

# Contents

|   |     |
|---|-----|
| <b>List of Figures</b>  | iii |
| <b>1 Abstract</b>   | 1   |
| <b>2 General introduction</b>   | 3   |
| 2.1 <i>In silico</i> calculations of biological systems . . . . .           | 3   |
| 2.1.1 Molecular dynamics . . . . .  | 4   |
| 2.1.2 Dealing with the potential . . . . .                                  | 6   |
| 2.1.3 Connection with statistical mechanics . . . . .                       | 8   |
| 2.1.4 Root Mean Square Deviation . . . . .                                  | 10  |
| 2.1.5 Root Mean Square Fluctuation . . . . .                                | 11  |
| 2.2 Potassium channels . . . . .  | 11  |
| 2.2.1 Biological introduction . . . . .                                     | 11  |
| <b>3 Analysis of W344R mutation</b>   | 15  |
| 3.1 Introduction . . . . .  | 15  |
| 3.1.1 Biological introduction . . . . .                                     | 15  |
| 3.1.2 Rosetta software . . . . .  | 19  |
| 3.2 Results . . . . .   | 22  |
| 3.3 Conclusions . . . . .   | 30  |
| <b>4 chapter3: <math>\alpha</math>-helix induction in SK2 by Calmodulin</b> | 37  |
| 4.1 Introduction . . . . .  | 37  |
| 4.1.1 Biological introduction . . . . .                                     | 37  |
| 4.1.2 enhanced sampling techniques . . . . .                                | 40  |
| 4.2 results . . . . .   | 45  |
| 4.3 Conclusions . . . . .   | 54  |
| <b>5 Chapter 4: Full channel simulations</b>                                | 57  |
| 5.1 Introduction . . . . .  | 57  |
| 5.2 simulation details and preparation . . . . .                            | 60  |
| 5.3 Riluzole 's pocket characterization . . . . .                           | 62  |
| 5.4 Network Analysis . . . . .  | 65  |
| 5.4.1 Introduction . . . . .  | 65  |

## Contents

---

|  |  |           |
|--|--|-----------|
| 5.4.2  | Definition of the network . . . . .                        | 66        |
| 5.4.3  | Current flow betweenness centrality . . . . .              | 68        |
| 5.5  | Current flow betweenness centrality calculations . . . . . | 69        |
| 5.5.1  | General features . . . . .                                 | 69        |
| 5.5.2  | Kv7.2 Apo . . . . .  | 70        |
| 5.5.3  | drugs . . . . .  | 72        |
| 5.5.4  | Conclusions . . . . .                                      | 76        |
| <b>Appendix A - Mathematical derivations</b> |  | <b>79</b> |
| <b>Appendix B - List of Abbreviations</b>    |  | <b>85</b> |
| <b>Bibliography</b>                          |  | <b>87</b> |

# List of Figures

|     |   |    |
|-----|---|----|
| 2.1 | Different bonded terms appearing in force fields.   | 7  |
| 3.1 | Details on the structure of channel Kv7.2, PDB ID: 7CR3.  | 17 |
| 3.2 | Chart flow with the protocol followed by the rosetta Flex ddG package.  | 32 |
| 3.3 | Experimental current densities versus rosetta $\Delta\Delta G$ of some variants of W344 together with the overlaid visualization of Native (N) and Tilted (T) configurations. | 33 |
| 3.4 | Insights into the binding of Native and Tilted configurations with calmodulin C lobe  | 34 |
| 3.5 | Molecular dynamics output of the calcium responsive domain alone and immerse in water.  | 35 |
| 3.6 | Double mutants simulations. Probability densities of the angles of the W344R with some variants on position 341   | 36 |
| 4.1 | Details on the structure of channel SK2.  | 39 |
| 4.2 | structural details of the SK2 fragment under study, PDB: 1KKD, with the secondary structure obtained from HREX REST2 simulations.   | 45 |
| 4.3 | Building protocol and secondary structure results of the HREX REST2 simulations for SK2 fragment bound to CaM C lobe.   | 48 |
| 4.4 | Depiction of the simulated systems and secondary structure results of the HREX REST2 simulations for the swapped SK2 fragment bound to CaM C lobe.                            | 50 |
| 4.5 | Hypothetical recognition mechanism of SK2 fragment with output of metadynamics simulations.   | 52 |
| 4.6 | secondary structure of some calmoduline targets immerse in water in absence of calmoduline.   | 56 |
| 5.1 | Graphical representation of the binding pocket of the drugs Retigabine and ZTZ240 in the channel Kv7.2.   | 58 |
| 5.2 | Graphical representation of the homology model built for the binding pocket of Riluzole in the channel SK4.   | 59 |
| 5.3 | Graphical representation of the full channel simulations of KCNQ2 channels apo and in presence of Retigabine and ZTZ240   | 60 |

