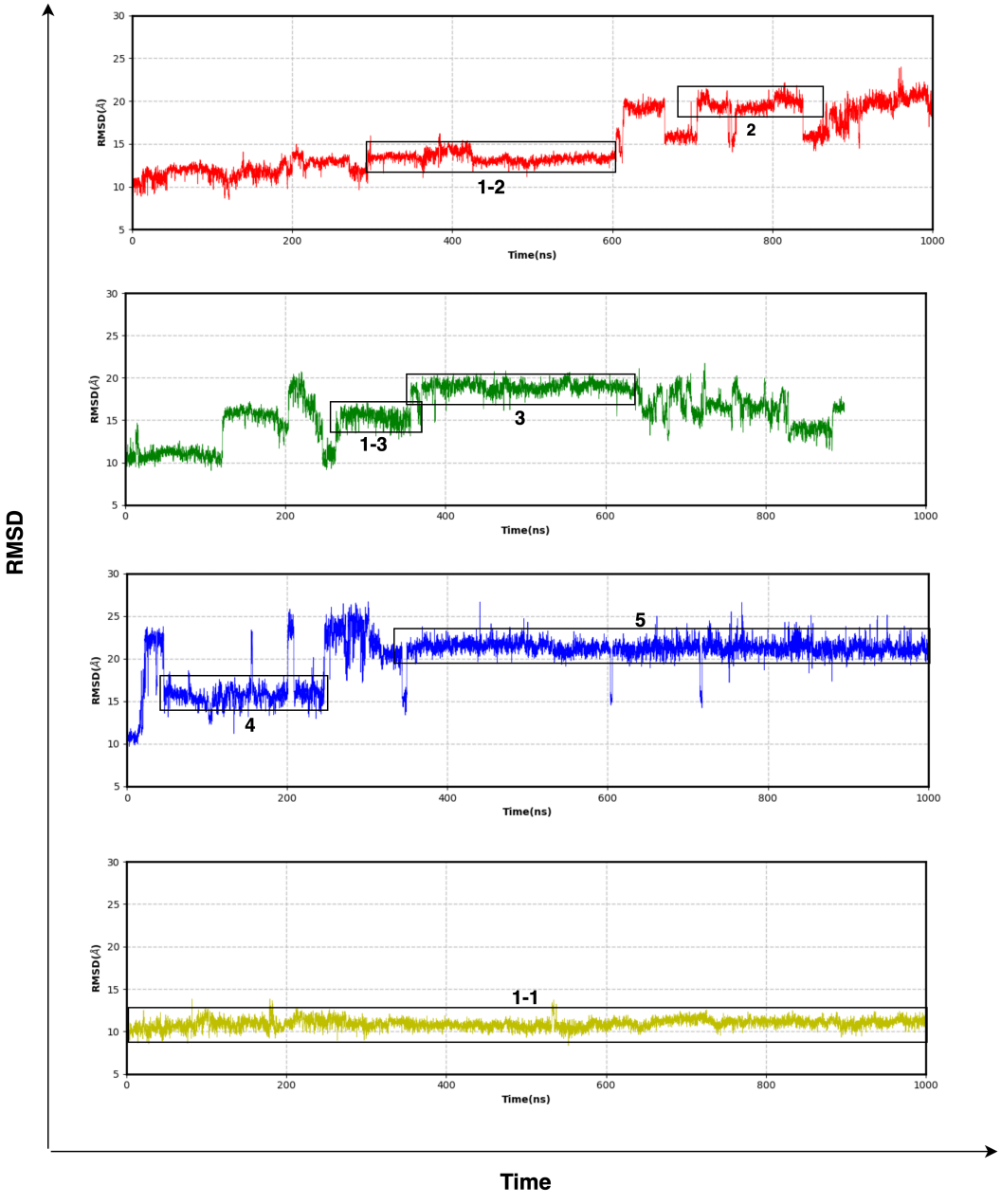


Figure 25: RMSD of p-diaminoazobenzene with the aligned protein structure in four parallel GaMD productions of 1$\mu$s each.

For p-diaminoazobenzene, a RMSD fluctuation of 1-2Å, corresponds to a change in its orientation in space maintaining the CoM in the same position. With a RMSD fluctuation of 3Å and higher, there is also translational motion involved moving away or towards a different binding

pocket. Figure 25 shows the ligand's RMSDs in the four 1$\mu$s GaMD production runs. The pockets identified with the energy maps were also found in the RMSDs profiles as stable regions for at least 100ns. The CoM of the DEKA ring corresponds to the (25,0) point in the energy maps and the (0,0) point in the RMSDs of the GaMD productions. Figure 26 shows the RMSD and the energy map of pocket 5 and pocket 1-1.

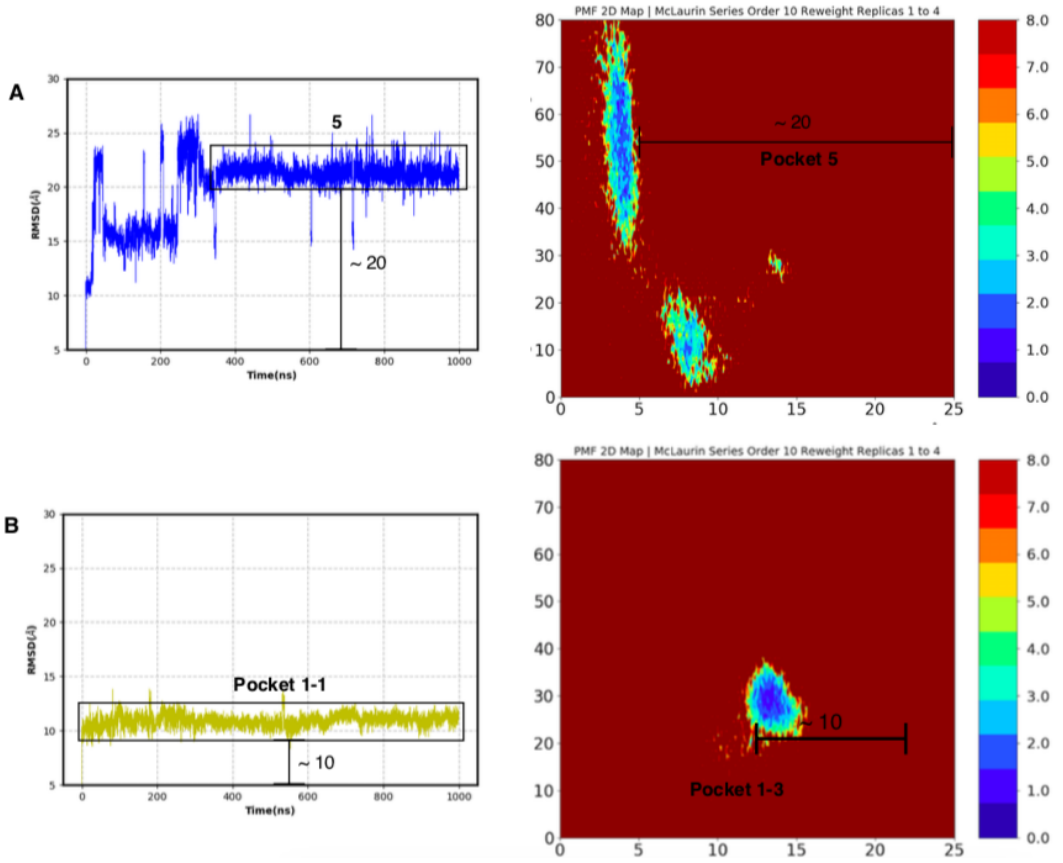


Figure 26: A)Binding of p-diaminoazobenzene to the Na$v$1.4 channel in pocket-5. B) Binding of p-diaminoazobenzene to the Na$v$1.4 channel in pocket 1-1.

Comparing the RMSD and the energy plot (26A), the ligand samples the space at 20Å distance from the DEKA, nearby the gate. Comparing the RMSD and the energy plot (26B), the ligand samples the space at 10Å distance from the DEKA ring. In pocket 1-1, the ligand has kept it's position and orientation in space throughout $\sim$ 1000ns. From the RMSDs, ligand's conformations with a flactuation of only 1-2Å for 100ns of more were selected for MMGBSA energy analysis. In figure 25, from the red plot, pocket 1-2 and 2 were selected, in the green plot pocket 1-3 and 3 were selected and additionally also a frame was extracted between 50 and 150ns. In the blue plot, pocket 4 and 5 were selected between 140-240ns and between 350-1000ns. Finally in the yellow plot, only one conformation was selected since the ligand kept a fluctuation of 1-2Å throughout the whole 1$\mu$s simulation. Table 3 below summarized the number of selected conformations in the four GaMD simulations.

Table 3: Number of conformations selected for MMGBSA energy analysis

| | **N of selected conformations** | **Timeframe (ns)** |
|---|---|---|
| GaMD-1 (red) | **2** | 400-600<br>650-830 |
| GaMD-2 (green) | **3** | 50-150<br>300-400<br>420-700 |
| GaMD-3 (blue) | **2** | 140-240<br>350-1000 |
| GaMD-4 (yellow) | **1** | 0-1000 |

From each timeframe shown in Table 3, a snapshot was extracted from the respective trajectories and employed as starting structure for further energy analysis.

#### 4.2.3 MMGBSA : Evaluation of the Binding Free Energy

The total binding free energy of p-diaminoazobenzene inside the channel was evalutated with MMGBSA. A total of three distinct pockets were found and the energy was pairwise decomposed to have quantitative values of the interactions of the ligand with the surrounding residues.
After the GaMD productions run, a total of 8 snapshots, shown in Table 3, were extracted from the simulations and 100ns of cMD production were run from each frame to check whether the binding of the ligand is stable with time in each conformation. The selection was based on both the energy maps and the RMSD of the GaMD productions. Figure 27 shows the RMSD of the seven cMD productions run.

