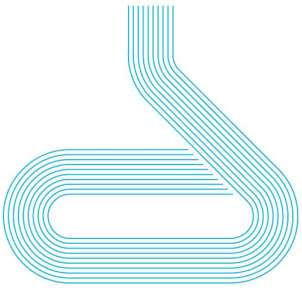


Universidade Vigo

Computational Study of Interactions of Na^+ and K^+ in Eukaryotic Ionic Channels

Trabajo Fin de Máster: Curso 2023-2024

Raúl Lago Saavedra

SUPERVISORS:
Nicolás Ramos Berdullas
Juan José Nogueira Pérez

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Theoretical Background

During the 1950's two seemingly uncorrelated scientific events occurred:

- On the one hand, the development and increasing accuracy of both biomolecular structure determination techniques (NMR, X-ray crystallography) and molecular structure prediction techniques from theoretical calculations, led into the elucidation of the molecular structure of the deoxyribonucleic acid (DNA) molecule, by Watson, Crick and Franklin [1]. This milestone marked the beginning of molecular biology, biologists and chemists were hand-to-hand studying biomolecules at an atomistic level. A paradigm shift was slowly occurring within biology, chemistry and medicine.
- On the other hand, the advent of Lambda calculus and the Turing machine, both of which were developed by Alan Turing and Alonzo Church, with inspiration drawn from Kurt Gödel [2], marked the start of modern computation, quicly followed by a period of sustained growth on the field. Given their capacity to perform a significant number of calculations, computers started to be developed, primarily, within the military and engineering fields. In the field of exact sciences, there were a plethora of potential applications for computers. However, at that time, machines were too large and too expensive to perform any calculations at the complexity level that physical, chemical or biological research requires.

In the following decades the Hohenberg-Kohn (HK) theorems and Kohn-Sham (KS) equations were published, defining the modern Density Functional Theory (DFT) [3], a simple method able to obtain the energy of any molecule with quantum mechanical accuracy. Even though DFT is greatly efficient, computational resources and technology available at the time were not powerful enough to calculate anything approaching the size of proteins, DNA/RNA, polysaccharides – *i.e.* biologically relevant macromolecules. Even so, DFT did shed some light towards the possibility of combining the previously *seemingly* uncorrelated events into a new field of research: **computational biochemistry** [4].

In the meantime, molecular biologists did not *stand-by doing nothing*. New biomolecules were being discovered and resolved every week. This provided an immense database for theoretical chemists, from which extract information about the conformation and atomic behaviour of these molecules. Not only giving them a reference frame to compare their results with, but also the capability to build force fields (FF). In contrast with other previous approaches, force fields (FF) were based on Molecular Mechanics (MM) rather than Quantum Mechanics (QM). While not offering the same accuracy, their execution times are much less

demanding. The accessibility and applicability of FF, was fundamental to convince experimentalists that theoretical and computational calculations are not only useful, but necessary. The increasing accuracy and reliance on computational chemistry helped the development of novel methodologies and techniques in the field. This, in conjunction with the significant advancements made in computer science, has led to the current situation, where the majority of publications in chemistry encompass both experimental and theoretical contributions.

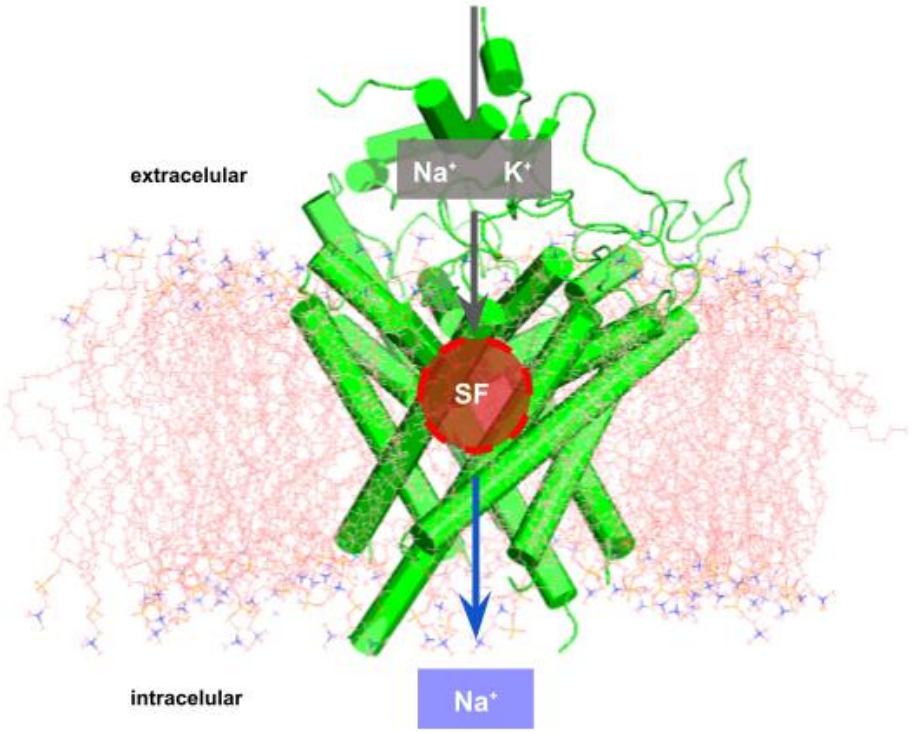


Figure 1.1: Schematic process of main function and selectivity of voltage-gated ion channels, in particular of Na_v. The SF is found towards the end of the channel, where K^+ are selected out and Na^+ are favoured to pass through. The behaviour of the channel does also depend on the voltage of the environment.

For the interest of this work it is worth paying special attention to membrane protein channels. These are essential for the correct functioning of every living being's homeostasis. These proteins' actions are key-elements on functional metabolism, as they take part in almost every (signal) transduction process [5]. Often, proteins of great size are comprised of subunits. The fundamental and most important on protein channels is the α subunit, responsible for the pore formation, while all other subunits (β,γ,δ,*etc.*), also called accessory subunits, mainly regulate the former's behaviour [6]. Even though, protein channels' molecular structures are more and more easily resolved, the action mechanism of many of them is not completely understood yet. In particular, ion channels represent a significant amount ($\sim$5.1%) of the total of unique genes of drug targeted membrane proteins [7]. These are associated with many significant diseases: Alzheimer's, epilepsy, Parkinson's, endocrine disorders, episodic ataxia, cardiac arrhythmia, (familial hemiplegic) migraine, among others [8, 9]. VVoltage-gated ion channels [Figure 1.1] are a subgroup of this category are

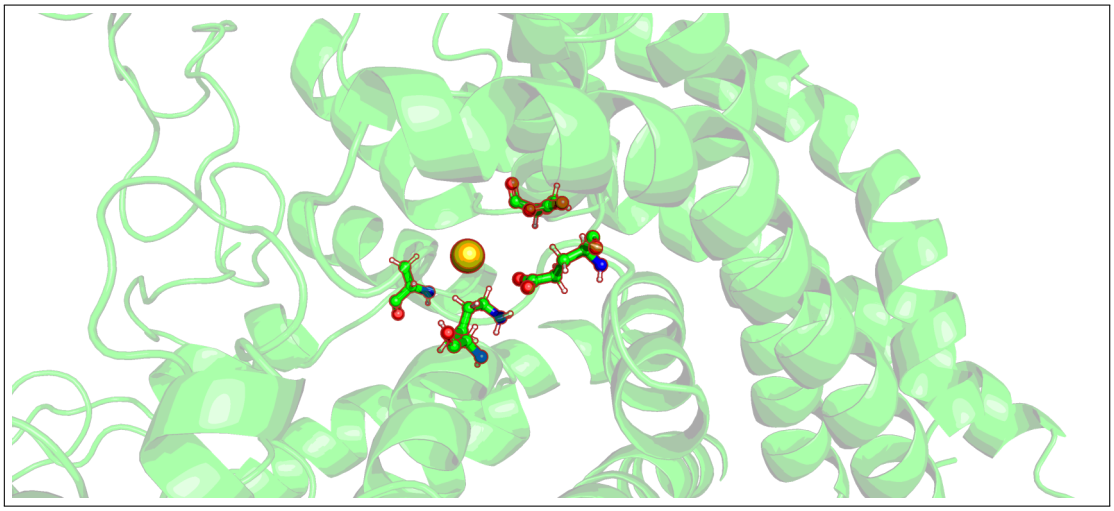


Figure 1.2: Zenithal projection of Selectivity Filter (Asp^{406}/Glu^{761}/Lys^{1244}/Ala^{1536}), with a Na^+ in its center, highlighted using `spheres` and `sticks` within the α subunit in `cartoon`.

a subgroup of this category. As the name implies, the activity of voltage-gated channels is regulated by changes in the membrane's voltage. These proteins are capable of detecting the transfer of an electrical charge through the cell membrane and regulate its activity accordingly. From the voltage-gated ion channels, there are three main ion-selective families: Voltage-Gated Sodium Channels (Na_v), Voltage-Gated Potassium Channels and Voltage-Gated Calcium Channels.

There are 9 known human Na_v, they all occur in the central or peripheral nervous system, except $Na_v1.4$ and $Na_v1.5$ that occur in skeletal and heart muscle, respectively [7]. These are called α subunits and are able to operate by themselves. However, there are 5 β subunit variants. Association between them is called a *complex*. For instance, Na_v complexes are usually 1α—1β or 1α—2β. On the one hand, the α subunit can transport ions by itself and has a selectivity towards sodium (Na^+). On the other hand, the β subunit (mainly) works as a channel modulator [10]. Additionally, the presence of β subunits regulates the α subunit gene expression (within the same tissue), and *vice-versa*. This work is focused on the recently obtained structure of human $Na_v1.4$-$\beta1$ complex, from a cryogenic electron microscopy (cryo-EM) analysis with a resolution of 3.2 Å. Additionally, the Asp^{406}/Glu^{761}/Lys^{1244}/Ala^{1536} (DEKA) region, corresponding to the selectivity filter (SF) confers sodium channels its selectivity [Figure 1.2], was resolved with a local resolution of 2.8 Å. Yet, residues 287 to 335 were not well resolved, the least conserved segment with other Na_v. The whole protein system consists of 2054 residues. The α subunit, with 1836 amino acids, is divided into 4 homologous but non-identical repeats (I-IV), each containing 6 transmembrane segments (S1-S6) [11]. The $Na_v1.4$ is highly conserved among all eukaryotic organisms, as shown by Uniprot Blast [12] results, notably within the *Hominidae* family, as the *Homo Sapiens* (Modern Human) protein showed an astonishing correlation with other hominids, obtaining a residue sequence identity higher than 99% and E-value of 0, which jointly show a null number of random alignments or consistent residue mismatches, when compared with the $Na_v1.4$ on *Gorilla Gorilla Gorilla* (Western Lowland Gorilla), *Pan Troglodytes* (Chimpanzee), *Pan Paniscus* (Bonobo) and *Pongo Pygmaeus Abelii* (Sumatran Orangutan). The SF, corresponding to the DEKA motif, is conserved homogeneously among all eukaryotic organism [13, 14].

Molecular level in-depth studies of how these proteins do operate are a necessary step to develop novel successful therapies. However, in order to study many of the protein properties, molecular dynamics (MD) simulations are needed. In contrast with static calculations, MD simulations are able to evolve a sys-

tem in time, employing several single-point calculations in combination with Newton's laws of motion. Fortunately, nowadays the computational resources in research institutions are sufficient to conduct such simulations, even at QM accuracy, commonly with DFT level of theory. Theoretical chemistry is by far the best way to study these systems' mechanism at an atomistic level, thus the immense relevance of computational calculations on the field. Particularly, few research has been done related to Na_v on eukaryotes. In contrast, some MD simulations have been successfully performed on Na_v on prokaryotes. The implication on human health and the amount of high resolution Na_v makes them very suitable for theoretical calculations [14]. This work aims to further explore and explain previously obtained MM-MD results, from calculations conducted by Nuria Anguita and Juanjo Nogueira at the MoBioChem research group of the Autonomous University of Madrid.

These previously performed simulations did not resulted as expected for de DEKA selectivity, as both Na^+ and K^+ passed through the selectivity filter with similar ease. The simulations were performed with a FF and applying a potential electric field to enhance the simulations. Both ion enter the channel in coordination with water molecules. In the near proximity of the SF, Na^+ lost an average of 1.5 waters, but gained one water molecule while exactly passing through the DEKA ring. On the other hand, K^+ loses 1.6 water molecules in the vicinity of the SF, and loses one additional water molecule when passing through the DEKA ring. Nevertheless, waters were not accurately simulated by the FF, as some frames of the simulation show equal charges facing each other (*e.g.* $-H^{\delta+}$ and Na^+).

1.2 Objectives

The general objective of this work is to characterize the interactions involved in the ion conduction mechanism of Na_v. To achieve this, the following specific objectives are pursued:

- To perform a benchmark of DFT functionals to chose an appropriate functional to model the intermolecular interactions present on the system.
- To study the microsolvation of Na^+ and K^+ and investigate their behaviour in terms of movement and interactions.
- To build a model composed of either Na^+ or K^+, water molecules and either glutamate or aspartate, representing the most relevant interactions between the solvated ion and the DEKA.
- To compute the interaction energy among all the components of the cluster model and characterize this interaction by an energy decomposition analysis technique

Chapter 2

Methodology

In this work the intermolecular interactions between solvated ions and amino acids have been characterized to shed light into the ion conduction mechanism in the eukaryotic voltage gated ion channels. Such a characterization has been carried out in a cluster model by applying a combination of different computational methods, namely, DFT, MD and EDA.