## List of Figures

---

|      |  |    |
|------|--|----|
| 5.4  | Graphical representation of the full channel simulations of KCNN4 channels apo and in presence of Riluzole . . . . .             | 61 |
| 5.5  | Characterization of the binding pocket of Riluzole in SK4 channel. . . . .   | 64 |
| 5.6  | Results obtained from the Network analysis for the $K_V7.2$ apo runs. . . . .  | 71 |
| 5.7  | Results obtained from the Network analysis for the $K_V7.2$ simulations in presence of ZTZ240 drug. . . . .                      | 73 |
| 5.8  | Results of the network analysis of the $K_V7.2$ channel in presence of ZTZ240 for the signal starting in the CaM C lobe. . . . . | 74 |
| 5.9  | Results of the network analysis technique for the $K_V7.2$ channel in presence of Retigabine. . . . .                            | 75 |
| 5.10 | Results of network analysis technique for the SK4 channel in presence of Riluzole. . . . .                                       | 77 |

# **Chapter 1**

## **Abstract**



# Chapter 2

## General introduction

### 2.1 *In silico* calculations of biological systems.

A simulation or an *in silico* calculation can be observed as a computer experiment, consisting in the application of a theoretical model to a target system by an algorithm, being the latter a set of static instructions. The action of the theory through the algorithm to the target system will arise an outcome that should be analyzed similarly to an experiment, which means that some validation will be required, as the theory that we introduce in the simulations will always be built to be as accurate as possible but very aware of our technical limitations, like for example the computational resources available. In the case of biology, typically, simulations will complement the information extracted from other experiments, and there it will display all its potential as they have a very different approach to the events under study, based on the movements and interactions of the atoms constituting the system, whereas experiments usually are indirect, meaning that they collect evidences or side effects related to the events of interest but sometimes missing the atomic mechanism.

Generalizing, two main different frameworks can be stabilised for biological computing simulations. When simulations are done together with experiments they can be used to validate them by seeing part of the target biological process or by computing the same quantities as in the experiments, elucidating the role of each part of the system with atomic resolution meanwhile, for example molecular dynamics simulations (MD) were able to observe some mechanisms and intermediate states of ionic channels [1, 2]. Conversely, if the simulations are done before experiments, they can be used as a first filter to direct the experiments and discard those that are less likely to succeed, one example is drug development, where docking simulations helped to identify new chemotypes relevant to opioid analgesic discoveries [3] or massively tested small ligands to an influenza receptor [4].

Frequently, the starting point of simulations is the structural information of the systems of interest, but for the calculations done in this work, atomic precision is needed and this is very challenging from an experimental perspective. Atomic struc-