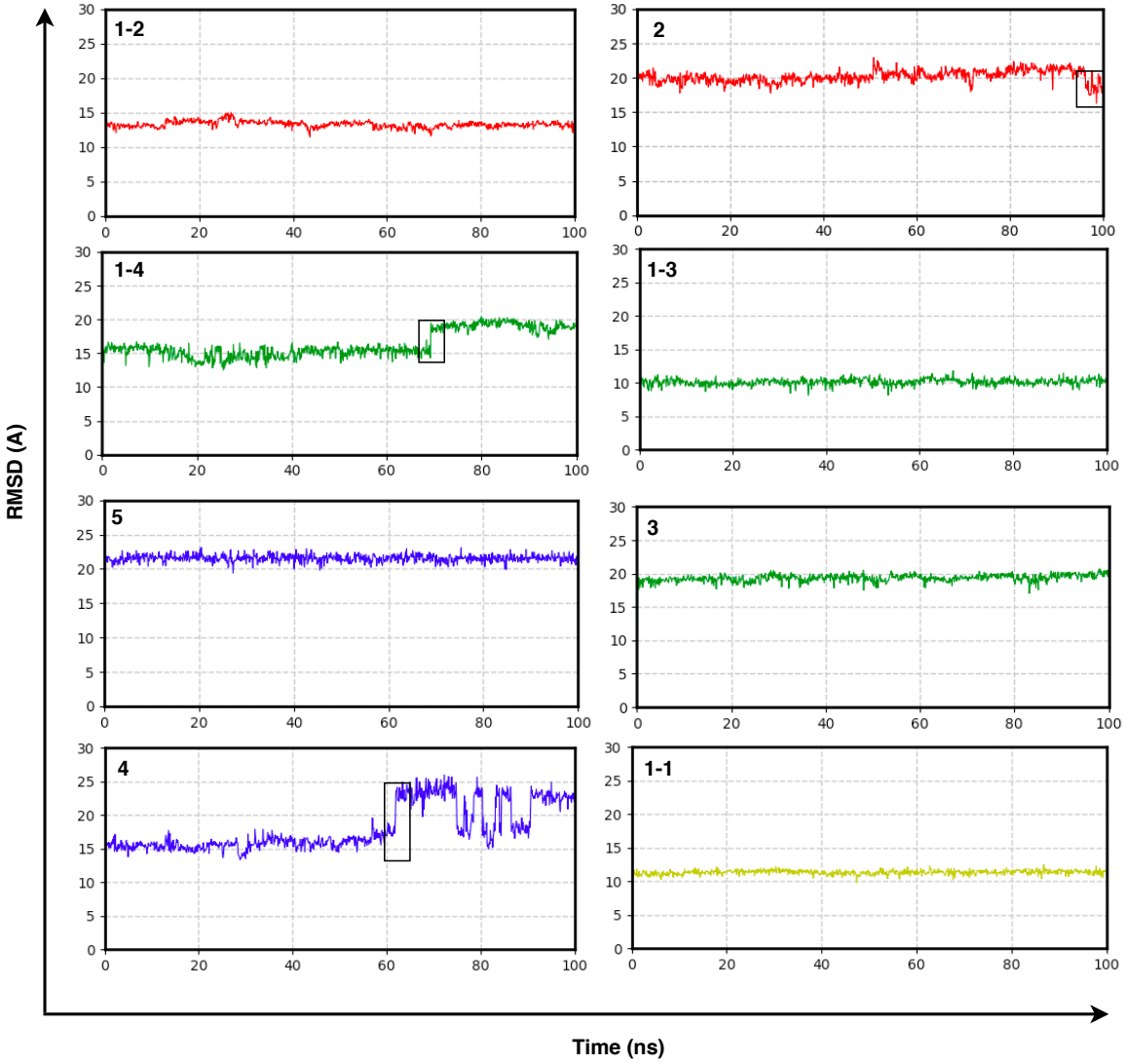


Figure 27: RMSD of p-diaminoazobenzene with aligned protein in seven cMD simulation. The black rectangles shows the shift in ligand's conformation in two simulations.

Three simulations, 1-4, 2 and 4 were discarded due to a fluctuation in RMSD value higher than 2Å, meaning that the ligand escaped from the original pocket and the binding free energy cannot be evaluated. The MMGBSA energy analysis was then performed on five trajectories.
The MMGBSA total binding energy was analyzed, arbitrarily with pocket 5, to look at the convergence of it with respect to the number of frames and time selected. Table 4 shows the total binding energy of 100ns simulation of pocket 5 with 20,50,100,200,400,500,1000 equidistant frames.

Table 4: Binding free energy computed for 100 ns taken several equidistant frames.

| | **Pocket-5** | | | | | | |
|---|---|---|---|---|---|---|---|
| | 20 frames | 50 frames | 100 frames | 200 frames | 400 frames | 500 frames | 1000 frames |
| **100ns** | -21.3420 | -21.9845 | **-22.4250** | -22.3540 | -22.5630 | -22.4330 | -22.6540 |

The free energy difference between 1000 frames and 100 frames is 0.2 kcal/mol, which for the purpose of this research can be neglected, optimizing the computational cost. Table 5 below shows the total binding free energy for 100 frames taken equidistantly in 10,20,40,60,80,100ns.

Table 5: Total binding energy for 100 frames taken in several equidistant timeframes.

| | **Pocket-5** | | | | | |
|---|---|---|---|---|---|---|
| | 10ns | 20ns | 40ns | 60ns | 80ns | 100ns |
| **100 frames** | -22.1050 | -22.3620 | -22.3670 | -22.2110 | -22.3440 | **-22.4250** |

The MMGBSA method is "frame dependent", meaning that the total energy of one frame in a favorable conformation could also give a lower energy than 1000 frames, fully avoiding dynamical effects. For such a reason, if the ligand remains bound in a distinct pose throughout the whole simulation, the total binding energy should be approximately the same independently of the time frame. All the calculations were run with 100 frames taking one frame at each 1 ns in 100ns simulation to account for dynamical effects throughout the entire simulation. From each production run, the total binding free energy of interaction between the channel and the ligand was obtained,

and shown in Table 6,

Table 6: Binding free energy of 100 ns of cMD production run

| **kcal/mol** | **Pocket(1-1)** | **Pocket(1-2)** | **Pocket(1-3)** | **Pocket(3)** | **Pocket(5)** |
|---|---|---|---|---|---|
| $\Delta G_{vdW}$ | -32.3099 | -28.4981 | -27.6722 | -32.0718 | -31.7602 |
| $\Delta G_{el}$ | -4.8541 | -6.4072 | -2.6470 | -8.6898 | -2.4325 |
| $\Delta G_{pol}$ | 15.6439 | 17.0365 | 12.4957 | 26.8407 | 15.9971 |
| $\Delta G_{np}$ | -4.2619 | -4.0657 | -3.5377 | -4.7405 | -4.2619 |
| $\Delta G_{tot}$ | **-25.7820** | -21.9345 | -21.3613 | -18.6614 | -22.4250 |

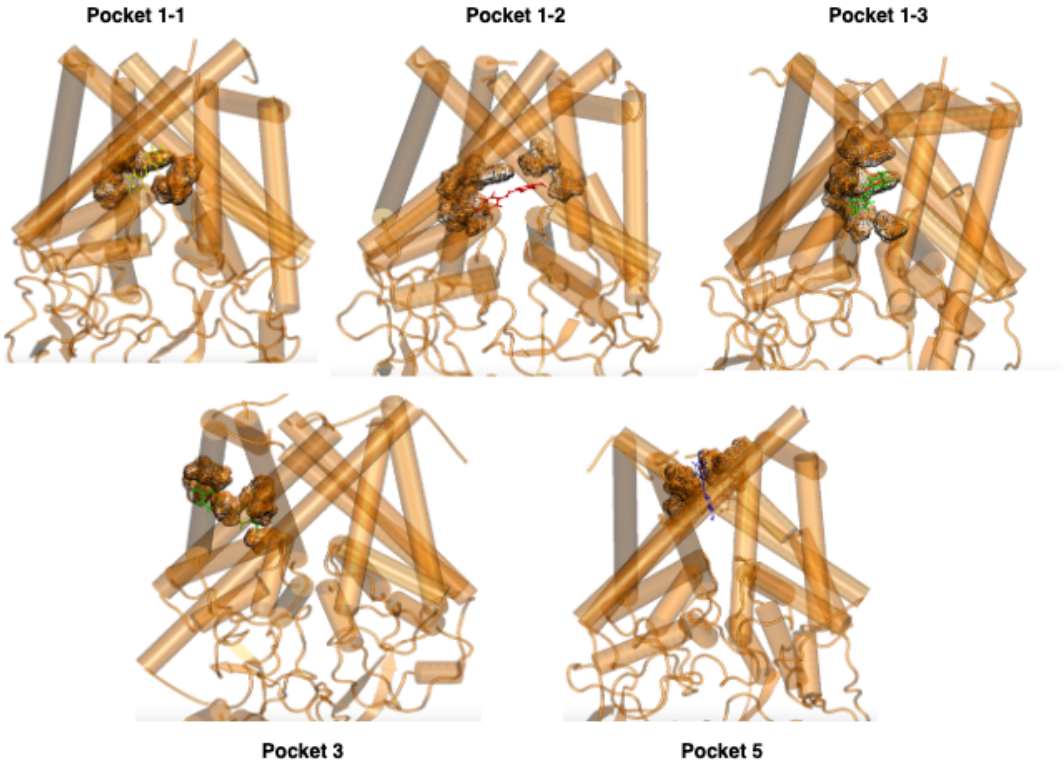


Figure 28: Cartoon representation of the five conformations found with the energy maps. Pocket 1-1,1-2 and 1-3 correspond to the same binding pocket but with a different orientation.