2.1 Electronic Structure

The initial quantum mechanical (QM) methods were based on wave function theory (WFT), which was inspired by Schrödinger's equation (2.1). The Hartree-Fock (HF) method, which was introduced in 1930 and is still used in the present day, marked a significant turning point in theoretical and computational chemistry. In order to overcome the many-body problem, the HF method was constructed upon the mean-field approximation, where only one electron is assessed at a time, while all other electrons form an *electronic field* that affects the one electron's behaviour. Consequently, the HF method is not exact, as the mean-field approximation does not account for any electron correlation at all. In order to address this significant disparity between the exact energy and the HF energy methodology in accounting for correlation effects, post-HF methodologies have been developed. Despite producing results that are closer to the exact energy, these methodologies are quite computationally expensive. One of the most commonly employed post-HF methodology is Møller-Plesset perturbation theory (MPPT). However, due to the high computational cost, it is often applied only up to the second order (MP2).

$$\hat{\mathrm{H}} \left|\psi\right\rangle = E \left|\psi\right\rangle \tag{2.1}$$

The most prevalent approximations in quantum mechanical (QM) methods employed in computational chemistry is the Born-Oppenheimer (BO) approximation. This states that the movement of electrons is decoupled from that of nuclei. The use of the BO approximation has minimal impact on calculations in most chemical systems, as nuclei have a *quasi*-infinite mass in comparison with electrons.

A wide variety of theoretical approaches exist for calculating a molecule's energy, all of which explicitly consider electrons. However, the treatment of electrons implies the consideration of two particularly notable electron properties: **exchange** and **correlation** (XC). The **exchange interaction**, also called Pauli repulsion when considering only fermions, arises when a system of indistinguishable particles, *e.g.* – electrons, are susceptive to exchange leading to an additional term in the energy that accounts for the anti-symmetry of the wave-function. The **electronic correlation** arises from the instantaneous interaction that occurs between

electrons, which can be further separated into dynamic correlation, that accounts for the electron-electron repulsion, and static correlation that arises from the near-degeneracy of a system's electronic configuration. As stated earlier, HF does not take into account correlation, though it completely and exactly calculates the exchange, thus (WFT) post-HF methods exclusively treat the error caused by the lack of correlation.

2.2 Density Functional Theory

2.2.1 Foundations

In contrast with WFT, on the early years of computational and theoretical chemistry, DFT based methodologies had been poorly researched. The only major breakthrough done with DFT before the HK theorems was the Thomas-Fermi-Dirac (TFD) model [3]. Nevertheless, DFT offered a refreshing point of view for theoretical chemistry, by using the electron density (ρ) the Hamiltonian would depend only on the positions of the nuclei and the total number of electrons (N).

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \tag{2.2}$$

The approach followed by the TFD model was to define the potential energy in a classical way and use a uniform electron gas (UEG), also called *jelium*, to calculate the kinetic energy. Consequently, early DFT models found widespread use in the solid-state physics community. In contrast to chemistry, where the fairly large errors in molecular calculations resulted in TFD having little to no impact. The model was, despite problematic, the best way of calculating a Hamiltonian using, virtually exclusively, the electronic density. However, this state of affairs was set to change when Hohenberg and Kohn (HK), in 1964, proved two theorems critical to establishing DFT as it is currently best known [15].

Hohenberg-Kohn Theorems

The total ρ completely and exactly determines all the ground-state properties of an N-electron system:

$$\nu_{ext} \rightarrow \psi_0 \rightarrow \rho \tag{2.3}$$

In essence, for a N-electron system, in an external potential (ν_{ext}), such as a random distribution of point charges, there is a unique ground-state wavefunction (ψ_0) and, consequently, a unique ρ. Ultimately, this means that, in a non-degenerate ground-state system, there shall be only one possible ρ for its ν_{ext} and thus a unique ψ_0 (2.4).

$$\rho \rightarrow \nu_{ext} \rightarrow \psi_0 \tag{2.4}$$

This constitutes de first HK theorem, and it is the most important premise to take in mind when working with DFT. Theoretically, from an ρ, it is possible to extract all the information that could be obtained from the wavefunction (ψ). This statement is proved by *reductio ad absurdum*. Any two possible external potentials (ν_a and ν_b) cannot be described by the same ρ, since this would imply the existence of two Hamiltonians ($\hat{\mathrm{H}}_a$ and $\hat{\mathrm{H}}_b$), that belong to two different ground-state wavefunction ($\psi_{0,a}$ and $\psi_{0,b}$), whose combination yield their respective ground-state energies ($E_{0,a}$ and $E_{0,b}$) (2.5), which are associated to the same ρ. Thus, the ψ_0 are not equal.

$$\begin{aligned} \langle \psi_{0,a} | \hat{\mathrm{H}}_a | \psi_{0,a} \rangle &= E_{0,a} \\ \langle \psi_{0,b} | \hat{\mathrm{H}}_b | \psi_{0,b} \rangle &= E_{0,b} \end{aligned} \tag{2.5}$$

Applying $\hat{\mathrm{H}}_a$ to $\psi_{0,a}$ cannot yield $E_{0,a}$. Actually, the obtained value must be higher than $E_{0,a}$ (2.6), due to the variational theorem of WFT [15].

$$E_{0,a} < \langle \psi_{0,b} | \, \hat{\mathrm{H}}_a \, | \psi_{0,b} \rangle \tag{2.6}$$

rewriting $\hat{\mathrm{H}}_a$ as $\hat{\mathrm{H}}_a - \hat{\mathrm{H}}_b + \hat{\mathrm{H}}_b$, leads to an inequality relating the ground-state energy of one wavefunction with the other (2.7).

$$\begin{aligned} E_{0,a} &< \langle \psi_{0,b} | \, \hat{\mathrm{H}}_a - \hat{\mathrm{H}}_b + \hat{\mathrm{H}}_b \, | \psi_{0,b} \rangle \\ &< \langle \psi_{0,b} | \, \hat{\mathrm{H}}_a - \hat{\mathrm{H}}_b \, | \psi_{0,b} \rangle + E_{0,b} \\ &< \int [\nu_a - \nu_b] \rho(\mathbf{r}) d\mathbf{r} + E_{0,b} \end{aligned} \tag{2.7}$$

If this same procedure is applied with $\hat{\mathrm{H}}_a$ and $\psi_{0,a}$, the following would be obtained:

$$E_{0,b} < \int [\nu_b - \nu_a] \rho(\mathbf{r}) d\mathbf{r} + E_{0,a} \tag{2.8}$$

Combining both inequalities gives rise to a contradiction in which the sum of the energies is greater that itself (2.9), as this cannot be possible, the premise that there are two possible external potentials that are described by the same density functional in a non-degenerate ground-state system is false, thus there must be one and only one ρ for each ψ_0, and *vice-versa*.

$$\begin{aligned} E_{0,a} + E_{0,b} &< \int [\nu_a - \nu_b] \rho(\mathbf{r}) d\mathbf{r} + \int [\nu_b - \nu_a] \rho(\mathbf{r}) d\mathbf{r} + E_{0,a} + E_{0,b} \\ E_{0,a} + E_{0,b} &< \cancel{\int [\nu_a - \nu_b] \rho(\mathbf{r}) d\mathbf{r}} - \cancel{\int [\nu_a - \nu_b] \rho(\mathbf{r}) d\mathbf{r}} + E_{0,a} + E_{0,b} \\ E_{0,a} + E_{0,b} &< E_{0,a} + E_{0,b} \end{aligned} \tag{2.9}$$

Hence, it is possible to build a *functional* (function of functions) that is able to obtain the energy of a chemical system, simple by knowing the number of electrons and the position of the nuclei. The second HK theorem proves that DFT can be variational (2.10), thereby indicating that the energy yielded by such functional would be exact at the ground-state. Consequently, from such functional ($F[\rho]$), the obtained energy will always be lower than the energy of $F[\rho']$, such that ρ' does not correctly correspond to the true ν_{ext} [3].

$$F[\rho'] + \int \nu_{ext} \rho'(\mathbf{r}) d\mathbf{r} \geq F[\rho] + \int \nu_{ext} \rho(\mathbf{r}) d\mathbf{r} = E_0 \tag{2.10}$$

Kohn-Sham Equations

The HK theorems proved that a functional that only uses ρ, as input, is able to yield the exact energy of a system. Not much later the Kohn and Sham (KS) equations were released. A set of equations that could be used to approach this problem. However, it is important to take in mind that the obtained functionals do not represent the *exact functional* proposed by HK.

The KS equations consider, at first, a system of non-interacting particles, from which an approximate (K_0) of the kinetic energy (K) of electrons is extracted from its orbitals (ϕ) (2.11). In order to approximate the

electron-electron potential energy (V_{ee}), the Coulomb energy (J) is used (2.12).

$$K_0 = \sum_i^N \langle\phi| - \frac{1}{2}\nabla^2 |\phi\rangle \tag{2.11}$$

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(1)\rho(2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \tag{2.12}$$

As these are approximations, combining terms (K_0 and J) into a single $F[\rho]$ does not yield the exact energy. The error obtained is known as the XC energy (E_{XC}).

$$F[\rho] = K_0 + J[\rho] + E_{\text{XC}}[\rho] \tag{2.13}$$

Thus, the E_{XC} could be calculated as the difference (2.14) between the exact energy value and the approximate value.

$$E_{\text{XC}} = (K[\rho] - K_0) + (V_{ee}[\rho] - J[\rho]) \tag{2.14}$$

Finally, the total KS energy would be:

$$E[\rho] = K_0 + \int \nu_{ext}\,\rho(\mathbf{r}) d\mathbf{r} + J[\rho] + E_{\text{XC}} \tag{2.15}$$

from which the only unknown value is E_{XC}, that represents a small fraction of the total energy. Further, orbitals can be minimized (2.16) by using a KS potential (ν_{KS}), consisting of ν_{ext}, the electronic potential (ν_{el}) and the functional derivative ($\frac{\delta E_{\text{XC}}}{\delta\rho}$) (2.17). Consequently, KS DFT can be carried out using the same basis sets already designed for WFT [3].

$$-\frac{1}{2}\nabla^2\phi_i + \nu_{\text{KS}}\phi_i = \epsilon_i\phi_i \tag{2.16}$$

$$\nu_{\text{KS}} = \nu_{ext} + \nu_{el} + \frac{\delta E_{\text{XC}}}{\delta\rho} \tag{2.17}$$

The first approach to the E_{XC} was the Local Density Approximation (LDA) based on the UEG XC energy adjusted to the local ρ (2.18).

$$E_{\text{XC}}^{\text{LDA}} = \int e_{\text{XC}}^{\text{UEG}}(\rho) \tag{2.18}$$

2.2.2 Jacob's Ladder

Even though, the exact E_{XC} was proved to exist, there is no guide on how to find it. Using LDA as the theoretical framework and starting point, Density Functional Approximation(s) (DFA) have been and are developed, by multiple theoretical chemistry research groups and institutions [16]. While DFT is a theoretical basis, variational and thus exact in their foundations, DFAs are merely the attempts to develop a functional based on DFT [3].

In order to improve the exchange energy (E_{X}) in DFAs, an added dependency on the electron density gradient ($\nabla\rho$) is made through what is known as the exchange enhancement factor (f_{X}) that itself utilizes the reduced density gradient (s) (**??**). All together produce the Generalized Gradient Approximation (GGA) (**??**). Correlation also improves by introducing the $\nabla\rho$, though it has a complex analytical form, not directly

correlated to physical reasoning. It is important to remember that, in principle, any DFA's correlation may be combined with any other DFA's exchange.

(2.19)

The next logical step is not only to account for the density gradient ($\nabla\rho$), but also the density Laplacian ($\nabla^2\rho$). Alternatively, the orbitals kinetic energy density (τ) (2.20) may be introduced instead of the $\nabla^2\rho$, as they carry, essentially, the same information.

$$\tau_\sigma = \frac{1}{2}\sum_i |\nabla\phi_{i\sigma}|^2 \tag{2.20}$$

The aforemetioned terms can be employed in in bo, Exchange and Correlation terms. Therefore, the GGAs that incorporate either τ, $\nabla^2\rho$ or both are called meta Generalized Gradient Approximationss (mGGAss) [3, 17].

$$E_{\text{XC}}^{\text{mGGA}} = \int e_{\text{XC}}(\rho, \nabla\rho, \tau) \tag{2.21}$$

On the other hand, the HF method yields the exact exchange. Though, as the KS framework is not identical to the one in the HF method, using the HF exchange (HF_X), as it is exact, could be a way to improve DFAs exchange. Indeed, it has been demonstrated that using a fraction of the HF exchange in GGA and mGGA functionals does improve the results. The DFAs that followed this approach are called Hybrid Functionals (**??**). The sum between the percentage of HF_X and E_X combined is usually close or equal to 100% ($a + b \approx 1$).

$$E_{\text{XC}} = a\text{HF}_\text{X} + bE_\text{X} + E_\text{C} \tag{2.22}$$

The last step implemented in DFAs to better results has been to introduce the MPPT, almost uniquely MP2, into the correlation energy (2.23), these are called Double-Hybrid Functionals. Though, as expected, these functionals are very computationally expensive [17].

$$E_{\text{XC}} = a_\text{X}\text{HF}_\text{X} + b_\text{X}E_\text{X} + a_\text{C}\text{MP2}_\text{C} + b_\text{C}E_\text{C} \tag{2.23}$$

This sort of *regatta* towards the exact functional has been coined as *Jacob's Ladder*, where LDAs stand at the bottom, GGAs represent the following step, mGGAs the next one, then Hybrid, then Double-Hybrid and at the top is the uknown exact XC-DFT functional. Each of these is currently referred as a *DFT Family* [16, 18].