tural determination of proteins was led by nuclear magnetic resonance (NMR), but this approach is limited by the fact that the target molecule needs to be crystallized, which can be troublesome [5], although the field is significantly growing [6]. In the last few years, structural determination is experiencing a revolution by recent advances in Cryo Electron Microscopy (Cryo-EM) [7], in a nutshell, one aqueous sample is frozen extremely quickly and its molecules adopt a vitreous phase that does not disrupt macromolecular structures, then, many images of the molecule are collected from different angles to reconstruct the three dimensional structure of the molecule, with this approach inconceivable large structures have been resolved to an outstanding resolution, like the ebola virus [8], eukaryotic ribosomes [9], or membrane proteins like ion channels [10]. Cryo-EM has the disadvantage that only information of stable domains can be obtained, since the molecule necessarily has to adopt the same conformation in most of the collected images in order to obtain a clear image of its three dimensional shape, this makes this structures to only contain static information, leaving the door open to complementary calculations capable of observe the dynamical processes experienced by the biological entities, as it is the case of molecular dynamics simulations (MD) [11]. In this introduction, we will focus in MD simulations, since they comprise most part of the calculations done for this thesis.

The last ingredient that has helped simulations, and in particular MD, to experience a mayor gain in popularity and visibility is the increasing computational power accessible to the research groups, putting aside the development of the so called high-performance computer (HPC) facilities, which is outstanding, it is noteworthy the evolution experienced by the two main MD software used in this thesis, NAMD [12] and GROMACS [13]; both have made enormous efforts in optimizing the resources of computers by allowing parallelization and the use of the graphic processing units (GPUs), so that the load of the calculations are balanced. These advances have been taking place throughout the time period of this thesis and have helped to democratized the technique as interesting simulations can be made by almost any research group with a workstation, in particular, our group has seen how our computational power has increased several times over.

### 2.1.1 Molecular dynamics

So, what is MD? In the following we will introduce the physics behind a molecular dynamics simulation of a biological system, trying to understand in the process the approximations used and the information that can be collected through a simulation. Essentially, MD simulation computes the time evolution of a system, to achieve that we compute Newton equation for each particle  $i$ :

$$\mathbf{F}_i = -\nabla V = m_i \ddot{\mathbf{r}}_i \quad (2.1)$$

In practice, this equation cannot be integrated (solved) as the number of particles in the system is usually enormous and computing the exact potential experienced by the  $i$ -th particle is not trivial at all, we will discuss later the shape of this potential, now we will focus on how to efficiently approximate the solutions of (Eq. 2.1), for

that purpose we will only assume that the potential only depends on the position of the particles constituting our system, since no time dependent perturbations will be introduced. So, for computing the time evolution of the system we can make use of a variant of the most common integration algorithm, which is the velocity Verlet algorithm [14], the main idea is to make a Taylor expansion for the positions at time  $(t + \Delta t)$ :

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\Delta t + 1/2\ddot{\mathbf{r}}(t)\Delta t^2 + O\Delta t^3 \quad (2.2)$$

that provides us with the positions at the following time step  $\mathbf{r}(t + \Delta t)$  with an error proportional to  $\Delta t^3$ , as positions, velocities and accelerations at time  $t$  are known. These new computed positions are used to defined the potential at the following step,  $V(\mathbf{r}(t + \Delta t))$ , since we have assumed that the potential only depends on the positions of the particles of the system, notice that the error in the potential results from the propagation of the error of the positions; the acceleration will be, from the application of (Eq. 2.1):

$$\mathbf{a}(t + \Delta t) = -1/m\nabla V(\mathbf{r}(t + \Delta t)) + O\Delta t^3 \quad (2.3)$$

For clarity, before writing the final expression for the velocities, let's see the taylor expansions of  $\mathbf{v}(t + \Delta t)$  and  $\mathbf{a}(t + \Delta t)$ :

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \mathbf{a}(t)\Delta t + \frac{1}{2}\dot{\mathbf{a}}\Delta t^2 + O\Delta t^3 \quad (2.4)$$

$$\mathbf{a}(t + \Delta t) = \mathbf{a}(t) + \dot{\mathbf{a}}\Delta t + O\Delta t^2 \quad (2.5)$$

separating  $\dot{\mathbf{a}}$  in (Eq. 2.5) and inserting in into Eq. 2.4), we get the expression for the velocities with an error of  $O\Delta t^3$ :

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{a}(t) + \mathbf{a}(t + \Delta t)}{2}\Delta t + O\Delta t^3 \quad (2.6)$$