The pocket 1-1 results in the lowest $\Delta G_{tot}$ followed by pocket 5, pocket 1-2 and 1-3, which are two different conformations inside pocket 1 and finally pocket 3. The simulations with the longest time in a conformation, shown by RMSD, result in the lowest binding free energy, giving an indication that the MMGBSA energy analysis is well suited form this system. The most important

contribution to the binding energy is the van der Waals interaction between the protein and the photoswitch. The ligand/protein electrostatic interactions and the polar part of the solvation energy also contribute favourably to the binding energy. The non polar contribution of solvation is always positive, i.e., non favourable.

A contact analysis was performed to identify the native and non-native contacts within a radius of 5Å from the ligand. Residues present within this radius in at least 2% of the frames were selected for the decomposition analysis. The MMGBSA method allows to further decompose the energy into pairwise-residue interactions. The decomposition was performed in each simulation and the residues with the highest contributions to be binding were identified and reported. Figure 29 below show a graphical representation of the five ligand's binding poses in the channel.

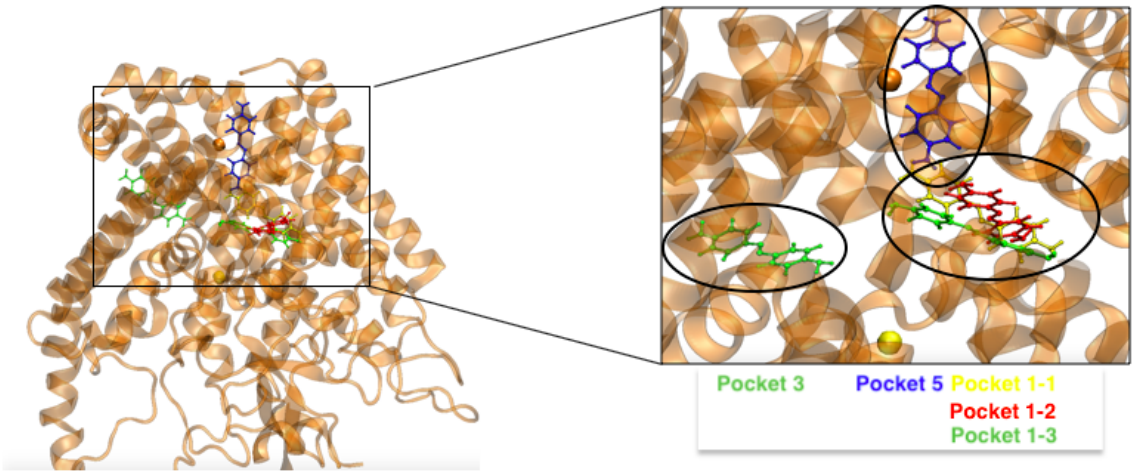


Figure 29: 3D representation of the Na$v$1.4 ion channel and the five ligand's poses. Three main pockets are circled in black. Three binding poses, 1-1,2-1 and 4-1 share the same pocket. One pocket, 3-1 is adjacent to the gate and one pocket accommodates the ligand between two domains.

Three binding poses, 1-1, 1-2 and 1-3 share the same pocket, but the total binding energy and the energy decomposition analysis shows different values and residues involved in the mechanism. Thus, a total of three pockets were identified; one pocket in the inner region of the protein sharing three distinct conformations 1-1, 1-2 and 1-3, one pocket pointing out of the transmembrane protein between two subunits, 3, and one pocket adjacent to the gate, 5, as shown in figure 29.

Table 7 below shows the free energy decomposition of pocket 1-1.

Table 7: Pairwise energy decomposition of p-diaminoazobenzene and the neighbouring residues. The six major interactions are displayed in bold with Lig587-Tyr571 being the highest contribution the binding free energy

| **Pocket 1-1** | $\Delta G_{vdW}$ | $\Delta G_{el}$ | $\Delta G_{pol}$ | $\Delta G_{np}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| Lig587-Phe154 | -0.4394 | -0.1127 | 0.1821 | -0.2917 | -0.6617 |
| Lig587-Phe284 | -0.4522 | -0.1058 | 0.0526 | -0.4291 | -0.9346 |
| **Lig587-Thr392** | -1.0995 | -0.2970 | -0.0180 | -0.8538 | **-2.2684** |
| **Lig587-Phe393** | -1.2543 | -0.0986 | 0.2544 | -0.9000 | **-1.9987** |
| Lig587-Leu437 | -0.8141 | -0.0131 | 0.1293 | -0.6856 | -1.3835 |
| Lig587-Ile517 | -0.6507 | -0.0337 | -0.1715 | -0.4391 | -1.2952 |
| **Lig587-Thr518** | -1.2463 | -0.0972 | -0.4544 | -1.0523 | **-2.8504** |
| Lig587-Ser520 | -0.4069 | 0.2387 | -0.1347 | -0.3731 | -0.6760 |
| Lig587-Ile567 | -0.7842 | 0.0182 | -0.0286 | -0.6242 | -1.4189 |
| **Lig587-Ser570** | -0.5364 | -1.4696 | -0.2595 | -0.5488 | **-2.8145** |
| **Lig587-Phe571** | -2.3029 | -0.3000 | 0.2546 | -1.5719 | **-3.9203** |
| **Lig587-Val574** | -1.2544 | -0.1030 | 0.1559 | -1.0806 | **-2.2821** |
| Lig587-Val575 | -0.6145 | -0.0922 | 0.1446 | -0.3578 | -0.9199 |

The major contribution to the non-covalent binding of the ligand is given by vdW interactions between p-diaminoazobenzene and Phe571, and likely, of hydrophobic nature. The benzyl groups creates VdW interactions also with Thr392, Phe393, Thr518 and Val574 while the amine group acts as hydrogen donor with Ser570.

Table 8 shows the pairwise energy decomposion of the ligand in pocket 1-2.

Table 8: Pairwise energy decomposition of p-diaminoazobenzene and the neighbouring residues. The six major interactions are displayed in bold with Lig587-Phe571 being the highest contribution to the binding free energy.

| **Pocket 1-2** | $\Delta G_{vdW}$ | $\Delta G_{el}$ | $\Delta G_{pol}$ | $\Delta G_{np}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| Lig587-Trp321 | -0.4587 | 0.0363 | 0.0620 | -0.3469 | -0.7072 |
| Lig587-Leu388 | -0.8488 | 0.0453 | 0.0834 | -0.6220 | -1.3420 |
| Lig587-Ala391 | -0.6278 | -0.1950 | 0.0960 | -0.5911 | -1.3179 |
| Lig587-Thr392 | -0.9184 | -0.1940 | 0.1731 | -0.7431 | -1.6825 |
| **Lig587-Ser433** | -0.3869 | -0.9526 | -0.2687 | -0.4819 | **-2.0903** |
| **Lig587-Leu437** | -1.0799 | 0.07947 | 0.0287 | -0.9702 | **-1.9419** |
| Lig587-Ile517 | -0.2693 | -0.0433 | -0.1290 | -0.1518 | -0.5935 |
| **Lig587-Thr518** | -0.6348 | -0.4723 | -0.0859 | -0.7543 | **-1.9474** |
| Lig587-Ser520 | -0.3476 | 0.1067 | -0.0288 | -0.3174 | -0.5872 |
| **Lig587-Ile567** | -1.0764 | -0.2187 | 0.0875 | -0.9370 | **-2.1447** |
| **Lig587-Ser570** | -0.4429 | -1.0646 | -0.3038 | -0.4559 | **-2.2675** |
| **Lig587-Phe571** | -2.4720 | 0.1090 | 0.1845 | -1.7693 | **-3.9478** |
| Lig587-Val574 | -0.3669 | -0.0642 | 0.0385 | -0.3102 | -0.7028 |
| Lig587-Tyr578 | -0.1736 | -0.1025 | 0.0358 | -0.1465 | -0.3868 |

The major contribution to the non-covalent binding of the ligand is given by hydrophobic interactions between ligand and Phe571. The benzyl groups creates VdW interactions also with Leu437, Ile567 and Thr518 while the amine group acts as hydrogen donor with Ser433 and Ser570 and the diazine group as hydrogen acceptor with Thr392 and Thr518.

Table 9 below shows the free energy decomposition of pocket 1-3,

Table 9: Pairwise energy decomposition of p-diaminoazobenzene and the neighbouring residues. The six major interactions are displayed in bold with Lig587-Phe571 being the highest contribution to the binding free energy.

| **Pocket-1-3** | $\Delta G_{vdW}$ | $\Delta G_{el}$ | $\Delta G_{pol}$ | $\Delta G_{np}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| **Lig587-Thr392** | -0.8952 | 0.2593 | -0.2922 | -0.8442 | **-1.7722** |
| Lig587-Phe393 | -0.2290 | 0.0881 | -0.0364 | -0.1315 | -0.3089 |
| Lig587-Leu437 | -0.4739 | 0.0405 | 0.0198 | -0.5338 | -0.9473 |
| Lig587-Ile517 | -0.5951 | 0.5754 | -0.3942 | -0.4366 | -0.8506 |
| Lig587-Thr518 | -0.5866 | -0.0111 | -0.0291 | -0.3786 | -1.0056 |
| Lig587-Ser520 | -0.7118 | 0.1406 | 0.0903 | -0.6485 | -1.1294 |
| Lig587-Trp523 | -0.2440 | -0.2593 | 0.2442 | -0.1613 | -0.4205 |
| **Lig587-Ile567** | -0.7649 | -0.7431 | -0.0231 | -0.6945 | **-2.2257** |
| **Lig587-Ser570** | -0.6584 | -1.2337 | -0.4888 | -0.5624 | **-2.9434** |
| **Lig587-Phe571** | -2.6794 | 0.1383 | 0.0458 | -1.8415 | **-4.3368** |
| **Lig587-Val574** | -1.2059 | -0.2097 | 0.0786 | -1.0280 | **-2.3650** |
| Lig587-Val575 | -0.6929 | -0.1954 | 0.2667 | -0.4250 | -1.046 |
| **Lig587-Tyr578** | -0.9384 | -0.4459 | 0.1053 | -0.9011 | **-2.180** |

Also in this case the major contribution to the non-covalent binding of the ligand is given by hydrophobic interactions between the ligand and Phe571. The benzyl group also strongly interacts via vdW interaction with Thr392, Val574 and Tyr578 and acts as H-bond donor with Ile567 and Ser570.