Nonexistence of an Absolute Reference Frame

As DFAs are neither exact nor variational there is no absolute reference frame to guide from when using exclusively DFAs. Not even increasing the basis set size would necessarily *improve the results*. Thus, in order to favour one functional over another, either there are performedsupporting calculations using

variational methods, or a previous benchmark, from which high level variational (costly post-HF) method calculations were performed, is consulted. The latter has become a very popular approach, as many DFAs excel in particular types of systems and fall behind in most other environments, it is much easier to select a few DFAs from a grand catalogued of already tested out functionals, than to systematically perform big benchmarks for each new study.

2.3 Interaction Energy

The interaction energy ($E_{\text{Interaction}}$) of any system's fragment with its environment is defined (2.24) as the difference between the system's energy (E_{AB}) with the fragment's energy (E_{A}) and the environment's energy (E_{B}), which is called supermolecular approach. In this scenario, the *environment* means anything that is not the *fragment*. Depending on the system, it might be more suitable to call both A and B *monomers* and AB the *complex*.

$$E_{\text{Interaction}} = E_{\text{AB}} - (E_{\text{A}} + E_{\text{B}}) \tag{2.24}$$

2.3.1 Basis Sets and Superposition Error

The application of quantum mechanical (QM) methods is contingent upon the use of orbitals (ϕ), which serve as the fundamental representation of electrons. Therefore, the development of a mathematically rigorous framework for describing these orbitals is of great importance. A common way to achieve a readily computable set of orbitals, a number of Gaussian functions (2.25) are assembled to form a basis function (2.26), which is then combined with multiple basis functions to constitute a basis set (BS) [17].

$$\chi = AAe^{-\mu r^2} \tag{2.25}$$

$$\phi_i \approx \sum_{\alpha} c_{\alpha i}\chi_{\alpha} \tag{2.26}$$

As BSs do not perfectly describe electronic orbitals, they are subject to associated errors. In particular, the basis set superposition error (BSSE) arises when studying the interaction energy between fragments of a system. When atoms between molecules are in close proximity, their basis functions may overlap, leading to an overestimation of the binding energy. Since the basis set of the complex AB is larger than the basis set of the individual moleculas A and B, when doing the energy difference EAB-EA-EB the interaction energy is overestimated due to the artifical lowering of E_{AB} with respect to E_{A} and E_{B}. Since the basis set of the complex AB is larger than the basis set of the individual moleculas A and B, when doing the energy difference (2.24) the interaction energy is overestimated due to the artifical lowering of E_{AB} with respect to E_{A} and E_{B}. One way to avoid this is by using the complete (infinite) basis set, which is a highly expensive approach. Conversely, the counterpoise (CP) method represents a considerably more efficient solution. In order to account for the BSSE when calculating interaction energies, the CP method calculates each monomer individually, while maintaining all the BSs of the system (2.27). Thus, the BSSE can be calculated and subsequently subtracted from the total energy of the system.

$$E_{\text{Interaction}} = E_{\text{AB}}^{\text{AB}} - (E_{\text{A}}^{\text{AB}} + E_{\text{B}}^{\text{AB}}) \tag{2.27}$$

2.3.2 Energy Decomposition Analysis

In order to study the interaction energy ($E_{\text{Interaction}}$) itself, it is useful to break down the energy into different contributions. According to perturbation theory, all intermolecular interactions contain the following fundamental physical contributions: [19]

- **Electrostatic**: Obtained from the *quasi*-classical Coulomb interaction.
- **Exchange**: Results from the tunnelling of electrons between interacting systems.
- **Repulsion**: Alludes to the Pauli exclusion principle that occurs between electrons at immediate distances.
- **Induction**: Originates from the polarization response of each molecule to the electrostatic field of any others.
- **Dispersion**: Arises from the correlated fluctuations of electron densities between near molecules.

These last two (**induction** and **dispersion**) form part of the polarization.

The Density based Energy Decomposition Analysis (EDA) method is a perturbation approach to the intermolecular interactions, initiated from eq.2.27. It manipulates the total energy expression (2.28), which depends on the one-electron density ($\rho(\mathbf{r})$) and the two-electron exchange correlation density ($\rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2)$), to obtain a proper energy decomposition.

$$\begin{aligned} E = -\frac{1}{2}\int \nabla^2 \rho(\mathbf{r}, \mathbf{r}')_{\mathbf{r}'=\mathbf{r}} d\mathbf{r} + \int \nu_N \rho(\mathbf{r}) d\mathbf{r} + \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}_1|} d\mathbf{r}_1 d\mathbf{r}_2 \\ + \frac{1}{2}\int\int \frac{\rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}_1|} d\mathbf{r}_1 d\mathbf{r}_2 + \sum_{i=1}^{N-1}\sum_{j>i}^{N} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} \end{aligned} \tag{2.28}$$

Firstly, the nuclear terms are subdivided into the monomer contributions (2.29). Secondly, the one-electron and two-electron exchange correlation densities are described as a combination of unperturbed terms of each monomer, and perturbation terms resulting from their interaction (2.30) .

$$\begin{aligned} \sum_{i=1}^{N-1}\sum_{j>i}^{N} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} = \sum_{i=1}^{N_A-1}\sum_{j>i}^{N_A} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} + \sum_{i=N_A}^{N-1}\sum_{j>i}^{N} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} \\ + \sum_{i=1}^{N_A}\sum_{j=N_A+1}^{N} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}, \qquad N_B = N - N_A \end{aligned} \tag{2.29}$$

$$\begin{aligned} \rho(\mathbf{r}) &= \rho^A(\mathbf{r}) + \rho^B(\mathbf{r}) + \Delta\rho(\mathbf{r})_{\text{polarization}} + \Delta\rho(\mathbf{r})_{\text{Pauli}} \\ \rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2) &= \rho^A_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2) + \rho^B_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2) + \Delta\rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2) + \rho^{AB}_{\text{X}}(\mathbf{r}_1, \mathbf{r}_2) \end{aligned} \tag{2.30}$$

On the one hand, the $\Delta\rho(\mathbf{r})_{\text{polarization}}$ and $\rho_{\text{XC}}(\mathbf{r}_1, \mathbf{r}_2)$ terms arise in the relaxation of the system's wave-function, therefore account for the polarization interaction. On the other hand, the $\Delta\rho(\mathbf{r})_{\text{Pauli}}$ and $\rho^{AB}_{\text{X}}(\mathbf{r}_1, \mathbf{r}_2)$ terms allude to the anti-symmetrization of the wave-function, in other words the exchange-repulsion interactions. Finally, the $E_{\text{Interaction}}$ is divided into its fundamental contributors: the electrostatic energy

($E_{\text{electrostatic}}$); the *Pauli* energy (E_{Pauli}), that is a combination of exchange and repulsion energies, and finally the polarization energy ($E_{\text{polarization}}$) term:

$$E_{\text{Interaction}} = E_{\text{electrostatic}} + E_{\text{Pauli}} + E_{\text{polarization}} \tag{2.31}$$

Additionally, ($E_{\text{polarization}}$) can be decomposed into two distinct contributions, namely induction ($E_{\text{induction}}$) and dispersion plus other contributors ($E_{\text{disp+res-pol}}$), through the use of second-order perturbation theory, arriving to the final expression:

$$E_{\text{Interaction}} = E_{\text{electrostatic}} + E_{\text{Pauli}} + E_{\text{induction}} + E_{\text{disp+res-pol}} \tag{2.32}$$

This last term ($E_{\text{disp+res-pol}}$) is considered to represent the dispersion energy as these other polarization contributors, in the case of weak non-covalent interactions, are negligible. [20]

2.4 Molecular Dynamics

2.4.1 Basic Concepts and Verlet integration algorithm

Molecular Dynamics (MD) simulations belong to the time-dependent methods, where the kinetic energy is strictly related to the temperature, thus molecules can only explore regions of the potential energy surface (PES) that are lower (or equal) to the kinetic energy. Take in mind that in MD (usually) energy is conserved. The movement of the nuclei are calculated using Newton's second equation (2.33), relating the force ($\text{F} = -\frac{dV}{dx}$), obtained through the potential energy (V) and the position of the particles (x), with mass (m) and acceleration ($a = \frac{d^2x}{dt^2}$).

$$-\frac{dV}{dx} = m\frac{d^2x}{dt^2} \tag{2.33}$$

In order to calculate the position of the nuclei on the following step, integration algorithms based on Newton's equations are used. The most common is the *Verlet* algorithm (2.35 & 2.36). Considering the Taylor expansion for the movement of the nuclei (2.34) using positive and a negative timesteps (Δt & $-\Delta t$) at the same time ($x_{i+\Delta t} + x_{i-\Delta t}$) even order terms cancel each other, while odd order terms sum.

$$\begin{aligned} x_{i+\Delta t} &= \sum_k \frac{1}{k!}\frac{d^k x}{dt^k}(\Delta t)^k \\ x_{i-\Delta t} &= \sum_k \frac{1}{k!}\frac{d^k x}{dt^k}(\Delta t)^k(-1)^k \end{aligned} \tag{2.34}$$

Thus jerk ($\frac{d^3x}{dt^3}$) is not being considered, the remaining hyperacceleration terms (now of order 4 or higher) are so small that they can be considered negligible (2.35). Though, the resulting equation needs the step previous to the initial one. This can be estimated (2.36) using a first order approximation of the Taylor

expansion (2.34) [17].

$$x_{i+\Delta t} = (2x_i - x_{i-\Delta t}) + \frac{d^2x}{dt^2}(\Delta t)^2 + \cancel{\frac{1}{12}\frac{d^4x}{dt^4}(\Delta t)^4 + \ldots} \tag{2.35}$$

$$x_{0-\Delta t} = x_0 - \frac{dx}{dt}(\Delta t) \tag{2.36}$$

2.4.2 Kinetic Energy and Temperature

In the context of studying biological systems, it is essential to define a temperature (T), as biological environments highly influenced by T. To maintain a specific T throughout an MD simulation, the kinetic energy (K) is adjusted to align with the equipartition theorem (2.38), or similar expressions.

$$K = \frac{1}{2}\sum_{i}^{N} m_i v_i^2 \tag{2.37}$$

$$K = \frac{3}{2}k_B T \tag{2.38}$$

However, using the Verlet algorithm, velocities (v_i) are not directly calculated, as the evolution of the atomic coordinates only depend on its acceleration (a_i), but are essential to calculate the K (2.37). Thus, the velocities have to be estimated (2.39) or calculated at each step during the simulation (2.40).

$$v_i = \frac{r_{i+\Delta t} - r_{i-\Delta t}}{\Delta t} \tag{2.39}$$

$$v_{i+\Delta t} = v_i + \frac{1}{2}(a_i + a_{i+\Delta t})\Delta t \tag{2.40}$$

Once the temperature is specified, the velocities are adjusted in order to maintain K following the equipartition theorem (2.38). The way in which the velocities are calibrated may vary, the many algorithms that serve this purpose have been coined as *thermostats*. A common way in which K suits the simulation desired temperature (T_{ref}) is adding a coupling parameter (λ) that defines the velocity scale factor in accordance with the simulated temperature (T_{sim}) at each step (2.41)

$$\text{Velocity scale factor} = \sqrt{1 + \frac{\Delta t}{\lambda}\left(\frac{T_{ref}}{T_{sim}} - 1\right)} \tag{2.41}$$

More complex thermostats, such as the the Nosé-Hoover, considers a *heat bath* around the systems whose $lambda$ varies dynamically along the simulation, producing a more thermodinamically accurate fluctuation of temperature [17].

2.4.3 *Ab initio* Potential Energy Calculations

As it was previously described MD propagate the nuclear positions by solving Newton's second equation in a time-discretized form using a potential energy function. When the latter is *ab initio* or QM based,

recurring to classical mechanics is justified in Born-Oppenheimer Molecular Dynamics (BOMD). It uses an approximation of the Born-Huang representation of the wavefunction (2.42), where it is decoupled into a direct product of the electronic (Φ) and nuclear (Ω) wavefunctions, as the mass of the nuclei is *quasi*-infinite in comparison with the mass of electrons (the basis of the BO approximation).

$$\psi(\mathbf{r}, \mathbf{R}, t) \approx \Omega(\mathbf{R}, t)\Phi(\mathbf{R}, \mathbf{r}) \tag{2.42}$$

When the classical limit is taken, the total energy is conserved without uncertainty. Thus the nuclei kinetic energy (K_{nuc}) may be calculated through the V (2.43). In conclusion, Newton's laws of motion may be applied (2.44) while using QM based potentials (V_{QM}).

$$K_{nuc} + V[\mathbf{R}, \mathbf{r}] = 0 \tag{2.43}$$

$$V_{\text{QM}}[\mathbf{R}, \mathbf{r}] \rightarrow M\frac{d^2\mathbf{R}}{dt^2} \tag{2.44}$$

Chapter 3

Computational Details

This chapter resumes the procedures followed and overall performed calculations. A fluxogram summarizing this chapter can be found at the end of it on Figure 3.5. Two microsolvation structures with 18 waters each were studied for both Na^+ and K^+. After analysing the results one of these structure was selected. From this solvation structure MD simulations were performed to better understand the behaviour of Na^+ and K^+ on a microsolvated environment. Later, a study of the ions with key-elements from the DEKA ring was performed. In the first place energy profiles were done, subsequently a more profound analysis of the interaction was carried out by means of EDA.