Where positions and velocities at step  $t$  is assumed to be known. At  $t = 0$  starting velocities need to be assigned, usually from a Boltzmann distribution for a desired temperature. In this derivation of the velocity-Verlet algorithm lies the essence of a MD simulation: velocities and positions are known at time  $t$ , therefore the position dependent potential is also known, with (Eq. 2.2) the positions at the following time step  $t + \Delta t$  are computed, these new positions define the new potential of the next time step, which is used to define the new accelerations with (Eq. 2.3), and the new velocities with (Eq. 2.6), finally, new velocities and accelerations are used to compute the coordinates at  $t + 2\Delta t$  again with (Eq. 2.2), this is repeated until the simulation is completed at time  $t + N\Delta t$ , providing us with a trajectory of the system with a time resolution of  $\Delta t$ . Notice that this simple and fast algorithm has an error associated with the truncation of the Taylor expansion that goes with  $\Delta t^3$ , where the time step is usually small.

### 2.1.2 Dealing with the potential

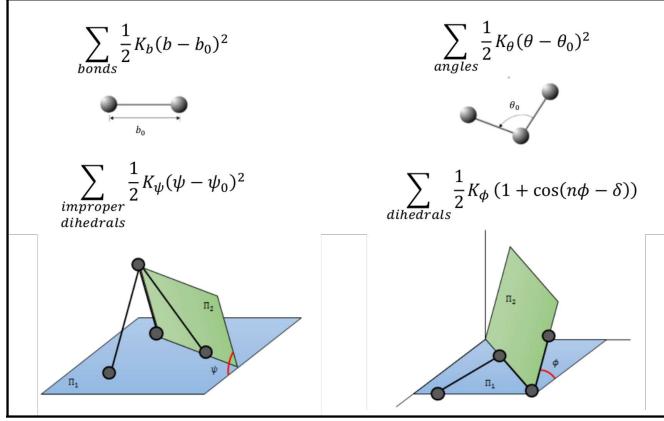
In the previous section we have seen that we have a very powerful algorithm to integrate the equations of motion and generate a trajectory of the system, one crucial question that we left open is how to compute the potential of our system. First of all, we have to determine the resolution of the model, which will establish the accuracy of the outcome of the simulation. For example, if we are modeling a chemical reaction where we want to reproduce the transfer of one electron between two molecules, we will need a quantum model to consider electrons, one of the most popular for that matter is density functional theory [15], but the complexity of the calculations and the addition of the active electrons to the total number of interacting particles substantially increases the computational cost for computing the potential of the system, in fact, in this thesis we will be dealing with biological systems immerse in water, with number of atoms ranging from the thousands to hundred thousands, making this approach impractical; furthermore, to assure numerical stability, the time step for the time evolution has to be 1-2 femtoseconds, in order to be 5-10 smaller than the vibration frequency of the C-H bond, the highest one that can be found in a biological system that explicitly includes all the atoms of the system (all-atom MD).

Therefore we have to develop an efficient way to compute potentials that fit our interests and lies in the sweet spot between accuracy and computational cost, this is usually accomplished by increasing the specificity of the model, as mentioned before we will work with biological systems and we will exploit their characteristics. One key feature is that the processes do not include chemical reactions or the breaking of covalent bonds, therefore a quantum level description of the systems will not be needed as electrons will not play any explicit role. Besides, the systems that we will consider in this thesis (i. e. proteins and cell membranes), are made out of a rather small number of building blocks, in the case of proteins these will be the 20 essential amino acids and for membranes we will use just one lipid, this means that there is a large number (but finite!) of different interactions that can take place in our system, so one intelligent approach is to parametrize all of them and put all these parameters in a list, also known as force field; leaving the prediction ability of the model to the quality of this set of parameters, which are obtained from quantum mechanical simulations and fitted to reproduce experimental properties of proteins or lipids, some of the most popular ones are AMBER [16], OPLS [17] or CHARMM [18], being the latter the one used in all simulations of this thesis.