Table 10 show the free energy decomposition of pocket-3,

Table 10: Pairwise energy decomposition of p-diaminoazobenzene and the neighbouring residues.The six major interactions are displayed in bold with Lig587-Leu17 being the highest contribution the binding free energy.

| **Pocket-3** | $\Delta G_{vdW}$ | $\Delta G_{el}$ | $\Delta G_{pol}$ | $\Delta G_{np}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| **Lig587-Leu17** | -1.3030 | 0.4912 | 0.3632 | -1.1242 | **-3.0617** |
| Lig587-Met21 | -0.8405 | 0.1304 | -0.1327 | -0.6061 | -1.4489 |
| **Lig587-Thr24** | -0.6752 | -0.3736 | -0.1210 | -0.6469 | **-1.8169** |
| Lig587-Phe118 | -0.8557 | 0.0274 | 0.2158 | -0.5220 | -1.1345 |
| **Lig587-Thr122** | -0.9424 | -0.6811 | 0.2372 | -0.9146 | **-2.3010** |
| **Lig587-Leu156** | -1.1585 | 0.0828 | 0.0684 | -1.030 | **-2.0382** |
| Lig587-Ile157 | -0.5685 | 0.1283 | -0.0543 | -0.5506 | -1.0452 |
| Lig587-Ile160 | -1.0473 | -0.1217 | 0.2825 | -0.8322 | -1.7187 |
| Lig587-Met273 | -0.8827 | -0.1659 | 0.0990 | -0.7162 | -1.6657 |
| **Lig587-Val274** | -1.2221 | -0.0116 | 0.0322 | -1.0668 | **-2.2684** |
| **Lig587-Asn277** | -1.1090 | -0.7900 | 0.1332 | -0.9779 | **-2.7438** |

The major contributor to the non-covalent binding of the ligand is given by hydrophobic interactions with Leu17. The benzyl groups also interacts via vdW interactions with Leu156, Val275 and Asn277 and acts as H-bond acceptor with Asn277 and donor with Leu17 and Thr122.

Table 11 below shows the free energy decomposition of pocket-5,

Table 11: Pairwise energy decomposition of p-diaminoazobenzene and the neighbouring residues. The six major interactions are displayed in bold with Lig587-Tyr578 being the highest contribution to the binding free energy.

| **Pocket-5** | $\Delta G_{vdW}$ | $\Delta G_{el}$ | $\Delta G_{pol}$ | $\Delta G_{np}$ | $\Delta G_{tot}$ |
|---|---|---|---|---|---|
| Lig587-Phe284 | -0.6917 | 0.0602 | -0.0057 | -0.5501 | -1.1874 |
| Lig587-Leu285 | -0.8391 | 0.1087 | 0.0711 | -0.7436 | -1.4028 |
| **Lig587-Leu288** | -1.1319 | -0.1704 | 0.2855 | -0.7417 | **-1.7585** |
| Lig587-Leu289 | -0.5673 | -0.0009 | 0.1321 | -0.5321 | -0.9682 |
| Lig587-Phe292 | -0.6574 | -0.0521 | 0.1084 | -0.5477 | -1.1488 |
| **Lig587-Ile441** | -1.0475 | -0.1091 | 0.1758 | -0.9147 | **-1.8957** |
| **Lig587-Ile445** | -1.4021 | -0.2214 | 0.3949 | -1.1207 | **-2.3494** |
| Lig587-Phe571 | -0.5102 | 0.1218 | -0.0576 | -0.4112 | -0.8573 |
| **Lig587-Val574** | -0.8511 | 0.2539 | -0.2799 | -0.7680 | **-1.6452** |
| **Lig587-Val575** | -1.4385 | 0.2227 | -0.0721 | -0.9881 | **-2.2761** |
| **Lig587-Tyr578** | -2.9685 | -0.3715 | 0.4207 | -1.8345 | **-4.7540** |
| Lig587-Ile579 | -0.8748 | -0.0171 | 0.1518 | -0.8129 | -1.1058 |
| Lig587-Ile583 | -0.5779 | -0.0847 | 0.2301 | -0.6732 | -1.4562 |

The major contribution to the non-covalent binding of the ligand is given by hydrophobic interactions between the benzyl groups of p-diaminoazobenzene and Tyr578. The ligand also interacts via vdW interactions with Ile441 Ile445 and Val575.
Within the five conformations, the ligand has always favorable van der Walls interactions with hydrophobic side chain aminoacids as phenylalanine and tyrosine (aromatics), and valine and isoleucine (aliphatic) and both van der Walls and electrostatic interactions with polar side chains aminoacids as Serine and threonine. The three conformations in pocket 1 have the strongest interactions with Phe571 and Ser570 but in pocket 1-1, the vdW interaction with Thr518 stabilizes the ligand yielding to the most favorable conformation. Pocket 1-1 was selected for additional QM/MM calculation on the basis of the total binding free energy being the lowest in the energy analysis.

## 4.3 cMD-GaMD Sampling Comparison

GaMD provides enhanced sampling of the binding pocket search for p-diaminoazobenzene in Na$v$1.4 channel. Four parallel cMD simulations of 1$\mu$s were performed from the same starting structure as for the GaMD simulations. An energy profile was performed, superimposing the total 4$\mu$s of cMD and 4$\mu$s of GaMD to compare the sampling between the two methods, as shown in Figure 30.

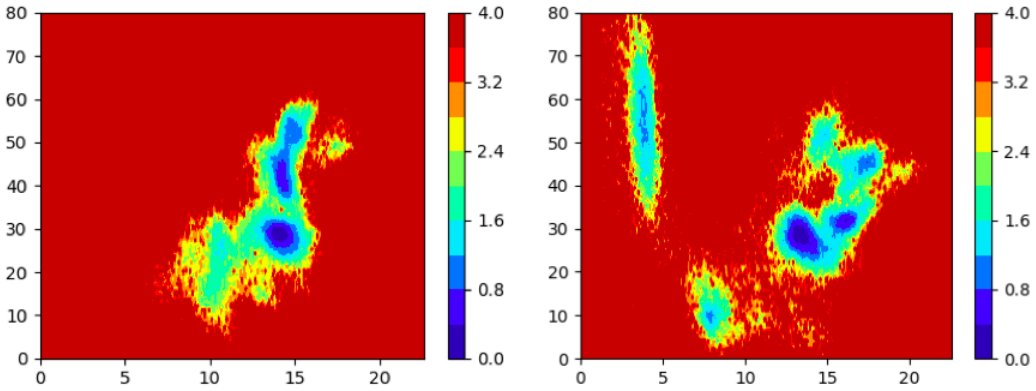


Figure 30: Energy profile of p-diaminoazobenzene CoM in 4$\mu$s of cMD (left) and 4$\mu$s of GaMD (right). The x-axis is the distance between the CoM of the $\alpha$ carbons of the gate and the CoM of p-diaminoazobenzene,and the y-axis corresponds to the inner product between two vectors, both starting at the gate CoM pointing at the DEKA ligand's CoMs.

The GaMD simulation allows to find an additional pocket at (5Å, 50°-60°), part of the phase space which was not sample in any of the cMD simulations. The applied boost potential allowed a jump over an high energy barrier allowing the ligand to find an additional pocket inside the channel compared to the cMD simulation. However, the cMD map has sampled a space at (10Å, 10°-30°) which was not sampled in the GaMD simulations. Both simulations have extensively sampled a similar part of the phase space, confirming the presence of the most stable pocket at (15Å, 30°). With this data, it is not possible to state that the GaMD has extensively sample the space more than cMD but an addition pocket was obtained which was not present in the cMD map. Recently, a new computational method based on GaMD, Ligand accelerated GaMD (LiGaMD) was developed to specifically simulate protein-ligand binding and unbinding processes. LiGaMD selectively boosts the ligand non-bonded interaction potential energy and enables improved enhanced sampling simulations over the previous GaMD[110]. This new method is in active development and will be tested in next simulations.