3.1 Microsolvation

3.1.1 Benchmark and Energy

Two distinct microsolvation structures were initially considered, comprising 18 waters each. Structures were obtained from other metal microsolvations [21]. The structures are named **A** [Fig.3.1a] and **B** [Fig.3.1b]. Structure **A** exhibits C3 symmetry, whereas structure **B** exhibits Ci symmetry. The first and second solvation spheres are formed by 6 and 12 water molecules, respectively. Consequently, the geometries under investigation were derived from these microsolvation structures, in which the original metal was replaced by Na^+ and K^+.

A comprehensive analysis has been conducted to identify all equilibrium structures and the most stable conformations in all microsolvation geometries. This analysis has been undertaken with consideration of a recent study on organic molecules microsolvated [22] in order to compare the accuracy and performance of different functionals. For this purpose, different types of functionals have been considered. Firstly, B3LYP and ωB97XD are Hybrid-GGA functionals. Secondly, M06-2X and PW6B95D3 are Hybrid-mGGA functionals. Lastly, PDE0-DH is a Double-Hybrid-GGA functional. All calculations have been carried out with the empirical dispersion correction **D3** of Grimme [**?**]. In order to fairly compare them, an optimization process with each functional was done from the same initial geometry [Fig.3.1]. Geometry optimization and vibrational analysis were conducted using Gaussian16. Additionally all optimization calculations were performed using cc-pVDZ basis set [23], except for K^+, treated with the LANL2DZ [24] combination of double zeta basis set and pseudo potentials.

Figure 3.1: Initial Microsolvation geometries, **A** and **B**, with C3 and Ci symmetry respectively. Each system has a total of 18 water molecules and the ion in its center.

3.1.2 Molecular Dynamics

Even though the previous calculations helped to choose a functional to perform posterior calculations, the MD simulations had to be performed at the B97-D4 def2-SV(P) level of theory, in order to reduce computation times. The functional was chosen with basis on its performance and computational efficiency [22]. From the optimized microsolvation structure of each ion, BOMD simulations were performed. The simulations were carried out using Orca (Version: 5.0.4). The MDs run for 10000 steps with a timestep of 0.5 fs, the recovered trajectories had a stride of 5 steps. The temperature was set to the biological standard of 310 K, using a Nosé-Hoover Thermostat that controlled the temperature every 10.0 fs. The system was contained within a spherical cell of 6.2 Å of radius. Prior to the analysis, the trajectory's first steps were cut off, until the energy was stable. Two main analysis were performed to the MD obtained:

1. **Root-Mean-Square Fluctuation** (RMSF):
   Evaluates how much an atom has moved (from its initial position) in average, along the simulation.
2. **Radial Distribution Function** (RDF):
   Describes how the density of a specific atom type (*i.e.* oxygen) varies in function of distance from a center point (*i.e.* the ion) in average, along the simulation.

Calculations were done with MDAnalysis (Version: 2.7.0), for the RMSF, and AMBER (Version: amber20), for the RDF.

3.2 Interaction Energy

In order to understand the interaction between ion and SF from the channel, the two main amino acids (glutamate and aspartate), to which the ion coordinates within the DEKA ring, had been studied. To do so, the second solvation sphere was removed, remaining only the first sphere. This first solvation sphere has a C3 symmetry, meaning that there are two sets, of three waters each, with significantly different distances. A potential energy profile (PEP) and subsequent interaction energy analysis were carried out.

3.2.1 Energy Profile

The variable to analyse throughout the PEP was the distance between a *selected monomer* and the ion. The so called *monomers* were: two differently distanced water molecules (closer and farther), the glutamate residue and the aspartate residue. This ends up in a total of eight different PEPs, as there are four different monomers and two ions. The PEPs were generated through a sequence of singlepoint calculations varying along a distance (ion-monomer), also called *scan*. This was achieved using a local script and Gaussian (Version: Gaussian16) at the M06-2X/cc-pVTZ level of theory (LANL2DZ for K^+). The BSSE was corrected using the `counterpoise` method from Gaussian. Scans were done with a distance of 0.2 Å per step. The water molecules scans were performed for a total of 50 steps, so a total displacement distance of 10.0 Å. The residue scans had a total of 80 steps, giving a final displacement distance of 16.0 Å. The initial distance from the ion at which waters were is 2.71 and 2.73 Å, for closer and farther in K^+, and 2.28 and 2.37 Å, for closer and farther in Na^+. The initial distance ion-carbon (from the side-chain carboxylate), for the residues, is 2.2 Å.

From these results a distance of minimum energy for each residue was revealed. Those minimum energy geometries were later used to perform the EDA calculations.

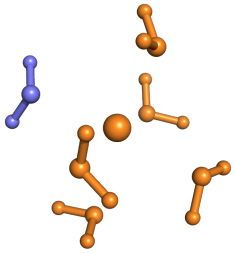
3.2.2 Energy Decomposition Analysis

The eight minimum energy geometries, obtained in the previous step, were subject to EDA calculations. For each of these, two main EDA calculations have been performed, in order to study the interactions between ion-environment and between monomer-environment [Figures 3.2 and 3.3], leaving a total of sixteen main EDA calculations. Additionally, the systems have been divided into discrete components in order to study the interaction between specific parts of the same system. The components include ion-monomer, ion-environment and monomer-environment; the monomer being either the residue or the selected water, while the environment is the remaining water molecules. This ends up in a total of 24 EDA calculations. These were carried out with EDA-NCI by Prof. Mandado [25].

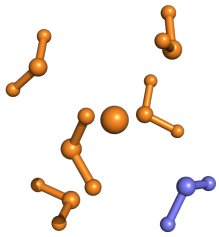
These calculations complement one another. The distance ion-residue of minimum energy is obtained from the scan. Afterwards, the structure is studied with EDA, obtaining the interaction energy between monomers. Thereby, this corresponds to the interaction energy of the minimum. If used as a reference it is possible to adjust all the absolute energy values from the PEP to represent instead the interaction energy using the following expression:

$$\begin{aligned} E_{\text{EDA}} &= (E_{min} - E_{ref}) c_{conv}^{te} \\ E_{ref} &= E_{min} - \frac{E_{\text{EDA}}}{c_{conv}^{te}} \end{aligned} \tag{3.1}$$

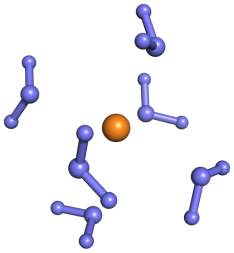
The c_{conv}^{te} represents the conversion factor between Hartree and kcal/mol.



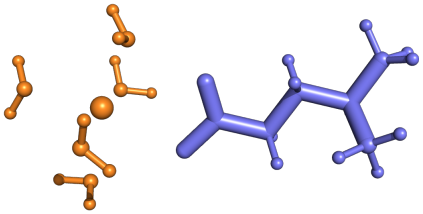
(a) **sCW**



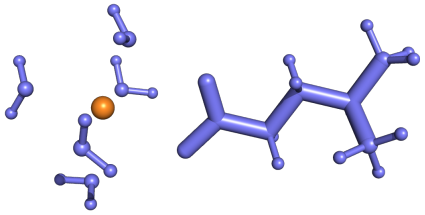
(b) **sFW**



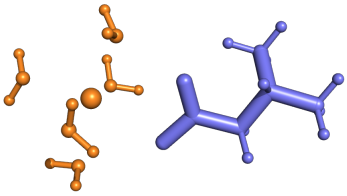
(c) **sIW**



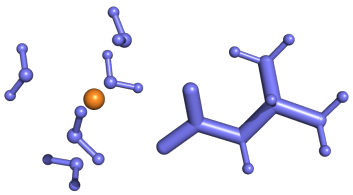
(d) **sGLU**



(e) **sIGLU**

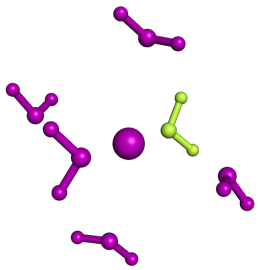


(f) **sASP**

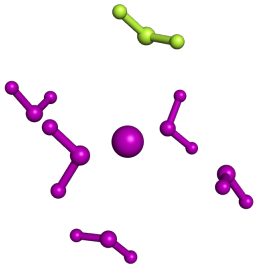


(g) **sIASP**

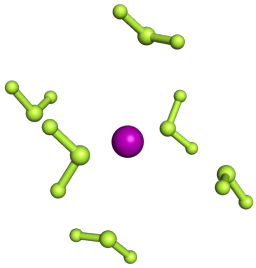
Figure 3.2: The different monomers used to perform EDA calculations for the all water (or one residue) of the Na^+ microsolvation. The systems show monomers as different colour atoms, stale blue or orange. EDA calculations were performed to obtain their interaction. Images (a) to (c) show the system containing only waters. Image (a) shows the closer water, while image (b) shows the farther water in a different colour. Image (c) represent the interaction between Na^+ and all waters. Images (d) and (f) represent the residue, glutamate and aspartate respectively, interaction with Na^+ and waters. Images (e) and (g) represent the Na^+ interaction with the residue, glutamate and aspartate respectively, and waters.



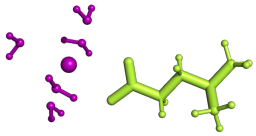
(a) **pCW**



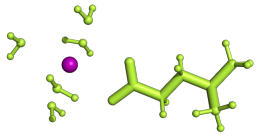
(b) **pFW**



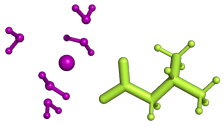
(c) **pIW**



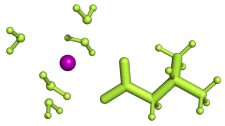
(d) **pGLU**



(e) **pIGLU**



(f) **pASP**

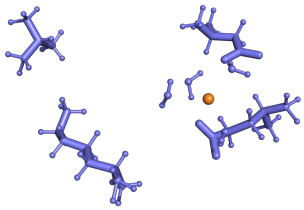


(g) **pIASP**

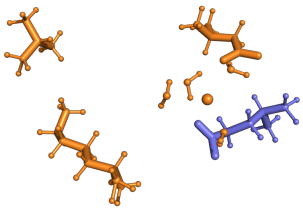
Figure 3.3: The different monomers used to perform EDA calculations for the all water (or one residue) of the K^+ microsolvation. The systems show monomers as different colour atoms, limon yellow or purple. EDA calculations were performed to obtain their interaction. Images (a) to (c) show the system containing only waters. Image (a) shows the closer water, while image (b) shows the farther water in a different colour. Image (c) represent the interaction between K^+ and all waters. Images (d) and (f) represent the residue, glutamate and aspartate respectively, interaction with K^+ and waters. Images (e) and (g) represent the K^+ interaction with the residue, glutamate and aspartate respectively, and waters.

3.3 DEKA System

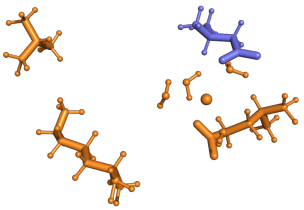
A frame from a previously performed MM-MD simulation, in which Na^+ passed through the DEKA, was subject to analysis. The system was stripped down to its essential components, with the DEKA system and the ion remaining. Subsequently, four waters were added, considering the simultaneous coordination of Na^+ with glutamate and aspartate from the DEKA ring. The water molecules were displayed in the position obtained from the optimized microsolvated structure. This structure was optimized considering two possibilities: constraining the DEKA and ion, or constraining only the DEKA. Calculations were carried out with Orca at the level of theory of M062X/cc-pVDZ. An EDA calculation was performed on the optimized geometry, considering different monomer combinations [Figure 3.4]. Additionally, the combinations between residue (either glutamate or aspartate), Na^+ and environment (everything left) were calculated.



(a) **DEKA-I**



(b) **DEKA-GLU**



(c) **DEKA-ASP**

Figure 3.4: The different monomers used to perform EDA calculations for the DEKA and $Na^+ \cdot 6\,H_2O$. The systems show monomers as different colour atoms, stale blue or orange. Image (a) shows Na^+ as one monomer and all remaining atoms as the other. Image (b) shows glutamate as monomer, and image (c) shows aspartate.