#### Force fields

Force fields are the mathematical tool used to compute the potential energy for biological systems, they consist in the combination of an additive energy function and a set of parameters. Most force fields share a common energy function so that they are referred to as Class I force fields [19]:

$$V(\mathbf{r}_1 \dots \mathbf{r}_n) = V_B(\mathbf{r}_1 \dots \mathbf{r}_n) + V_{NB}(\mathbf{r}_1 \dots \mathbf{r}_n) \quad (2.7)$$



**Figure 2.1:** Different bonded terms appearing in force fields.

Where the bonded term  $V_B(\mathbf{r}_1 \dots \mathbf{r}_n)$  is the sum of some simple geometrical terms:

$$\sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{improper dih} \frac{1}{2} K_\psi (\psi - \psi_0)^2 + \sum_{dihedrals} \frac{1}{2} K_\phi (1 + \cos(n\phi - \delta)) \quad (2.8)$$

These terms are depicted in (Fig 2.1) and refer to the bond length  $b$ , angle  $\theta$ , torsion angle or improper dihedral  $\phi$ , and out of plane distortion or dihedral  $\psi$ , finally  $K_b, b_0, K_\theta, \theta_0, K_\psi, \psi_0, K_\phi, n$  and  $\delta$  are parameters stored in the force field, this set of parameters takes into account the specificity of the interacting atoms, considering not only the different nature of those, this is if the particle is an carbon or an oxygen, but also the environment surrounding it, for instance differentiating carbon atoms belonging to protein backbone from others in the side chains.

The non bonded term  $V_{NB}(\mathbf{r}_1 \dots \mathbf{r}_n)$  is the sum of a Lennard-Jones type potential that accounts for the Van der Waals interactions and a culombic term for the electrostatic ones:

$$V_{NB}(\mathbf{r}_1 \dots \mathbf{r}_n) = \sum_{pairs(i,j)} \left[ \varepsilon_{ij} \left( \frac{\sigma_{ij}^{12}}{r_{i,j}^{12}} - \frac{\sigma_{ij}^6}{r_{i,j}^6} \right) + \frac{q_i q_j}{4\pi \varepsilon r_{ij}} \right] \quad (2.9)$$

where  $\varepsilon_{ij}$  and  $\sigma_{ij}$  are the well depth and the minimum energy distance of the Lennard-Jones potential, respectively, and  $q_i$  and  $q_j$  are the charges of the atoms of the pair and  $\varepsilon$  the dielectric constant. Usually, force fields store  $\varepsilon_i$  and  $\sigma_i$  for each atom type, and for each interacting pair of atoms  $i$  and  $j$ ,  $\varepsilon_{ij}$  and  $\sigma_{ij}$  are obtained by applying the so-called Lorentz-berthelot combining rules,  $\sigma_{ij} = 1/2(\sigma_i + \sigma_j)$  and  $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ .

Assuming that the shape of the energy terms in (Eqs. 2.8 and 2.9) and all possible parameter are stored in the force field, it is left for the software to know which atoms are interacting and the subsequent parameters needed in each time step, for

that matter, MD software follows two different approaches for the bonded and nonbonded interactions. For the bonded terms, the bonded interactions that will take place throughout the simulation have to be provided to the software before the simulation starts and are usually referred to as the topology of the system; since we are not allowing bonds to break, the terms referring to angles, torsions, etc. will hold throughout the simulation, even though their value will depend on the positions of the particles and will be different at each time step, this is justified as we know that proteins or lipids form covalent bonds that are energetically very costly to break. Hence, it is mandatory to generate the topology of the system with the information of the number of bonded terms and the atom types forming these terms, determining the number of terms appearing in the sums of (Eq. 2.8) and the parameters needed for each one, so that the software only has to introduce the value of the positions at each time in (Eq. 2.8) to compute the bonded terms of the potential.