## 4.4 Quantum Chemical Calculations

### 4.4.1 Vertical Excitations of Azobenzene in Polarizable Continuum Model

The vertical excitation energies of p-diaminoazobezene inside pocket 1-1 are computed, including Phe571, the highest contributor to the binding of the ligand, in the QM region. Due to the size of the system, DFT is employed because of its computational efficiency over wavefunction-based methods. In one of the most extensive comparison of 614 ES and 29 functionals, published in 2009[111], pure local functionals as LDA or GGA were found to underestimate the values of $\lambda_{max}$ while range separated hybrid functionals as CAM-B3LYP and $\omega$B97X-D and LC-$\omega$PBE lead to the opposite error, overestimating the values of $\lambda_{max}$[111]. It was concluded that the $n \rightarrow \pi^*$ transitions are rather insensitive to the selected functional and tends to be more precisely described than the $\pi \rightarrow \pi^*$ transition. Another extensive ES-$\lambda^{max}$ benchmark has been also published, focussing on "real life" dyes and reporting B3LYP being the most accurate functional for low lying excited states with $n \rightarrow \pi^*$ (0.09eV) and $\pi \rightarrow \pi^*$ (0.24eV) of mean absolute error (MAE). Finally another comprehensive study has compared the performance of global and range separated hybrid functionals in organic dyes including also azobenzene. The study has shown that in the $n \rightarrow \pi^*$ transitions of nitroso and thiocarbonyl compounds, all hybrid functionals give similar vertical energies while in the $n \rightarrow \pi^*$ of 15 anthraquinones, range separated hybrid functionals are further away from the experimental $\lambda_{max}$ then their global counterpart. Range separated functionals as CAM-B3LYP and $\omega$B97x-D have improved accuracy for the description of Rydberg excitations and significant improvements on the charge transfer description[97, 112].
The geometry of trans-azobenzene was optimized at B3LYP/cc-pvdz level of theory. From the optimized structure, the singlet vertical excitations of the trans isomer were calculated in PCM . The first transition n $\rightarrow \pi^*$ is symmetry forbidden and therefore has a very weak oscillator strength, while the second transition, $\pi \rightarrow \pi^*$ is much more intense. Table 12 below shows the excitation energies for these two transitions.

Table 12: Excitation energies (in eV) of $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions for azobenzene in PCM. The oscillator strengths are shown in parentheses. [a] Ref. [113]

| | **Azobenzene** | | | |
|---|---|---|---|---|
| | B3LYP-D3 | CAM-B3LYP-D3 | wB97X-D | exp[a] |
| n $\rightarrow \pi^*$ | 2.65(0.00) | 2.82(0.00) | 2.82(0.00) | 2.79 |
| $\pi \rightarrow \pi^*$ | 3.76(1.18) | 4.06(1.17) | 4.07(1.17) | 3.95 |

The vertical transition energies calculated with the two range separated functionals (CAM-B3LYP-D3/cc-pvdz and $\omega$B97X-D/cc-pvdz) are fairly close to the experimental values with a slight overestimation of the value in both transitions. The global hybrid functional B3LYP-D3/cc-pvdz underestimates the values of both transitions. The symmetry was assigned by visual inspection of the molecular orbitals (MOs). Evaluation of the MOs shows that the first transition originates

from the lone pair on central nitrogens while the second transition is delocalized in the entire molecule as shown in Figure 31,

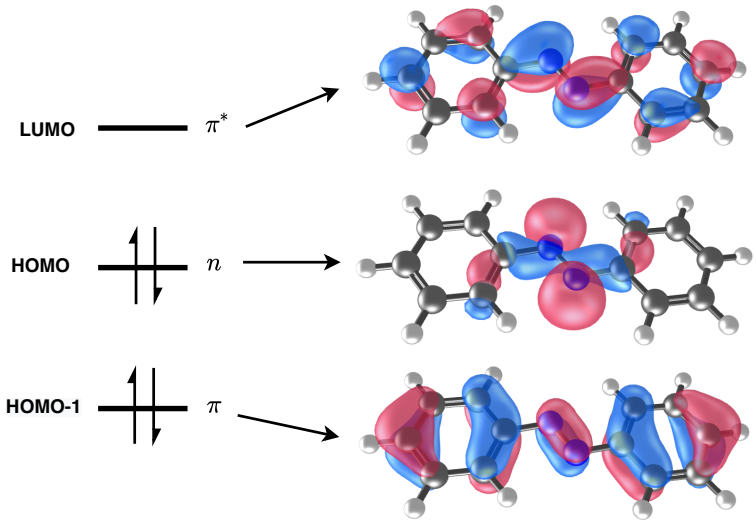


Figure 31: MO representation of the $\pi$, $n$ and $\pi^*$ orbitals.

### 4.4.2 Vertical Excitations of P-diaminoazobenzene in Polarizable Continuum Model

The geometry of trans-p-diaminoazobenzene was optimized at B3LYP/cc-pvdz level of theory in the same manner of the unsubstituted azobenzene and the singlet vertical excitations were calculated. Also in the amino derivatives, the first transition n $\rightarrow \pi^*$ is symmetry forbidden and therefore has a very weak oscillator strength, while the second transition, $\pi \rightarrow \pi^*$ is much more intense. Table 13 below shows the excitation energies for this two transitions,

Table 13: Excitation energies of n $\rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions for p-diaminoazobenzene in PCM. The oscillator strengths are shown in parentheses. [b] Ref. [31]

| | **P-diaminoazobenzene** | | | |
|---|---|---|---|---|
| | B3LYP-D3 | CAM-B3LYP-D3 | wB97X-D | exp[b] |
| n $\rightarrow \pi^*$ | 2.89(0.00) | 3.05(0.00) | 3.02(0.00) | - |
| $\pi \rightarrow \pi^*$ | 3.06(1.70) | 3.32(1.60) | 3.34(1.57) | 3.16 |

The global hybrid functional B3LYP-D3 gives a close estimation to the experimental value with an underestimation of 0.1 eV while the the range separated functionals give an overestimation of 0.16-0.18 eV. Also for p-diaminoazobenzene, the symmetry of the MOs was assigned by visual inspection. The first transition originates from the lone pair on central nitrogens while the second transition is delocalized in the entire molecule as shown in Figure 32,

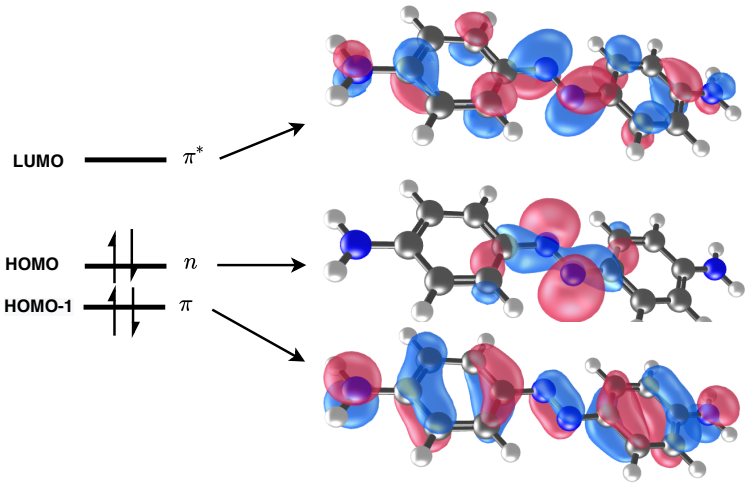


Figure 32: MO representation of the $\pi$, $n$ and $\pi^*$ orbitals.

The presence of electron donating groups at the para position of the benzyl rings creates a large red shift in the $\pi \rightarrow \pi^*$ absorption, almost leading to an overlap between the two transitions. On the basis of a qualitative argument, an increase in the average electron density of the antibonding state, reduces the effective bond order of the N=N bond and thus reduces the barrier height of isomerization. With a relation to the molecular orbitals, the ground electronic state of azobenzene, where the HOMO is doubly occupied, is characterized by a bond order of ∼2 for the N=N bond. The first excited state where the LUMO is occupied by one electron, is characterized instead by a reduction of the electron density and a lower bond order for N=N. Tables 14 and 15 below show the difference in vertical excitation energies between azobenzene and p-diaminoazobenzene at B3LYP-D3/cc-pvdz and CAM-B3LYP-D3/cc-pvdz level of theory.

Table 14: Comparison of excitation energies (in eV) of $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions for azobenzene and p-diaminoazobenzene in PCM.

| | **B3LYP-D3** | | |
|---|---|---|---|
| | $n \rightarrow \pi^*$ | $\pi \rightarrow \pi^*$ | $(\pi \rightarrow \pi^*)$ - $(n \rightarrow \pi^*)$ |
| **Azobenzene** | 2.65 | 3.76 | 1.11 |
| **P-diaminoazobenzene** | 2.89 | 3.06 | 0.17 |

Table 15: Comparison of excitation energies of n $\rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions for azobenzene and p-diaminoazobenzene

| | **CAM-B3LYP-D3** | | |
|---|---|---|---|
| | $n \rightarrow \pi^*$ | $\pi \rightarrow \pi^*$ | $(\pi \rightarrow \pi^*)$ - $(n \rightarrow \pi^*)$ |
| **Azobenzene** | 2.82 | 4.06 | 1.24 |
| **P-diaminoazobenzene** | 3.05 | 3.32 | 0.27 |

Comparing the n $\rightarrow \pi^*$ of azobenzene and p-diaminoazobenzene in both B3LYP-D3 and CAM-B3LYP-D3, the energy gap increases of 0.23 eV with the addition of the amino substituents. In the $\pi \rightarrow \pi^*$ absorption instead, the presence of the amino groups, induces a larger red shift of $\sim$ 0.7 eV in both B3LYP and CAM-B3LYP. The smaller energy gap between the states, makes the amino derivative more suitable to be used as cis/trans photoswitch.

### 4.4.3 Vertical excitations of p-diaminoazobenzene in the Na$v$1.4 channel

The vertical transition energies to the valence excited states were computed in a QM/MM calculation with p-diaminoazobenzene and Phe571 included in the QM region. The binding pocket with the lowest total binding free energy was selected and the coordinates of the entire system were extracted. From the energy decomposition analysis, Phe571 was selected and included in the QM region with p-diaminoazobenzene since they presented the strongest interaction.