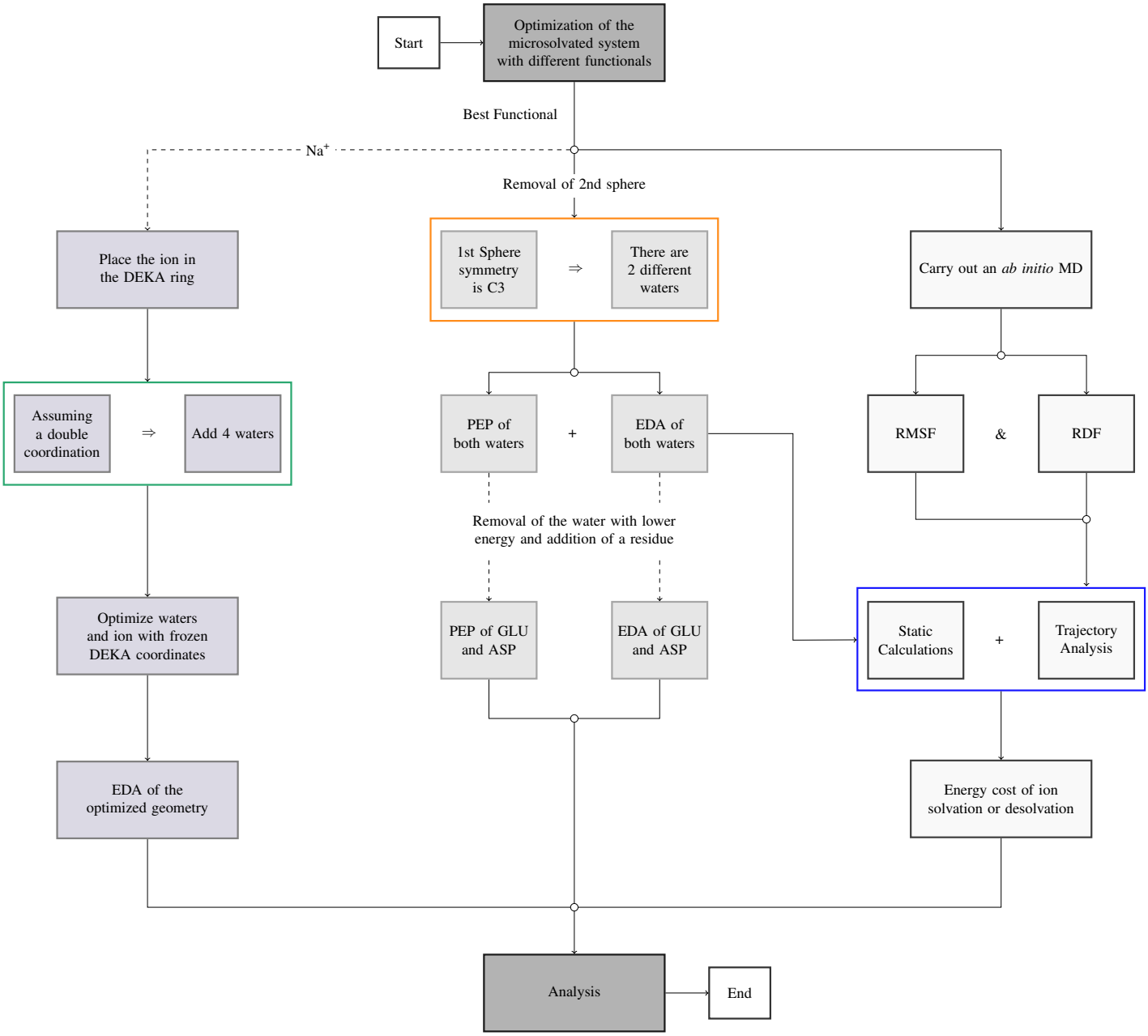


Figure 3.5: Simplified procedure showing the steps followed. It has been divided into 3 columns. The most left shows calculations performed with the DEKA ring. The center column shows the steps mainly related with ion interaction, either with water or amino acids. Finally, the most right shows the steps related to the microsolvation of the ions, mainly the MD simulations.

Chapter 4

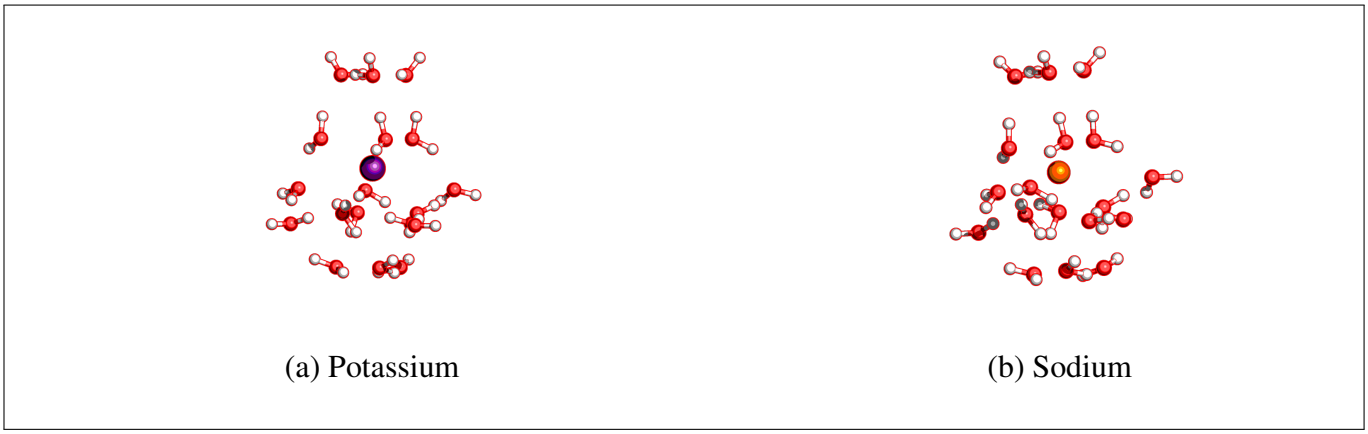
Results and Discussion

4.1 Microsolvation

Optimized Structure and Energy

Table 4.1: Energy obtained from microsolvation structures **A** and **B** for both ions, with the DFAs selected. **Interaction Energies serve purely an indicative purpose, there was not effectuated any BSSE correction.

Ion	Structure	DFA	Total Energy (Hartree)	Interaction Energy** (kcal/mol)
Sodium	**B**	B3LYP	-1538.3100	-205.6115
		ωB97XD	-1539.8709	-193.4718
		M06-2X	-1537.6521	-195.0744
		PW6B95D3	-1537.8496	-198.7861
		PDE0-DH	-1536.5118	-193.4347
	A	B3LYP	-1538.3262	-183.0167
		ωB97XD	-1537.8658	-171.9342
		M06-2X	-1537.6604	-178.0829
		PW6B95D3	-1539.8803	-178.3332
		PDE0-DH	-1536.5252	-201.8063
Potassium	**B**	B3LYP	-1404.1087	-155.2483
		ωB97XD	-1405.5072	-146.3026
		M06-2X	-1403.4342	-145.8030
		PW6B95D3	-1403.6788	-148.9537
		PDE0-DH	-1402.4232	-143.8446
	A	B3LYP	-1404.1277	-135.5230
		ωB97XD	-1403.6974	-126.6155
		M06-2X	-1403.4464	-130.9379
		PW6B95D3	–	–
		PDE0-DH	-1402.4392	-124.3341



(a) Potassium (b) Sodium

Figure 4.1: Optimization of Structure **A** of the Second Microsolvation Sphere of both Potassium (a) and Sodium (b), from the original structure **A**, using M06-2X/cc-pVDZ.

The initial microsolvation structures [Fig.3.1] were succesfully optimized. The energy obtained from the optimization from the original microsolvation structures [Figure 3.1] using the DFAs listed on **3.1 Microsolvation** are shown on Table 4.1, Unfortunately the structure **A** with K^+ optimization did not converge when using PW6B95D3. Results show that structure **A** presents the lower absolute energy in comparison with structure **B**, both for Na^+ and K^+. Henceforth, subsequent analysis will be performed with basis on structure **A**. The ion-water distances from the first sphere, on the K^+ are 2.71 and 2.73 Å , while on Na^+ are 2.28 and 2.37 Å. A certain degree of consistency across all functionals is observed:

- There are lower interaction energy values for Na^+ than for K^+, as expected.
- Calculations with K^+ show a less energetically negative value than Na^+. This results serves exclusively to confirm the consistency in results within each individual functional. Calculations with K^+ were done combining basis sets, thus it is not possible to conclude any physical-chemical property when comparing absolute energies with Na^+ systems, treated with only one basis set.

Table 4.2: CPU Time spent on the same system's single point energy calculation compared between DFAs. Time is expressed as 00h00'00" meaning hours–minutes–seconds in that particular order. All calculations were performed on the same cluster's node, minimizing the variation on computational resources available.

DFA	CPU Time
B3LYP	08'18"
ωB97XD	12'06"
M06-2X	09'22"
PW6B95D3	22'50"
PBE0-DH	2h28'02"

Despite the fact that all functionals were found to perform correctly, M06-2X has previously been shown to provide excellent performance when studying non-covalent interactions without the introduction of empirical dispersion corrections [22, 26, 27], particularly in the context of bio-organic compounds.Furthermore, M06-2X has been demonstrated to offer an excellent performance-to-computational cost ratio [Table 4.2]. Consequently, the following non-dynamical calculations were conducted using M06-2X.

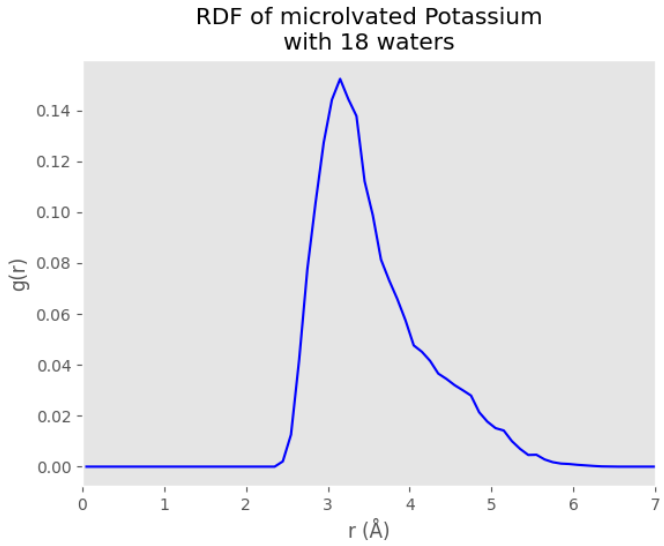
Molecular Dynamics

The MD simulations show that the distance between waters and Na^+ is shorter than with K^+, as it was expected. Moreover, throughout the simulations, there was a continuous exchange between first and second sphere waters. However, this process was much more noticeable for K^+ than Na^+ [Table 4.3]. The selectivity of DEKA towards Na^+ may be attributed to the ion inclination to maintain a *hydrated state* [6].

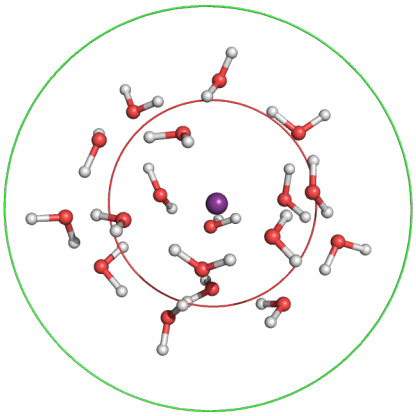
Table 4.3: Value for the RMSF (Å) of Oxygen atoms are displayed. Oxygen atoms are labelled with their respective ID. Atoms from 2 to 7 correspond to the first microsolvation sphere, while all others correspond to the second microsolvation sphere.

	Oxygen's ID	**Sodium**	**Potassium**
First Sphere	2	0.66	1.01
	3	0.60	0.80
	4	0.83	0.93
	5	0.75	0.92
	6	0.52	0.61
	7	0.74	0.71
	Average	0.67	0.82
Second Sphere	20	1.27	1.27
	21	0.51	0.61
	26	1.22	0.96
	27	0.78	0.53
	32	0.74	0.79
	33	1.16	0.78
	38	1.01	0.72
	39	0.41	0.90
	44	0.97	1.35
	45	0.55	0.50
	50	0.60	1.09
	51	0.60	0.80
	Average	0.77	0.82

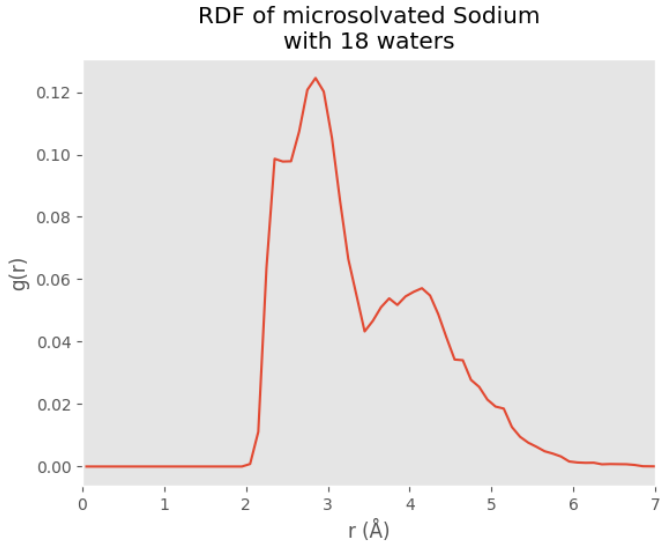
The values obtained for the RMSF in the Na^+ simulation are lower than those obtained in the K^+, particularly within the first solvation sphere, showing a maximum values of 0.83 and 0.93 Å, but also average values of 0.67 and 0.82 Å, respectively. The second sphere also presented, in general, smaller values for Na^+ than K^+, showing maximum values of 1.27 and 1.35 Å, but also average values of 0.77 and 0.82 Å, respectively. The fact that Na^+ waters present lower RMSF values indicates a stronger ion-water interaction.



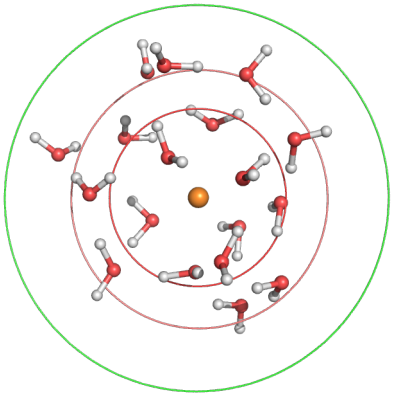
(a)



(b)



(c)



(d)

Figure 4.2: RDF of Oxygen atoms during the K^+ (a) and Na^+ (c) MD simulations. Images (b) and (d), for K^+ and Na^+, show the 1000th frame of the simulation, both with a green circle with 6.2 Å of radius, representing the limit of the cell. Image (b) red circle has 3.15 Å, while red circles at image (d) have 2.85 and 4.15 Å, matching their respective RDFs' peaks.