For the nonbonded interactions, in principle, all atoms interact with each other, making the number of interactions scale with  $N(N - 1)$ , being  $N$  the total number of atoms (and usually a big number), to speed up the calculations, nonbonded interactions are further differentiated into short ranged and long ranged by a cutoff. For each atom, to establish the atoms that are within the cutoff for and avoiding computing the distance to the rest at each time step, atoms separated by a distance slightly larger than the cutoff will be stored in a list that is updated with a frequency of 10-20 steps, so only distance differences of the atoms in this list are needed for subsequent computation of the interaction potential with (Eq. 2.9). For the long range interactions, the potential is computed by means of the particle mesh Ewald summation, that scales with  $N \log(N)$  since its foundation lies in performing a Fourier transform to compute the potential through the Poisson equation in the reciprocal space of the positions.

Summarizing, we can sketch how a molecular dynamics software works, coordinates and atom type of each particle are introduced in the force field to define the potential; this potential is used to compute the coordinates and velocities of the system at the following time step in a Verlet like algorithm that moves the particles defining the new positions and therefore the new potential repeating it for all the MD simulation.

This introduction of molecular dynamics is a simplified version of what actually happens inside a simulation, many subtleties followed by different softwares to reduce the computational cost and more information on the possible options allowed by MD software can be found, for example, in the [GROMACS manual](#) [20] or the [NAMD's user guide](#) [21].

### 2.1.3 Connection with statistical mechanics

Now that we have the time evolution of a system and the MD simulation is completed, the next step is to post process the simulation and extract information out of it, which is more difficult than expected, as in most cases it is not possible to derive conclusions just from the visualization of the obtained trajectories. On the one hand, MD can be logically cumbersome as a trajectory generates a lot of data, namely positions and velocities for all particles, and this information has to be properly managed to derive other properties of the system. On the other hand, a theoretical background in statis-

tical physics is required, as it is the mathematical tool needed to extract mean values of the target quantities out of the time evolution of the system, since the equations of the velocity Verlet algorithm have an associated error proportional to  $\Delta t^3$ , which means that as the simulation goes on the error on the magnitudes computed also grow, making the outcome of the simulations not deterministic (or chaotic). We will introduce now some basic concepts of statistical mechanics that will help us understand the mathematics behind the calculation of any property and the inherent approximations lying underneath.

For each system, a phase space,  $\Gamma$ , can be defined, and it is formed by all plausible combinations of positions  $q_i$  and momenta  $p_i$  that a system of  $N$  particles can have:

$$\Gamma \equiv \{q_i, p_i\} \quad i = 1 \dots 3N \quad (2.10)$$

A state is defined as a point of  $\Gamma$ , this states allocates one value for each coordinate and momenta, and the ensemble of all states constitutes the phase space. We assume that  $\Gamma$  will be static in time, this is that no external forces or time dependent perturbations will be introduced. Each state will have associated a probability of materializing defined by the density probability function  $\rho(q, p)$  in the phase space:

$$\rho(\{q_i\}, \{p_i\}) \equiv \rho(q, p) \quad (2.11)$$

so that:

$$\rho(\{q_i\}, \{p_i\}) dq_1 \dots dq_{3N} dp_1 \dots dp_{3N} \equiv \rho(q, p) dq dp \quad (2.12)$$

Is the probability that the system lies in an infinitesimal volume of phase space. In biology, a smaller space than  $\Gamma$  is usually defined as there are some states that will not be physically accessible and we usually do not include them in our simulations, an example of those states can be the ones that have covalent bond breaks, mathematically we can say that we set the probability of these states to zero; this new smaller space is usually called configurational space.

The combination of all points in the phase space plus the probability function contains all the information of the system and it is called ensemble. If our system is at thermal equilibrium, this is that has a constant number of particles, volume and temperature ( $N, V, T$ ), it is said to be in the canonical ensemble and its probability function is defined as  $\rho = e^{-H(\{q_i\}, \{p_i\})/k_B T}$ , where  $k_B$  the Boltzmann constant and  $T$  the temperature and  $H(\{q_i\}, \{p_i\}) \equiv H(q, p)$  is the Hamiltonian of the system  $H(q, p) = K(q, p) + V(q, p)$ , which is the sum of the kinetic term  $K(q, p)$  and  $V(q, p)$ . With this definitions we can write the first postulate of statistical physics, the mean value of any magnitude that depends on the coordinates and momenta is defined as:

$$\langle A \rangle = \frac{1}{Z} \int \rho(q, p) A(q, p) dq dp \quad (2.13)$$

Where we have normalized the value by dividing it by the partition function ( $Z$ ), which is the integral of the probability density throughout all phase space. The idea

behind (Eq. 2.13) is that in order to know the mean value of a quantity, we need to compute it at each point of the phase space and weight it with the probability associated to the appearance of that state, as it must be considered that there are high energy states that will not contribute to the average, this integral is in general not feasible, since the number of elements in the ensemble is supposed to be big and the number of integration variables explodes with the number of particles. However, MD simulations follow another approach to evaluate the mean values, this is, if we let the system evolve a sufficiently large time it will be able to explore all the phase space yielding the time average:

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(q(t), p(t)) dt \quad (2.14)$$

furthermore, under these circumstances the probability of a state  $P(q, p)$  will be the fraction of time that spends in that state, meaning that even the free energy  $F$  can be computed with this approach:

$$\rho(q, p) dp dq = P(q, p) dp dq \Rightarrow F = -k_B T \ln(P) \quad (2.15)$$

And projecting the free energies on some coordinates can lead to free energy profiles that can describe the energetic barriers separating some events of interest. Finally, a system is said to be ergodic if it satisfies that the time average is equal to the mean value:

$$\langle A \rangle = \bar{A} \quad (2.16)$$

As we are not able to compute ensemble averages using (Eq 2.13), in MD simulations we have to assume that the system is ergodic and the time average will arise reliable results. This hypothesis lies under all molecular dynamics simulations and it is very tricky to deal with because it is not possible to know before hand if this assumption is acceptable, this is the reason because molecular dynamics software's manuals will not specify how long you have to run your simulation in order to estimate physical quantities, it will depend on the size of the portion of interest of the configurational space and the time scale needed by the system to explore it. So what can we do to validate the simulations? In general, there are a few rules that molecular dynamics works follow, like doing different replicas of the same simulation, or checking if the computed magnitude is robust over the simulation time, but what usually works best is having some experimental data that somehow validates your results.

To conclude this part of the introduction, we will define a couple of magnitudes that will be computed repeatedly in almost all simulations of this thesis, as they serve as indicators of what is actually happening in a simulation without the need of carefully visualizing all frames, which can be really cumbersome in some situations, also they will also serve as examples of the previously commented statistical quantities.

## 2.1.4 Root Mean Square Deviation

The first indicator is called Root Mean Square Deviation (RMSD), and consist in computing the average deviation of the coordinates of a set of atoms  $\mathbf{r}_i(t)$  at each time

$t$  with respect to a reference structure  $\mathbf{r}_i(t_0)$ :

$$RMSD(t) = \sqrt{\frac{1}{N} \sum_i^N (\mathbf{r}_i(t) - \mathbf{r}_i(t_0))^2} \quad (2.17)$$

So that average deviation of the structure throughout the simulation time can be plotted, usually, solid rigid rotations and translations are subtracted so that if our RMSD has steep increases or major changes through time will mean that our system is changing and something is happening, however a plateau in general will mean that the simulation is stable through time as no conformational changes are taking place.

### 2.1.5 Root Mean Square Fluctuation

The root mean square fluctuation (RMSF) is the time average of the RMSD, so it yields a number per set of atoms, instead of a time series. In this thesis, this magnitude is usually presented as a plot of its value for each amino acid of the sequence, so it can be observed that amino acids belonging to  $\alpha$ -helices have less mobility than those belonging to non structured domains. Notice that again solid state movements are also subtracted from the RMSF.

## 2.2 Potassium channels

### 2.2.1 Biological introduction

This work will focus on potassium channels belonging to the families KCNN and KCNQ, so in the following they will be briefly introduced from a biological perspective, to characterize them, or at least to have a perspective of the biology involved, we will try to answer the following questions: How are they synthesized?, where are they located? and what is their function?. But before we do that we will make a brief introduction of proteins.