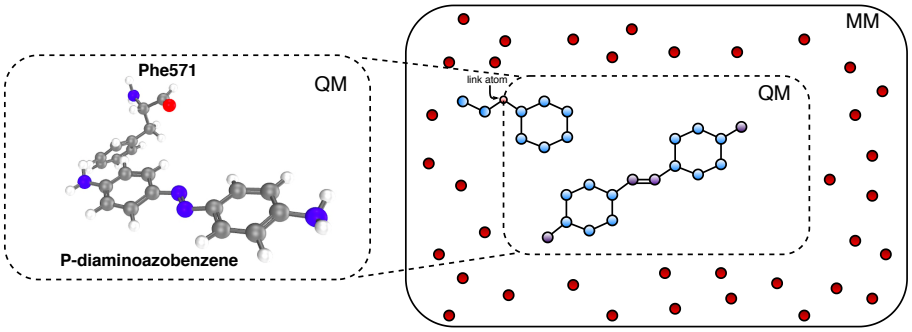


Figure 33: Graphical representation of the QM/MM regions. The QM region incorporated p-diaminoazobenzene and Phe571 where between the N-C peptide peptide bond, a link H atom was added in the boundary between the two regions. In the MM region, the red circle represents the surrounding (water, ions,protein,membrane) in point charges which are included in the QM Hamiltonian as one-electron operator terms.

From 20 equidistant snapshot of the 100ns cMD trajectory, a singlet vertical excitation of the QM region was calculated in the presence of the MM environment at B3LYP-D3/cc-pvdz, CAM-B3LYP-D3/cc-pvdz and $\omega$B97X-D/cc-pvdz level of theory. The graphs below show the gaussian convoluted vertical excitation energies for the 20 geometries with B3LYP-D3, CAM-B3LYP-D3 and $\omega$B97X-D.

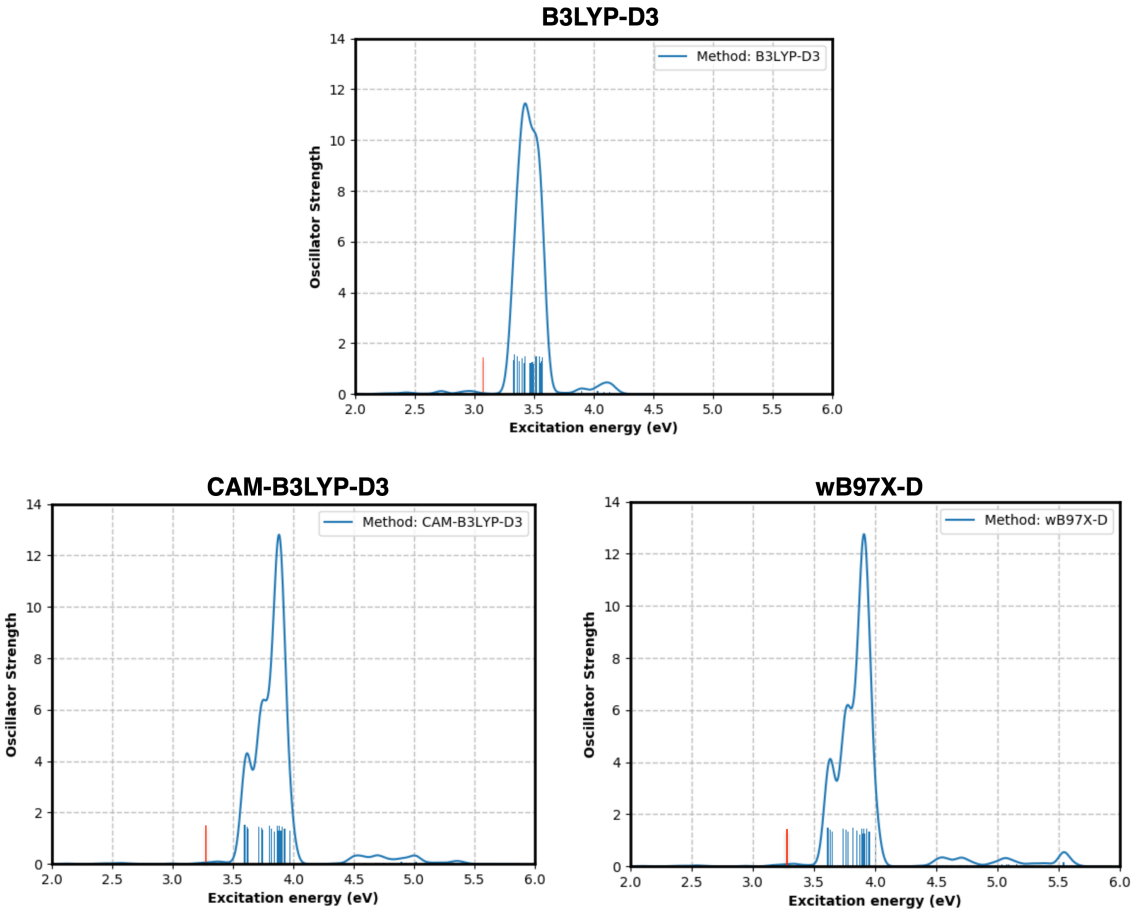


Figure 34: Gaussian fitted vertical excitation energies of 20 geometries of GaMD-4 binding pocket with p-diaminoazobenzene and Phe571 in the QM region. The red line represents the vertical excitation energy of p-diaminoazobenzene in PCM.

The two range separated functionals show a similar behaviour in both the excitation energies and the oscillator strength, with an average value of 3.85 eV. B3LYP-D3, presents a lower excitation energy of 3.45eV. B3LYP-D3 presents a shift of $\sim 0.40$ eV towards blue light while CAM-B3LYP-D3 and $\omega$B97X-D present a shift of $\sim 0.5$ eV also towards blue light, when the calculation is performed with the inclusion of the MM environment. The shift is due to the environment but also caused by the vibrational motion which is not present in the PCM calculations. Additional calculations will be performed to investigate in more detail the effect of the environment and to characterize better the nature of the excited states.

# 5 Conclusions

The ability of photoswitches to have a structural and functional change upon irradiation, makes them important compounds to be used in photopharmacology. Ion channels are ideal targets for photoswitches since the transport of ions through these membrane proteins can be regulated with light. In this work, the interaction of a molecular photoswitch in a human's brain $Na_v1.4$ channel was investigated using conventional and accelerated MD to search for suitable binding pockets for p-diaminoazobenzene. An end state free energy method was employed to characterize the binding free energy of the ligand and the pairwise decomposition energy of its interactions with surrounding residues. Finally, the vertical excitation energies were computed with TD-DFT at PCM and QM/MM models with B3LYP-D3/cc-pvdz, CAM-B3LYP/cc-pvdz and *omega*B97X-D/cc-pvdz level of theory. The MMGBSA free energy analysis shows the presence of three distinct binding pockets, near the intracellular gate, the DEKA filter and in intercalation between two subunits of the $\alpha$ domain. The major contribution to the non-covalent binding of the ligand is given by hydrophobic interactions with the non-polar side chain aminoacids as phenylalanine, threonine and tyrosine. Comparison of cMD and GaMD sampling has shown that GaMD allows to find an additional pocket, pocket 5, but also the cMD simulation has sampled a space which was not sampled by GaMD, thus, it is not possible to state that GaMD has enhanced the sampling. The vertical transition energies of azobenzene in PCM are very close to the experimental value of $\lambda_{max}$ with both global and range separated functionals. In p-diaminoazobenzene, the addition of electron donating groups at the para position creates a large red shift in the $\pi \rightarrow \pi^*$ absorption contributing to a smaller gap with the $n \rightarrow \pi^*$. The addition of the MM environment and Phe571 in the QM region has shown a blue shift of 0.4-0.5 eV for all three functionals compared to the PCM calculation of p-diaminoazobenzene. In further investigations, the vertical excitation energy of only p-diaminoazobenzene in the QM region will be computed to differentiate the effect of the MM environment from the presence of Phe571 in the QM region and study the extent of charge transfer occurring in this reaction. Further calculations will be performed to investigate in more detail the effect of the environment and to characterize the nature of the excites states.

# References


[1] Marie A. Chisholm-Burns. *Pharmacotherapy principles & practice*. McGraw-Hill Medical, New York, 2008.

[2] Overington JP, Al-Lazikani B, and Hopkins AL. How many drug targets are there? 5(12):993–6, 2006.

[3] Edwards IR and Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. 356(9237):1255–9, 2000.

[4] Lehár J, Krueger AS, Avery W, Heilbut AM, Johansen LM, Price ER, Rickles RJ, Short GF 3rd, Staunton JE, Jin X, Lee MS, Zimmermann GR, and Borisy AA. Synergistic drug combinations tend to improve therapeutically relevant selectivity. 27(7):659–66, 2009.

[5] DiMasi JA, Feldman L, Seckler A, and Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. 87(3):272–7, 2010.

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

[7] Carlet J, Collignon P, Goldmann D, Goossens H, Gyssens IC, Harbarth S, Jarlier V, Levy SB, N'Doye B, Pittet D, Richtmann R, Seto WH, van der Meer JW, and Voss A. Society's failure to protect a precious resource: antibiotics. 378(9788):369–71, 2011.

[8] Hopkins AL. Network pharmacology: the next paradigm in drug discovery. 4(11):682–90, 2008.

[9] Szymański W, Beierle JM, Kistemaker HA, Velema WA, and Feringa BL. Reversible photocontrol of biological systems by the incorporation of molecular photoswitches. 113(8):6114–78, 2013.

[10] Beharry AA and Woolley GA. Azobenzene photoswitches for biomolecules. 40(8):4422–37, 2011.

[11] 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.

[12] Hüll K, Morstein J, and Trauner D. In vivo photopharmacology. 118(21):10710–10747, 2018.

[13] Alexandre Mourot, Timm Fehrentz, Yves Le Feuvre, Caleb M Smith, Christian Herold, Deniz Dalkara, Frédéric Nagy, Dirk Trauner, and Richard H Kramer. Rapid optical control of nociception with an ion-channel photoswitch. *Nature Methods*, 9(4):396–402, 2012.