The RDF plots [Figure 4.2] show that both spheres are somewhat maintained throughout the Na^+ simulation, as its graph show two peaks. However, spheres merge together for K^+, as its graph only shows one peak. Additionally, the radius at which the (first) peak is reached is bigger for K^+, at 3.15 Å, than for Na^+, at 2.85 Å. The first non-zero value is obtained at a distance of 2.45 Å and $g(r)$ value of 0.0021 for K^+, and at 2.05 Å and 0.0007 of $g(r)$ value. This results are consistent with other calculations performed at a higher level of theory [28, 29].

4.2 Interaction Energy

In order to study the interaction energy between ion and residues or water molecules, PEPs and EDA calculations were performed. From the optimized microsolvation geometry, the second sphere was completely removed, as it could not possibly fit the Na_v. The remaining first sphere, with six waters, shows a C3 symmetry, resulting in two differently distanced (to the ion) water molecule groups [Figure 4.1]. One water molecule of each group was individually studied. These water molecules were named closer and farther. The water with lowest energy, consequently the easiest to desolvate, was the farther [Tables 4.4 and 4.5], as expected. This water was removed, and either glutamate or aspartate residues were set on its place, building the set of systems studied [Figures 3.2 and 3.3].

The scan calculations yielded the PEP as a function of distance towards the ion. Both K^+ and Na^+ systems obtained the minimum energy distance on the glutamate and aspartate at 3.4 and 3.2 Å, respectively. In contrast, the water solvation sphere was maintained for the EDA calculations [Tables 4.4 and 4.5]. These structures are not fully optimized, which may lead to some inaccuracy on the representation of the results.

Table 4.4: Main results obtained from EDA calculations performed on the four Na^+ systems obtained through the microsolvation sphere optimization and the scans. Systems maintain the same nomenclature used before and shown at Figure 3.2. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy.

Energy (kcal/mol)	**sCW**	**sFW**	**sIW**	**sGLU**	**sIGLU**	**sASP**	**sIASP**
Electrostatic	-14.409	-12.228	-104.300	-90.351	-193.790	-90.217	-192.428
Pauli	10.264	10.587	36.681	17.013	32.361	15.538	32.269
Dispersion	0.418	-1.304	-10.074	-16.223	-39.746	-15.757	-39.561
Induction	-6.354	-5.599	-22.104	-0.527	5.340	-0.207	5.188
Total	-10.072	-8.536	-99.716	-90.015	-195.675	-90.569	-194.374

Table 4.5: Main results obtained from EDA calculations performed on the four K^+ systems obtained through the microsolvation sphere optimization and the scans. Systems maintain the nomenclature used before and shown at Figure 3.3. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy.

Energy (kcal/mol)	**pCW**	**pFW**	**pIW**	**pGLU**	**pIGLU**	**pASP**	**pIASP**
Electrostatic	-12.699	-11.574	-83.524	-92.473	-177.613	-97.500	-183.775
Pauli	6.295	6.032	32.227	5.170	32.319	18.300	37.814
Dispersion	-2.866	-2.727	-10.367	-4.490	-15.744	-12.966	-16.541
Induction	-1.002	-1.028	-11.671	-1.903	-10.738	-2.441	-11.84
Total	-10.263	-9.289	-73.275	-93.621	-171.636	-94.53	-174.201

The subsequent EDA calculations on the water solvation sphere, when analysing the ion-waters (**IW**) interactions, yielded a more negative value for Na^+ than for K^+, with total interaction energies of -99.716 and -73.275 kcal/mol [Tables 4.4 and 4.5], in agreement with the MD simulations. The interaction energy of the closer water larger than the farther water, for Na^+ and K^+ respectively, with -10.072 and -10.263 kcal/-mol for the closer water, and -8.536 and -9.289 kcal/mol. The interaction energy of glutamate (**GLU**) and

aspartate (**ASP**) is very close for both ions, though closer in Na^+ systems ($|E_{\textbf{sGLU}} - E_{\textbf{sASP}}| = 0.554$ kcal/-mol), than in K^+ systems ($|E_{\textbf{pGLU}} - E_{\textbf{pASP}}| = 0.909$ kcal/mol). This interaction will be referred as binding energy. On the other hand, when comparing the interaction energies between same residues but different ions, the K^+ systems show a higher affinity between solvated ion and residue ($E_{\textbf{pGLU}} - E_{\textbf{sGLU}} = -3.606$ kcal/mol) than the Na^+ systems ($E_{\textbf{pASP}} - E_{\textbf{sASP}} = -3.961$ kcal/mol). This also occurs when considering the desolvation and residue binding simultanously. The energy necessary to desolvate a water molecule is the exact opposite of the interaction energy between solvated ion and that specific water. Therefore, for Na^+ is 10.072 kcal/mol, and for K^+ is 8.536 kcal/mol. Combining these energies (desolvation and binding) does also favour K^+ over Na^+, both for glutamate binding (4.1) and aspartate binding (4.2).

$$\begin{aligned} Na^+ &\rightarrow E_{Na^+\text{-desolv-glu\_bind}} = E_{\textbf{sGLU}} - E_{\textbf{sFW}} = -81.479 \text{ kcal/mol} \\ K^+ &\rightarrow E_{K^+\text{-desolv-glu\_bind}} = E_{\textbf{pGLU}} - E_{\textbf{pFW}} = -84.332 \text{ kcal/mol} \end{aligned} \tag{4.1}$$

$$\begin{aligned} Na^+ &\rightarrow E_{Na^+\text{-desolv-asp\_bind}} = E_{\textbf{sASP}} - E_{\textbf{sFW}} = -82.033 \text{ kcal/mol} \\ K^+ &\rightarrow E_{K^+\text{-desolv-asp\_bind}} = E_{\textbf{pASP}} - E_{\textbf{pFW}} = -85.241 \text{ kcal/mol} \end{aligned} \tag{4.2}$$

The difference in energy, for the desolvation-binding process, between ions within the same residue are of $E_{K^+\text{-desolv-glu\_bind}} - E_{Na^+\text{-desolv-glu\_bind}} = -2.843$ kcal/mol, for glutamate, and $E_{K^+\text{-desolv-asp\_bind}} - E_{Na^+\text{-desolv-asp\_bind}} = -3.762$, favouring K^+.

Table 4.6: Results of EDA calculations perfermed on the divided systems of Na^+. All energy values are displayed in kcal/mol. The division consist on the study of monomer-ion, monomer-environment and ion-environment. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy

	sCW			**sFW**		
Energy	Closer-Ion	Closer-Env	Ion-Env	Farther-Ion	Farther-Env	Ion-Env
Electrostatic	-23.835	6.105	-88.136	-19.716	4.24	-92.266
Pauli	8.501	1.745	30.331	6.582	3.742	32.338
Dispersion	-1.441	-2.017	-8.434	-1.569	-3.106	-8.042
Induction	-5.086	-0.609	-18.838	-4.474	-0.847	-19.81
Total	-21.843	5.22	-85.008	-19.162	4.025	-87.709
	sGLU			**sASP**		
Energy	Glutamate-Ion	Glutamate-Env	Ion-Env	Aspartate-Ion	Aspartate-Env	Ion-Env
Electrostatic	-120.166	18.407	-92.266	-118.477	17.097	-92.266
Pauli	3.764	12.493	32.338	3.635	11.228	32.338
Dispersion	-20.787	-7.378	-8.042	-20.353	-7.203	-8.042
Induction	10.075	-3.908	-19.81	9.691	-3.814	-19.81
Total	-127.011	19.598	-87.709	-125.402	17.294	-87.709

These systems were divided to analyse specific interactions [Tables 4.6 and 4.7]. Such division may cause some inaccuracies on the results. The ion-environment interaction, in **FW**, **GLU** and **ASP**, yielded the exact same result, as systems are equivalent. The residue-environment interaction is of special interest, resulting on higher repulsion (Pauli) values for Na^+ in comparison with K^+. This may be explained by the shorter distance at which water molecules are found in Na^+ solvation sphere. The low values for the

Table 4.7: Results of EDA calculations perfermed on the divided systems of K^+. All energy values are displayed in kcal/mol. The division consist on the study of monomer-ion, monomer-environment and ion-environment. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy.

	pCW			**pFW**		
Energy	Closer-Ion	Closer-Env	Ion-Env	Farther-Ion	Farther-Env	Ion-Env
Electrostatic	-16.578	2.326	-70.061	-14.357	1.451	-72.028
Pauli	5.967	0.274	27.086	5.653	0.314	27.337
Dispersion	-2.163	-0.96	-8.781	-2.031	-0.995	-8.682
Induction	-1.912	-0.277	-9.754	-1.885	-0.308	-9.657
Total	-14.673	1.362	-61.459	-12.61	0.463	-63.317
	pGLU			**pASP**		
Energy	Glutamate-Ion	Glutamate-Env	Ion-Env	Aspartate-Ion	Aspartate-Env	Ion-Env
Electrostatic	-117.85	18.149	-72.028	-126.369	19.543	-72.028
Pauli	7.971	2.641	27.337	5.172	4.139	27.337
Dispersion	-8.128	-3.177	-8.682	-0.738	-4.104	-8.682
Induction	-0.929	-2.99	-9.995	-0.607	-3.234	-9.657
Total	-118.84	14.611	-63.317	-122.442	16.332	-63.317

repulsion energy in K^+, at the residue-environment interaction, could rise inside the protein channel, as water molecules may be closer to other residues, either from DEKA or not.

This is further corroborated by the interaction ion-residue, deeply dominated by the electrostatic energy, resulting in more stable values for Na^+. The difference between the ion-residue interaction values between ions is of $E_{\textbf{pGLU}\_glu-env} - E_{\textbf{sGLU}\_glu-env} = -8.171$ kcal/mol, for glutamate, and $E_{\textbf{pGLU}\_asp-env} - E_{\textbf{sGLU}\_asp-env} = -2.960$ kcal/mol, for aspartate. So a higher desolvation may contribute to reduce the repulsion energy in Na^+, favouring its coordination with DEKA over K^+.

4.3 DEKA System

Table 4.8: Results of EDA calculations performed on the DEKA system following the the nomenclature employed at Figure 3.4. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy.

Energy (kcal/mol)	**DEKA-I**	**DEKA-GLU**	**DEKA-ASP**
Electrostatic	-208.941	-123.964	-86.726
Pauli	30.991	66.421	40.035
Dispersion	-44.73	-42.741	-30.586
Induction	9.836	-10.551	-7.854
Total	-212.67	-110.745	-85.063

The energy cost of desolvation and binding between glutamate and aspartate, within Na^+, is very similar ($\Delta E = 0.554$ kcal/mol). Thus, a simultaneous coordination of both residues was considered. A frame of

a Na^+ passing through the DEKA ring, from a previous MM-MD simulation, was used as an initial position. From geometry all waters were removed and four waters, taken from the first solvation sphere of the optimized microsolvated structure, were added. This step was done because the waters' positions obtained with the FF during the MD did not seem chemically accurate. Consequently, this new structure was optimized twice: constraining the DEKA position and ion or only the DEKA. The former was rejected as the ion-residue interaction was compromised by the initial water molecules disposition. Thus, the optimization constraining exclusively the DEKA [Figure 3.4] was employed. The system was subject to EDA calculations [Table 4.8 and 4.9]. The energy of glutamate-ion interaction (-121.631 kcal/mol) on the DEKA is higher than that of the glutamate-ion interaction extracted from **sGLU** (-127.011 kcal/mol), but still lower than the glutamate-ion from pGLU (-118.840 kcal/mol). The interaction residue-environment presents a high repulsion value which is countered by a strong polarization energy.

Table 4.9: Results of EDA calculations performed on the divided DEKA system. The divisions consist on residue-sodium, residue-environment and sodium-environment. It is important to take in consideration that 'Dispersion' represents all the non-induction polarization energy.

	DEKA-GLU			**DEKA-ASP**		
Energy	Glutamate-Ion	Glutamate-Env	Ion-Env	Aspartate-Ion	Aspartate-Env	Ion-Env
Electrostatic	-116.648	-19.658	-110.459	-77.981	-13.491	-130.532
Pauli	8.866	57.539	26.898	0.080	42.014	30.389
Dispersion	-17.349	-30.135	-16.593	-6.934	-22.65	-22.611
Induction	3.400	-13.626	-9.325	3.149	-9.669	-10.019
Total	-121.632	-5.876	-109.39	-81.620	-3.793	-132.665

The EDA calculations on the DEKA ring reveal that, overall, the predominant interactions are electrostatic. There was a higher interaction with glutamate than with aspartate, though this is not necessarily significant as the initial position of glutamate in that particular frame was favouring the interaction with the ion, with a distance $-COO_{glu}$ to Na^+ of 5.2 Å, in comparison with the initial distance $-COO_{asp}$ to Na^+ of 6.6 Å. The final distances obtained with the optimized structure were of 3.1 and 4.9 Å, for glutamate and aspartate respectively.

The coordination with the DEKA ring may be favoured towards Na^+ when considering the repulsion from the protein structure. Due to the limit space of the channel, the shorter distances between ion and water molecules in Na^+ could be determinant to explain the DEKA selectivity. Overall, K^+ presents lower energies of desolvation-binding with either glutamate or aspartate, though this changes when exclusively considering the interaction between ion and residue, without waters.

Further calculations to elucidate the DEKA selectivity mechanism could consist on: using more frames with the DEKA ring; performing calculations of K^+ at the DEKA ring or performing calculations on the DEKA ring with different amounts of water. The study of the interaction with the two remaining residues from DEKA, alanine and lysine, could be very helpful. The lysine residue has a positively charged side-chain, that could destabilize the solvation sphere or even repel the passing ion. As K^+ has a less stable solvation sphere this effects may affect it more intensely, than Na^+. Employing different methodologies could be helpful, for instance alchemical methods in combination with previous calculations could reveal the different behaviours of Na^+ and K^+ within similar situations.