[14] Matthew R Banghart, Alexandre Mourot, Doris L Fortin, Jennifer Z Yao, Richard H Kramer, and Dirk Trauner. Photochromic blockers of voltage-gated potassium channels. *Angewandte Chemie (International ed. in English)*, 48(48):9097–9101, 2009.

[15] Andrew A. Beharry and G. Andrew Woolley. Azobenzene photoswitches for biomolecules. *Chem. Soc. Rev.*, 40:4422–4437, 2011.

[16] Linda E. Watson. Azobenzene : aspects, applications, and research. 2017.

[17] Burdette S. C. and Bandara H. M. D. Photoisomerization in different classes of azobenzene. 41(5):1809–1825, 2012.

[18] C. J. Brown. A refinement of the crystal structure of azobenzene. 21(1):146–152, 1966.

[19] Crecca CR and Roitberg AE. Theoretical study of the isomerization mechanism of azobenzene and disubstituted azobenzene derivatives. 110(26):8188–203, 2006.

[20] Burdette S. C. and Bandara H. M. D. Photoisomerization in different classes of azobenzene. 41(5):1809–1825, 2012.

[21] Petr Klán and Jakob Wirz. Photochemistry of organic compounds : from concepts to practice. 2009.

[22] Tatsuya Fujino, Sergei Yu. Arzhantsev, and Tahei Tahara. Femtosecond time-resolved fluorescence study of photoisomerization of trans -azobenzene. 105(35):8123–8129, 2001.

[23] William M. Horspool and Pill-Soon. Song. *CRC handbook of organic photochemistry and photobiology*. CRC Press, Boca Raton, 1995.

[24] Banghart M, Borges K, Isacoff E, Trauner D, and Kramer RH. Light-activated ion channels for remote control of neuronal firing. 7(12):1381–6, 2004.

[25] Toyohiko Yatagai, Daisuke Barada, Masahide Itoh, and Kenji Harada. Generation of surface relief hologram and nano structure on azobenzen polymer films and its numerical analysis by moving particle method. In Yury Denisyuk, Ventseslav Sainov, and Elena Stoykova, editors, *Holography 2005: International Conference on Holography, Optical Recording, and Processing of Information*, volume 6252, pages 187 – 191. International Society for Optics and Photonics, SPIE, 2006.

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

[27] John L. Magee, William Shand, and Henry Eyring. Non-adiabatic reactions. rotation about the double bond *. 63(3):677–688, 1941.

[28] Xiaogong Wang. Azo polymers : synthesis, functions and applications. 2017.

[29] Tatsuya Fujino, Sergei Yu. Arzhantsev, and Tahei Tahara. Femtosecond/picosecond time-resolved spectroscopy of trans- azobenzene: Isomerization mechanism following s2(*) s0photoexcitation. *Bulletin of the Chemical Society of Japan*, 75(5):1031–1040, 2002.

[30] Hermann. Rau, Gerhard. Greiner, Guenter. Gauglitz, and Herbert. Meier. Photochemical quantum yields in the a (+h.nu.) .dblarw. b (+h.nu.,.delta.) system when only the spectrum of a is known. *The Journal of Physical Chemistry*, 94(17):6523–6524, 08 1990.

[31] A. A. Blevins and G. J. Blanchard. Effect of positional substitution on the optical response of symmetrically disubstituted azobenzene derivatives. *The Journal of Physical Chemistry B*, 108(16):4962–4968, 2004.

[32] Christina R. Crecca and Adrian E. Roitberg. Theoretical study of the isomerization mechanism of azobenzene and disubstituted azobenzene derivatives. *The Journal of Physical Chemistry A*, 110(26):8188–8203, 2006. PMID: 16805507.

[33] Bertil Hille. *Ion channels of excitable membranes*. Sinauer, Sunderland, Mass., 2001.

[34] 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.

[35] Xiaojing Pan, Zhangqiang Li, Qiang Zhou, Huaizong Shen, Kun Wu, Xiaoshuang Huang, Jiaofeng Chen, Juanrong Zhang, Xuechen Zhu, Jianlin Lei, Wei Xiong, Haipeng Gong, Bailong Xiao, and Nieng Yan. Structure of the human voltage-gated sodium channel nav1.4 in complex with 1. *Science*, 362(6412), 2018.

[36] William A. Catterall, Alan L. Goldin, and Stephen G. Waxman. International union of pharmacology. xlvii. nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacological Reviews*, 57(4):397–409, 2005.

[37] Xiaojing Pan, Zhangqiang Li, Qiang Zhou, Huaizong Shen, Kun Wu, Xiaoshuang Huang, Jiaofeng Chen, Juanrong Zhang, Xuechen Zhu, Jianlin Lei, Wei Xiong, Haipeng Gong, Bailong Xiao, and Nieng Yan. Structure of the human voltage-gated sodium channel nav1.4 in complex with 1. *Science*, 362(6412), 2018.

[38] Stephen C. Cannon. Physiologic principles underlying ion channelopathies. *Neurotherapeutics*, 4(2):174–183, 2007.

[39] Pierre Paoletti, Graham C. R. Ellis-Davies, and Alexandre Mourot. Optical control of neuronal ion channels and receptors. *Nature Reviews Neuroscience*, 20(9):514–532, 2019.

[40] K L Choi, R W Aldrich, and G Yellen. Tetraethylammonium blockade distinguishes two inactivation mechanisms in voltage-activated k+ channels. *Proceedings of the National Academy of Sciences*, 88(12):5092–5095, 1991.

[41] van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NF, and Yu HB. Biomolecular modeling: Goals, problems, perspectives. 45(25):4064–92, 2006.

[42] Veronica Salmaso and Stefano Moro. Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. *Frontiers in Pharmacology*, 9:923, 2018.

[43] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967.

[44] W.F. van Gunsteren and H.J.C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Molecular Physics*, 34(5):1311–1327, 1977.

[45] Stewart A. Adcock and J. Andrew McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615, 05 2006.

[46] 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 [article v1.0]. *Living Journal of Computational Molecular Science*, 1(1):5957, 11 2018.

[47] 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.

[48] Stewart A. Adcock and J. Andrew McCammon. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, 106(5):1589–1615, 05 2006.

[49] Sprenger KG, Jaeger VW, and Pfaendtner J. The general amber force field (gaff) can accurately predict thermodynamic and transport properties of many ionic liquids. 119(18):5882–95, 2015.

[50] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[51] Camley BA, Lerner MG, Pastor RW, and Brown FL. Strong influence of periodic boundary conditions on lateral diffusion in lipid bilayer membranes. 143(24):243113, 2015.

[52] Jiangang Chen and K. N. Houk. Molecular modeling: principles and applications by andrew r. leach. addison wesley longman limited: essex, england, 1996. 595 pp. isbn 0-582-23933-8. $35. *Journal of Chemical Information and Computer Sciences*, 38(5):939–939, 1998.

[53] Jan von Plato. Boltzmann's ergodic hypothesis. *Archive for History of Exact Sciences*, 42(1):71–89, 1991.

[54] Ruslan L. Davidchack, Richard Handel, and M. V. Tretyakov. Langevin thermostat for rigid body dynamics. *The Journal of Chemical Physics*, 130(23):234101, 2009.

[55] 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(8):3684–3690, 1984.

[56] Roland Faller and Juan J. de Pablo. Constant pressure hybrid molecular dynamics–monte carlo simulations. *The Journal of Chemical Physics*, 116(1):55–59, 2002.

[57] 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.

[58] Vojtech Spiwok, Zoran Sucur, and Petr Hosek. Enhanced sampling techniques in biomolecular simulations. *Biotechnology Advances*, 33(6, Part 2):1130 – 1140, 2015. BioTech 2014 and 6th Czech-Swiss Biotechnology Symposium.

[59] Donald Hamelberg, John Mongan, and J. Andrew McCammon. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics*, 120(24):11919–11929, 2004.

[60] 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. PMID: 26300708.

[61] 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. PMID: 25061441.

[62] Apurba Bhattarai and Yinglong Miao. Gaussian accelerated molecular dynamics for elucidation of drug pathways. *Expert Opinion on Drug Discovery*, 13(11):1055–1065, 2018. PMID: 30371112.

[63] 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. PMID: 25061441.

[64] P. W. Atkins and Julio de. Paula. *Atkins' physical chemistry*. Oxford University Press, Oxford, 2014.

[65] Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, and Hou T. End-point binding free energy calculation with mm/pbsa and mm/gbsa: Strategies and applications in drug design. 119(16):9478–9508, 2019.

[66] Michael R. Shirts. *Best Practices in Free Energy Calculations for Drug Design*, pages 425–467. Springer New York, New York, NY, 2012.

[67] Xuan-Yu Meng, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2):146–157, 06 2011.

[68] Michael R. Shirts, David L. Mobley, and John D. Chodera. Chapter 4 alchemical free energy calculations: Ready for prime time? volume 3 of *Annual Reports in Computational Chemistry*, pages 41 – 59. Elsevier, 2007.

[69] Nadine Homeyer, Friederike Stoll, Alexander Hillisch, and Holger Gohlke. Binding free energy calculations for lead optimization: Assessment of their accuracy in an industrial drug design context. *Journal of Chemical Theory and Computation*, 10(8):3331–3344, 2014. PMID: 26588302.

[70] Williams-Noonan BJ, Yuriev E, and Chalmers DK. Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry. 61(3):638–649, 2018.

[71] Samuel Genheden and Ulf Ryde. The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. 10(5):449–461, 2015.

[72] Miller BR 3rd, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, and Roitberg AE. Mmpbsa.py: An efficient program for end-state free energy calculations. 8(9):3314–21, 2012.

[73] Samuel Genheden. Mm/gbsa and lie estimates of host–guest affinities: dependence on charges and solvation model. *Journal of Computer-Aided Molecular Design*, 25(11):1085–1093, 2011.

[74] Gerrit Groenhof. *Introduction to QM/MM Simulations*, pages 43–66. Humana Press, Totowa, NJ, 2013.

[75] Hans Martin Senn and Walter Thiel. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

[76] Attila Szabo and Neil S. Ostlund. *Modern quantum chemistry : introduction to advanced electronic structure theory*. Dover Publications, Mineola, N.Y., 1996.

[77] Wolfram. Koch and Max C. Holthausen. *A chemist's guide to density functional theory*. Wiley-VCH, Weinheim ;, 2001.

[78] Perdew JP, Ruzsinszky A, Tao J, Staroverov VN, Scuseria GE, and Csonka GI. Prescription for the design and selection of density functional approximations: more constraint satisfaction with fewer fits. 123(6):62201, 2005.

[79] Burke K. Perspective on density functional theory. 136(15):150901, 2012.

[80] A. V. Arbuznikov. Hybrid exchange correlation functionals and potentials: Concept elaboration. *Journal of Structural Chemistry*, 48(1):S1–S31, 2007.

[81] Teepanis Chachiyo and Hathaithip Chachiyo. Understanding electron correlation energy through density functional theory. *Computational and Theoretical Chemistry*, 1172:112669, 2020.

[82] A. D. Becke and M. R. Roussel. Exchange holes in inhomogeneous systems: A coordinate-space model. 39(8):3761–3767, 1989.

[83] Axel D. Becke. A new mixing of hartree–fock and local density-functional theories. 98(2):1372–1377, 1993.

[84] John P. Perdew and Karla Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conference Proceedings*, 577(1):1–20, 2001.

[85] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988.

[86] Chengteh Lee and Carlos Sosa. Local density component of the lee–yang–parr correlation energy functional. *The Journal of Chemical Physics*, 100(12):9018–9024, 1994.

[87] Kieron Burke, John P. Perdew, and Yue Wang. *Derivation of a Generalized Gradient Approximation: The PW91 Density Functional*, pages 81–111. Springer US, Boston, MA, 1998.

[88] Axel D. Becke. Density-functional thermochemistry. i. the effect of the exchange-only gradient correction. *The Journal of Chemical Physics*, 96(3):2155–2160, 1992.

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

[90] 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.

[91] Jeng-Da Chai and Martin Head-Gordon. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics*, 128(8):084106, 2008.

[92] You-Sheng Lin, Guan-De Li, Shan-Ping Mao, and Jeng-Da Chai. Long-range corrected hybrid density functionals with improved dispersion corrections. *Journal of Chemical Theory and Computation*, 9(1):263–272, 2013. PMID: 26589028.

[93] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620, 2008.

[94] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. 52(12):997–1000, 1984.

[95] M. Thiele, E. K. U. Gross, and S. Kümmel. Adiabatic approximation in nonperturbative time-dependent density-functional theory. *Phys. Rev. Lett.*, 100:153004, Apr 2008.

[96] MARK E. CASIDA. *Time-Dependent Density Functional Response Theory for Molecules*, pages 155–192.

[97] Denis Jacquemin, Eric A. Perpète, Gustavo E. Scuseria, Ilaria Ciofini, and Carlo Adamo. Td-dft performance for the visible absorption spectra of organic dyes: Conventional versus long-range hybrids. *Journal of Chemical Theory and Computation*, 4(1):123–135, 01 2008.

[98] Daniel R. Roe and Thomas E. 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. PMID: 26583988.

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

[100] 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, 01 2000.

[101] Seonghoon Kim, Jumin Lee, Sunhwan Jo, and Wonpil Im. Charmm-gui ligand reader & modeler. 112(3 SUPP 1):289a, 2017.

[102] Marcus D. Hanwell, Donald E. Curtis, David C. Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, 2012.

[103] 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.

[104] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. 25(2):247, 2007.

[105] David A. Case, Thomas E. Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz Jr., Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.

[106] Jeffery B. Klauda, Richard M. Venable, J. Alfredo Freites, Joseph W. O'Connor, Douglas J. Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D. MacKerell, and Richard W. Pastor. Update of the charmm all-atom additive force field for lipids: Validation on six lipid types. *The Journal of Physical Chemistry B*, 114(23):7830–7843, 06 2010.

[107] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[108] 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(3):327 – 341, 1977.

[109] Yihan Shao, Zhengting Gan, Evgeny Epifanovsky, Andrew T.B. Gilbert, Michael Wormit, Joerg Kussmann, Adrian W. Lange, Andrew Behn, Jia Deng, Xintian Feng, Debashree Ghosh, Matthew Goldey, Paul R. Horn, Leif D. Jacobson, Ilya Kaliman, Rustam Z. Khaliullin, Tomasz Kuś, Arie Landau, Jie Liu, Emil I. Proynov, Young Min Rhee, Ryan M. Richard, Mary A. Rohrdanz, Ryan P. Steele, Eric J. Sundstrom, H. Lee Woodcock III, Paul M. Zimmerman, Dmitry Zuev, Ben Albrecht, Ethan Alguire, Brian Austin, Gregory J. O. Beran, Yves A. Bernard, Eric Berquist, Kai Brandhorst, Ksenia B. Bravaya, Shawn T. Brown, David Casanova, Chun-Min Chang, Yunqing Chen, Siu Hung Chien, Kristina D. Closser, Deborah L. Crittenden, Michael Diedenhofen, Robert A. DiStasio Jr., Hainam Do, Anthony D. Dutoi, Richard G. Edgar, Shervin Fatehi, Laszlo Fusti-Molnar, An Ghysels, Anna Golubeva-Zadorozhnaya, Joseph Gomes, Magnus W.D. Hanson-Heine, Philipp H.P. Harbach, Andreas W. Hauser, Edward G. Hohenstein, Zachary C. Holden, Thomas-C. Jagau, Hyunjun Ji, Benjamin Kaduk, Kirill Khistyaev, Jaehoon Kim, Jihan Kim, Rollin A. King, Phil Klunzinger, Dmytro Kosenkov, Tim Kowalczyk, Caroline M. Krauter, Ka Un Lao, Adèle D. Laurent, Keith V. Lawler, Sergey V. Levchenko, Ching Yeh Lin, Fenglai Liu, Ester Livshits, Rohini C. Lochan, Arne Luenser, Prashant Manohar, Samuel F. Manzer, Shan-Ping Mao, Narbe Mardirossian, Aleksandr V. Marenich, Simon A. Maurer, Nicholas J. Mayhall, Eric Neuscamman, C. Melania Oana, Roberto Olivares-Amaya, Darragh P. O'Neill, John A. Parkhill, Trilisa M. Perrine, Roberto Peverati, Alexander Prociuk, Dirk R. Rehn, Edina Rosta, Nicholas J. Russ, Shaama M. Sharada, Sandeep Sharma, David W. Small, Alexander Sodt, Tamar Stein, David Stück, Yu-Chuan Su, Alex J.W. Thom, Takashi Tsuchimochi, Vitalii Vanovschi, Leslie Vogt, Oleg Vydrov, Tao Wang, Mark A. Watson, Jan Wenzel, Alec White, Christopher F. Williams, Jun Yang, Sina Yeganeh, Shane R. Yost, Zhi-Qiang You, Igor Ying Zhang, Xing Zhang, Yan Zhao, Bernard R. Brooks, Garnet K.L. Chan, Daniel M. Chipman, Christopher J. Cramer, William A. Goddard III, Mark S. Gordon, Warren J. Hehre, Andreas Klamt, Henry F. Schaefer III, Michael W. Schmidt, C. David Sherrill, Donald G. Truhlar, Arieh Warshel, Xin Xu, Alán Aspuru-Guzik, Roi Baer, Alexis T. Bell, Nicholas A. Besley, Jeng-Da Chai, Andreas Dreuw, Barry D. Dunietz, Thomas R. Furlani, Steven R. Gwaltney, Chao-Ping Hsu, Yousung Jung, Jing Kong, Daniel S. Lambrecht, WanZhen Liang, Christian Ochsenfeld, Vitaly A. Rassolov, Lyudmila V. Slipchenko, Joseph E. Subotnik, Troy Van Voorhis, John M. Herbert, Anna I. Krylov, Peter M.W. Gill, and Martin Head-Gordon. Advances in molecular quantum chemistry contained in the q-chem 4 program package. *Molecular Physics*, 113(2):184–215, 2015.

[110] Yinglong Miao, Apurba Bhattarai, and Jinan Wang. Ligand gaussian accelerated molecular dynamics (ligamd): Characterization of ligand binding thermodynamics and kinetics. *bioRxiv*, 2020.

[111] Denis Jacquemin, Valérie Wathelet, Eric A. Perpète, and Carlo Adamo. Extensive td-dft benchmark: Singlet-excited states of organic molecules. *Journal of Chemical Theory and Computation*, 5(9):2420–2435, 09 2009.

[112] Michael J. G. Peach, Peter Benfield, Trygve Helgaker, and David J. Tozer. Excitation energies in density functional theory: An evaluation and a diagnostic test. *The Journal of Chemical Physics*, 128(4):044118, 2008.

[113] H.-J. Timpe. Photochromism – molecules and systems. herausgeber: Dürr, h., bouas-laurent, h. 1. auflage, 1068 s. amsterdam, oxford, new york, tokyo: Elsevier, 1990. schriftenreihe: Studies in organic chemistry, 40. isbn 0-444-87432-1. *Journal für Praktische Chemie*, 333(5):811–812, 1991.