In conclusion, the calculations did not yield a comprehensive understanding of the DEKA selectivity mech-

anism. However, they did suggest several potential hypothetical mechanisms that could be investigated further in the future. It has been demonstrated that the structure and properties of the Na^+ and K^+ microsolvation are markedly distinct, which could serve as a pivotal factor in the Na_v selectivity mechanism. Additionally, the binding to glutamate or aspartate amino acids results in disparate energies between ions, predisposing the K^+ coordination. This could undergo a shift when contemplating alternative scenarios, and thus, this could be a subject of further investigation.

Chapter 5

Conclusion

The $Na_v1.4\beta$ is the latest Na_v that has been resolved. A cryo-EM analysis set up a final resolution of 3.2 Å for the protein, and 2.8 Åwithin its SF. The SF is conformed by four amino acids, the so called DEKA ring, ubiquos among all eukaryotic Na_v. Previous enhanced MM-MD simulations of the $Na_v1.4$ have been inconclusive on the matter, not showing any preference of Na^+ over K^+. Thus, a study of intermolecular interactions using QM methodologies is here performed. In order to simulate the $Na_v1.4$, key-elements from the DEKA system, namely glutamate and aspartate, have been employed.

The microsolvated Na^+ and K^+ were optimized with a number of DFAs has been performed. The M06-2X functional was finally selected as the most appropiate. Subsequent MD simulations were performed with the microsolvated ions. These calculations elucidated the behaviour of the solvation sphere and showed that water molecules tend to stay closer to Na^+ than K^+.

A model was build from the optimised microsolvated structure. These models were subject to scans which revealed the energy over the distance towards the ion from glutamate and aspartate, which defined the models to be employed on subsequent calculations.

The EDA calculations were carried out to further investigate the interaction energy. From these calculations, the desolvation of a single water (the less energetically bound) was studied for both ions, revealing an energy cost higher for Na^+ than for K^+. In addition, the binding energy for K^+ and either glutamate or aspartate residues is lower than those obtained for Na^+. The overall desolvation and following binding showed a preference for K^+ over Na^+. However, when the energy contributions are analysed separatedly, it is noticeable that the repulsion energy is always higher for Na^+. Furthermore, the analysis of the divided systems shows that the ion-residue interaction is always more stable in Na^+ than K^+.

Finally, a frame from the original MM-MD of Na^+ simulation was studied with EDA. A double coordination between Na^+ with glutamate and aspartate was considered. The overall results show that the interactions are dominated by the electrostatic energy. Different coordination schemes with the DEKA, changing the number of water molecules in close proximity to the ion, may take an important role on its selectivity.

Overall calculations are insufficient to give a clear mechanism for the selectivity of the DEKA ring. However, conservation of the solvation sphere may be very important for ion selectivity. Calculations indicate that Na^+ is more favoured than K^+ to maintain the aforementioned solvation sphere while passing through the DEKA. Many factors may contribute to this behaviour, in this work the binding to glutamate or aspartate is particularly emphasised. However, interactions between water molecules and lysine (also on the DEKA

ring) may be relevant. As the lysine side chain is positively charged, a wide solvation sphere could lead to a stronger interaction between water and lysine. This interaction may destabilise the ion solvation as both the ion and the residue are positively charged.

Bibliography

[1] Samuel Schindler. Model, theory, and evidence in the discovery of the dna structure. *The British Journal for the Philosophy of Science*, 59(4):619–658, December 2008.

[2] Jack Copeland. *On Computable Numbers,with an Application to the Entscheidungsproblem*, chapter 1, pages 58–72. Oxford University Press, 2004.

[3] Axel D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140(18), April 2014.

[4] Natércia F. Brás, Rui P.P. Neves, Nuno M.F.S.A. Cerqueira, Sergio F. Sousa, Pedro A. Fernandes, and Maria J. Ramos. *Computational Biochemistry*. Elsevier, 2015.

[5] Justin A. MacDonald. *Signal Transduction Pathways and the Control of Cellular Responses to External Stimuli*, chapter 4, pages 87–123. Wiley, 2004.

[6] Georgi V. Petkov. *Ion Channels*, chapter 16, pages 387–427. Elsevier, 2009.

[7] David S. Goodsell, Christine Zardecki, Luigi Di Costanzo, Jose M. Duarte, Brian P. Hudson, Irina Persikova, Joan Segura, Chenghua Shao, Maria Voigt, John D. Westbrook, Jasmine Y. Young, and Stephen K. Burley. Rcsb protein data bank: Enabling biomedical research and drug discovery. *Protein Science*, 29(1):52–65, November 2019.

[8] Christian A. Hubner. Ion channel diseases. *Human Molecular Genetics*, 11(20):2435–2445, October 2002.

[9] Beata Dworakowska and Krzysztof Dołowy. Ion channels-related diseases. *Acta Biochimica Polonica*, 47(3):685–703, 2000.

[10] Heather A. O'Malley and Lori L. Isom. Sodium channel subunits: Emerging targets in channelopathies. *Annual Review of Physiology*, 77(1):481–504, February 2015.

[11] 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 na v 1.4 in complex with 1. *Science*, 362(6412), October 2018.

[12] Alex Bateman, Maria-Jesus Martin, Sandra Orchard, Michele Magrane, Rahat Agivetova, Shadab Ahmad, Emanuele Alpi, Emily H Bowler-Barnett, Ramona Britto, Borisas Bursteinas, Hema Bye-A-Jee, Ray Coetzee, Austra Cukura, Alan Da Silva, Paul Denny, Tunca Dogan, ThankGod Ebenezer, Jun Fan, Leyla Garcia Castro, Penelope Garmiri, George Georghiou, Leonardo Gonzales, Emma Hatton-Ellis, Abdulrahman Hussein, Alexandr Ignatchenko, Giuseppe Insana, Rizwan Ishtiaq, Pet-

teri Jokinen, Vishal Joshi, Dushyanth Jyothi, Antonia Lock, Rodrigo Lopez, Aurelien Luciani, Jie Luo, Yvonne Lussi, Alistair MacDougall, Fabio Madeira, Mahdi Mahmoudy, Manuela Menchi, Alok Mishra, Katie Moulang, Andrew Nightingale, Carla Susana Oliveira, Sangya Pundir, Guoying Qi, Shriya Raj, Daniel Rice, Milagros Rodriguez Lopez, Rabie Saidi, Joseph Sampson, Tony Sawford, Elena Speretta, Edward Turner, Nidhi Tyagi, Preethi Vasudev, Vladimir Volynkin, Kate Warner, Xavier Watkins, Rossana Zaru, Hermann Zellner, Alan Bridge, Sylvain Poux, Nicole Redaschi, Lucila Aimo, Ghislaine Argoud-Puy, Andrea Auchincloss, Kristian Axelsen, Parit Bansal, Delphine Baratin, Marie-Claude Blatter, Jerven Bolleman, Emmanuel Boutet, Lionel Breuza, Cristina Casals-Casas, Edouard de Castro, Kamal Chikh Echioukh, Elisabeth Coudert, Beatrice Cuche, Mikael Doche, Dolnide Dornevil, Anne Estreicher, Maria Livia Famiglietti, Marc Feuermann, Elisabeth Gasteiger, Sebastien Gehant, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz-Gumowski, Ursula Hinz, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Guillaume Keller, Arnaud Kerhornou, Vicente Lara, Philippe Le Mercier, Damien Lieberherr, Thierry Lombardot, Xavier Martin, Patrick Masson, Anne Morgat, Teresa Batista Neto, Salvo Paesano, Ivo Pedruzzi, Sandrine Pilbout, Lucille Pourcel, Monica Pozzato, Manuela Pruess, Catherine Rivoire, Christian Sigrist, Karin Sonesson, Andre Stutz, Shyamala Sundaram, Michael Tognolli, Laure Verbregue, Cathy H Wu, Cecilia N Arighi, Leslie Arminski, Chuming Chen, Yongxing Chen, John S Garavelli, Hongzhan Huang, Kati Laiho, Peter McGarvey, Darren A Natale, Karen Ross, C R Vinayaka, Qinghua Wang, Yuqi Wang, Lai-Su Yeh, Jian Zhang, Patrick Ruch, and Douglas Teodoro. Uniprot: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489, November 2020.

[13] Isabelle Favre, Edward Moczydlowski, and Laurent Schild. On the structural basis for ionic selectivity among na+, k+, and ca2+ in the voltage-gated sodium channel. *Biophysical Journal*, 71(6):3110–3125, December 1996.

[14] Carmen Domene. *Computer Simulation of Ion Channels*, chapter 6, pages 1737–1745. The Royal Society of Chemistry, December 2016.

[15] Christopher J. Cramer. *Essentials of Computational Chemistry*. Wiley, 2004.

[16] `https://psicode.org/psi4manual/master/dft_byfunctional.html#all-available`. cited: 10/06/2024.

[17] Frank Jensen. *Introduction to Computational Chemistry*. Wiley, 2006.

[18] Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, June 2017.

[19] Alberto Otero de la Roza and Gino A. DiLabio. *Non-Covalent Interactions in Quantum Chemistry and Physics Theory and Applications*. Elsevier Science Technology Books, 2017.

[20] Marcos Mandado and José M. Hermida-Ramón. Electron density based partitioning scheme of interaction energies. *Journal of Chemical Theory and Computation*, 7(3):633–641, February 2011.

[21] Jan Moens, Goedele Roos, Pablo Jaque, Frank DeProft, and Paul Geerlings. Can electrophilicity act as a measure of the redox potential of first-row transition metal ions? *Chemistry – A European Journal*, 13(33):9331–9343, November 2007.

[22] Christoph Plett, Stefan Grimme, and Andreas Hansen. Conformational energies of biomolecules in solution: Extending the mpconf196 benchmark with explicit water molecules. *Journal of Computational Chemistry*, 45(7):419–429, November 2023.

[23] Kirk A. Peterson and Thom H. Dunning. Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms al–ar, and the first row atoms b–ne revisited. *The Journal of Chemical Physics*, 117(23):10548–10560, December 2002.

[24] S. Chiodo, N. Russo, and E. Sicilia. Lanl2dz basis sets recontracted in the framework of density functional theory. *The Journal of Chemical Physics*, 125(10), September 2006.

[25] `https://github.com/marcos-mandado/EDA-NCI`. cited: 11/06/2024.

[26] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1–3):215–241, July 2007.

[27] Nicolás Ramos-Berdullas, Ignacio Pérez-Juste, Christian Van Alsenoy, and Marcos Mandado. Theoretical study of the adsorption of aromatic units on carbon allotropes including explicit (empirical) dft dispersion corrections and implicitly dispersion-corrected functionals: the pyridine case. *Physical Chemistry Chemical Physics*, 17(1):575–587, 2015.

[28] S. Sikander Azam, Thomas S. Hofer, Bernhard R. Randolf, and Bernd M. Rode. Hydration of sodium(i) and potassium(i) revisited: A comparative qm/mm and qmcf md simulation study of weakly hydrated ions. *The Journal of Physical Chemistry A*, 113(9):1827–1834, February 2009.

[29] Anan Tongraar, Klaus R. Liedl, and Bernd M. Rode. Bornoppenheimer ab initio qm/mm dynamics simulations of na+and k+in water: from structure making to structure breaking effects. *The Journal of Physical Chemistry A*, 102(50):10340–10347, November 1998.

[30] `https://www.rcsb.org/structure/6AGF`. cited: 11/06/2024.

Supplementary Material and Additional Information

List of Acronyms

→ **B**

- **BS**: Basis Set
- **BSSE**: Basis Set Superposition Error
- **BO**: Born-Oppenheimer

→ **C**

- **CPU**: Central Processing Unit
- **cryo-EM**: Cryogenic Electron Microscopy

→ **D**

- **DEKA**: Asp^{406}/Glu^{761}/Lys^{1244}/Ala^{1536}
- **DFT**: Density Functional Theory
- **DFA**: Density Functional Approximation
- **DNA**: Deoxyribonucleic Acid

→ **E**

- **EDA**: Energy Decomposition Analysis

→ **F**

- **FF**: Force Field

→ **G**

- **GGA**: Generalized Gradient Approximation

→ **H**

- **HF**: Hartree-Fock
- **HK**: Hohenberg-Kohn

→ **K**

- $\mathbf{K^+}$: Potassium Cation

- **KS**: Kohn-Sham

  → **L**
- **LDA**: Local Density Approximation

  → **M**
- **MD**: Molecular Dynamics
- **MM**: Molecular Mechanics
- **MP2**: Møller-Plesset of Second order
- **MPPT**: Møller-Plesset Perturbation Theory
- **mGGA**: meta Generalized Gradient Approximation

  → **N**
- **Na^+**: Sodium Cation
- **Na_v**: Voltage-Gated Sodium Channel
- **NCI**: Non-Covalent Interactions
- **NMR**: Nuclear Magnetic Resonance

  → **P**
- **PEP**: Potential Energy Curve
- **PES**: Potential Energy Surface

  → **Q**
- **QM**: Quantum Mechanics
- **QC**: Quadratic Convergence

  → **R**
- **RDF**: Radial Distribution Function
- **RMSF**: Root-Mean-Square Fluctuation
- **RNA**: Ribonucleic Acid

  → **S**
- **SF**: Selectivity Filter

  → **T**
- **TFD**: Thomas-Fermi-Dirac

  → **U**
- **UEG**: Uniform Electron Gas

  → **W**

- **WFT**: Wave Function Theory

  $\rightarrow$ **X**

- **XC**: Exchange-Correlation

Protein Sequence

Here is the $Na_v1.4$-$\beta 1$ amino acid sequence, chain A corresponds to the α subunit and chain B to the β subunit, displayed in FASTA format:

> 6AGF_1|Chain A|Sodium channel protein **type** 4 subunit alpha|Homo sapiens (9606)\\
MASWSHPQFEKGGGARGGSGGGSWSHPQFEKGFDYKDDDDKGTMARPSLCTLVPLGPECLRPFTRESLAAIEQRAVEEEARL
QRNKQMEIEEPERKPRSDLEAGKNLPMIYGDPPPEVIGIPLEDLDPYYSNKKTFIVLNKGKAIFRFSATPALYLLSPFSVVR
RGAIKVLIHALFSMFIMITILTNCVFMTMSDPPPWSKNVEYTFTGIYTFESLIKILARGFCVDDFTFLRDPWNWLDFSVIMM
AYLTEFVDLGNISALRTFRVLRALKTITVIPGLKTIVGALIQSVKKLSDVMILTVFCLSVFALVGLQLFMGNLRQKCVRWPP
PFNDTNTTWYSNDTWYGNDTWYGNEMWYGNDSWYANDTWNSHASWATNDTFDWDAYISDEGNFYFLEGSNDALLCGNSSDAG
HCPEGYECIKTGRNPNYGYTSYDTFSWAFLALFRLMTQDYWENLFQLTLRAAGKTYMIFFVVIIFLGSFYLINLILAVVAMA
YAEQNEATLAEDKEKEEEFQQMLEKFKKHQEELEKAKAAQALEGGEADGDPAHGKDCNGSLDTSQGEKGAPRQSSSGDSGIS
DAMEELEEAHQKCPPWWYKCAHKVLIWNCCAPWLKFKNIIHLIVMDPFVDLGITICIVLNTLFMAMEHYPMTEHFDNVLTVG
NLVFTGIFTAEMVLKLIAMDPYEYFQQGWNIFDSIIVTLSLVELGLANVQGLSVLRSFRLLRVFKLAKSWPTLNMLIKIIGN
SVGALGNLTLVLAIIVFIFAVVGMQLFGKSYKECVCKIALDCNLPRWHMHDFFHSFLIVFRILCGEWIETMWDCMEVAGQAM
CLTVFLMVMVIGNLVVLNLFLALLLSSFSADSLAASDEDGEMNNLQIAIGRIKLGIGFAKAFLLGLLHGKILSPKDIMLSLG
EADGAGEAGEAGETAPEDEKKEPPEEDLKKDNHILNHMGLADGPPSSLELDHLNFINNPYLTIQVPIASEESDLEMPTEEET
DTFSEPEDSKKPPQPLYDGNSSVCSTADYKPPEEDPEEQAEENPEGEQPEECFTEACVQRWPCLYVDISQGRGKKWWTLRRA
CFKIVEHNWFETFIVFMILLSSGALAFEDIYIEQRRVIRTILEYADKVFTYIFIMEMLLKWVAYGFKVYFTNAWCWLDFLIV
DVSIISLVANWLGYSELGPIKSLRTLRALRPLRALSRFEGMRVVVNALLGAIPSIMNVLLVCLIFWLIFSIMGVNLFAGKFY
YCINTTTSERFDISEVNNKSECESLMHTGQVRWLNVKVNYDNVGLGYLSLLQVATFKGWMDIMYAAVDSREKEEQPQYEVNL
YMYLYFVIFIIFGSFFTLNLFIGVIIDNFNQQKKKLGGKDIFMTEEQKKYYNAMKKLGSKKPQKPIPRPQNKIQGMVYDLVT
KQAFDITIMILICLNMVTMMVETDNQSQLKVDILYNINMIFIIIFTGECVLKMLALRQYYFTVGWNIFDFVVVILSIVGLAL
SDLIQKYFVSPTLFRVIRLARIGRVLRLIRGAKGIRTLLFALMMSLPALFNIGLLLFLVMFIYSIFGMSNFAYVKKESGIDD
MFNFETFGNSIICLFEITTSAGWDGLLNPILNSGPPDCDPNLENPGTSVKGDCGNPSIGICFFCSYIIISFLIVVNMYIAII
LENFNVATEESSEPLGEDDFEMFYETWEKFDPDATQFIAYSRLSDFVDTLQEPLRIAKPNKIKLITLDLPMVPGDKIHCLDI
LFALTKEVLGDSGEMDALKQTMEEKFMAANPSKVSYEPITTTLKRKHEEVCAIKIQRAYRRHLLQRSMKQASYMYRHSHDGS
GDDAPEKEGLLANTMSKMYGHENGNSSSPSPEEKGEAGDAGPTMGLMPISPSDTAWPPAPPPGQTVRPGVKESLV
> 6AGF_2|Chain B|Sodium channel subunit beta-1|Homo sapiens (9606)\\
MGRLLALVVGAALVSSACGGCVEVDSETEAVYGMTFKILCISCKRRSETNAETFTEWTFRQKGTEEFVKILRYENEVLQLEE
DERFEGRVVWNGSRGTKDLQDLSIFITNVTYNHSGDYECHVYRLLFFENYEHNTSVVKKIHIEVVDKANRDMASIVSEIMMY
VLIVVLTIWLVAEMIYCYKKIAAATETAAQENASEYLAITSESKENCTGVQVAE

The protein can be obtained through the Protein Data Bank (PDB) web with the ID: **6AGF** [30].

Equipment and Software

Calculations, images and data were obtained using multiple software, here is a list of all the software used and their respective versions:

- Gaussian - Version: Gaussian16
- Orca - Version: 5.0.4
- AMBER - Version: amber20
- MDAnalysis - Version: 2.7.0
- Obabel - Version: 3.1.1
- Pymol - Version: 2.5.0
- Molden - Version: 6.2

Calculations were done with **Universidade de Vigo** computational resources.

Images

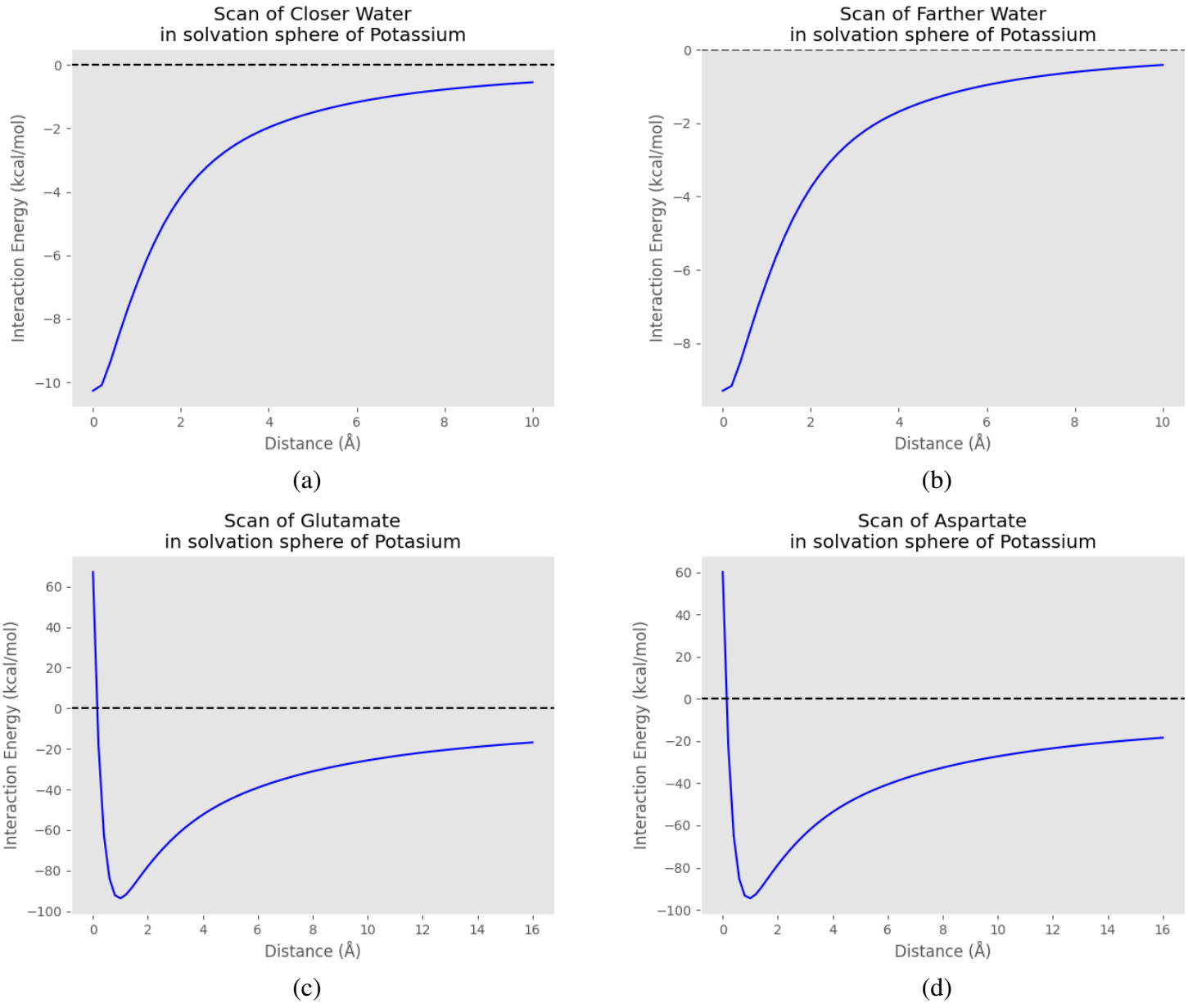


Figure 5.1: Results from scans on K^+ systems, that are shown in Figure 3.3. The x-axis represents the displacement from the original position, written as 'Distance'.

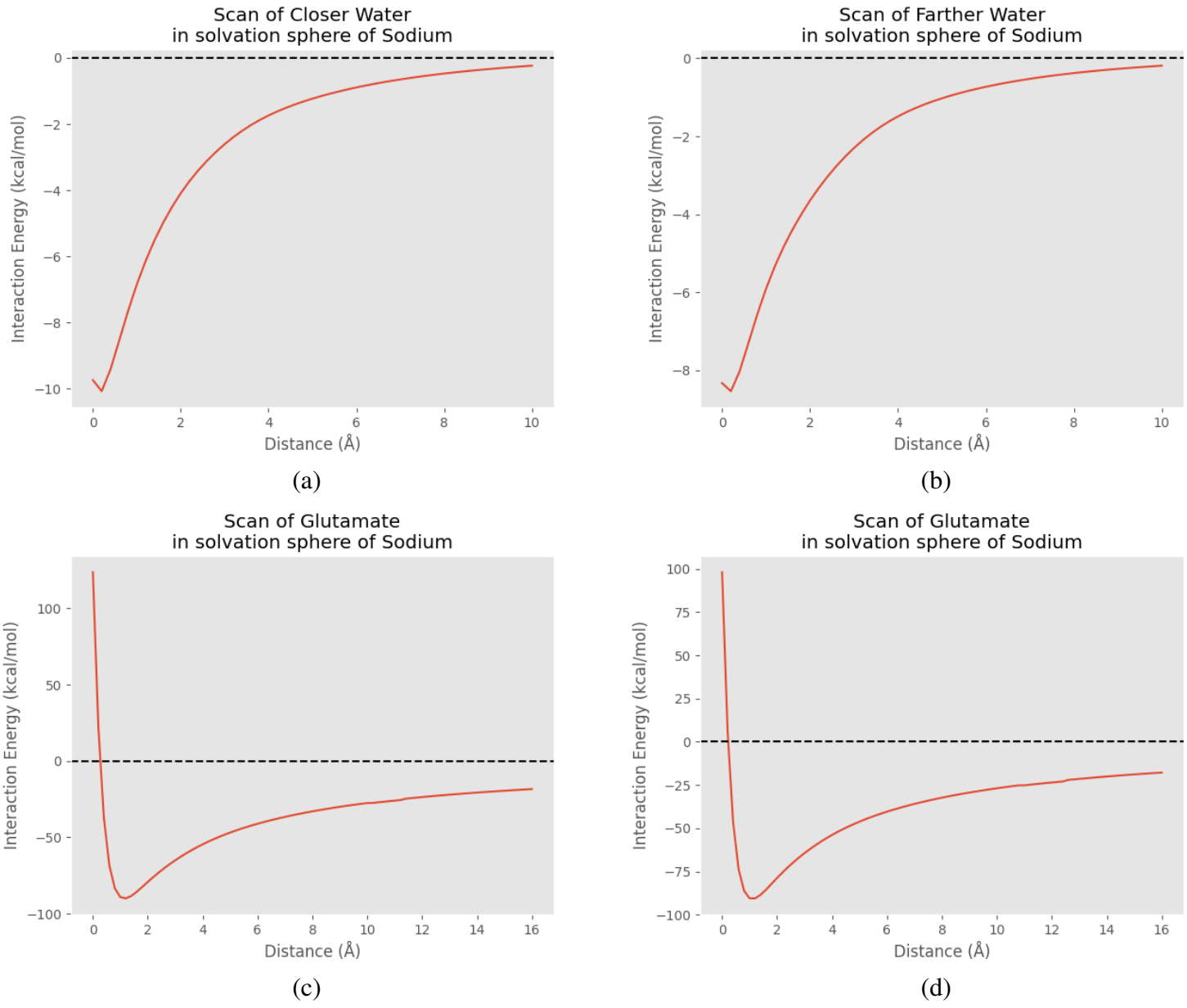


Figure 5.2: Results from scans on Na^+ systems, that are shown in Figure 3.2. The x-axis represents the displacement from the original position, written as 'Distance'.