Proteins are polypeptide chains, this means that they are formed by several amino acids joined by peptide bonds; these are formed when the carboxyl group of one amino acid covalently bonds to the amino group of the following amino acid, losing in the process a water molecule. The chain is elongated repeating this process until the whole protein is formed. As a consequence, the first amino acid keeps its amino group intact and the last one its carboxyl group, therefore the chain is said to extend from its amino terminus (N terminus) to its carboxyl terminus (C terminus). This polypeptide chain is usually referred to as the sum of the backbone and the side chains, being the backbone the succession of the  $NH - C_\alpha H - C' = O$ , where the  $NH$  is the remain of the amino group and  $H - C' = O$  of the acid one after the peptide bond formation, and the  $C_\alpha$  is the alpha carbon of the amino acid, to this carbon is attached the side chains of the amino acids, the 20 amino acids that appear in proteins (also known as canonical) are distinguished by its side chain.

## Potassium channels

---

The structure of proteins can be very intricate and it is usually separated in four terms, the primary structure takes into account the sequence of the protein, the secondary structure considers the local geometric arrangement of the protein, which can be ordered (alpha helices or beta sheets) or disordered (loops, turns or bend), both alpha helices and beta sheets are stabilized by hydrogen bonds. The tertiary structure refers to the assembly of several secondary structures to form domains, which are regions that are able to perform chemical or physical tasks. Finally, Quaternary structure considers the assembly of different tertiary domains [22].

### Synthesis and folding

Protein synthesis is a core process in each living cell as it consists in the creation of proteins, the biological entities that perform all the tasks needed for the proper functioning of the cells. The whole synthesis process is very intricate and here we will only introduce the basics to follow the results obtained in this thesis. The information required for the synthesis lies in the genes and it is stored in a polymer called deoxyribonucleic acid (DNA), all information is stored in a four letter code, made by four nucleotides abbreviated by A, T, C and G. To carry the information stored in the DNA to the ribosomes, the macromolecular machines that perform the synthesis, a process called transcription is followed, it consist in replicate the information stored in the double-stranded DNA to a single stranded ribonucleic acid (RNA), which travels through the cell until it is eventually found by the ribosome. Inside the ribosome the next step of protein synthesis takes place, called translation, it is the process in which the ribosome reads the information carried by the RNA to assemble a sequence of amino acids that will become a protein [23]. While the translation is taking place, the resulting nascent peptide chain is exiting the ribosomal exit tunnel into the cytosol within the cell, the addition of amino acids to the nascent chain is a slow process considering that it lies in the nanoscale, the synthesis rate is around 6 amino acids per second in eukaryotic cells and around 20 in *E.coli* bacteria [24].

The last step of the synthesis is the folding of the nascent chain into the wild type structure, this process start in the ribosomal exit tunnel and takes place vectorially from the N to the C terminus, it has been showed that the appearance of small  $\alpha$ -helices inside the last part of the tunnel, also known as vestibule, is allowed, while the formation of tertiary structures is prevented; these peculiarities of the ribosome in the first steps of the folding process of a protein might be of critical importance [25].

### Ionic channels are membrane proteins

All living cells are enveloped by a membrane, its main purpose is to prevent diffusion between the exterior environment and cytoplasm, as both environments have an aqueous nature, this is achieved by constituting a hydrophobic area, usually by the assemble of many lipids into a bilayer. Each lipid consists in a polar phosphate group from which two hydrophobic chains come out, so that the hydrophilic heads face the water and the chains of the two layers constitute the hydrophobic surface. As a proof of the importance of membrane proteins in cells, it has been stated that they comprise

nearly 30% of all proteins in eukaryotic cells [6]. Membrane proteins can be classified in two main groups: peripheral and integral; peripheral proteins are bound to the membrane indirectly by integral proteins or directly by interactions with the polar heads of the phospholipids, on the other hand, integral proteins have one or more segments embedded in the membrane, a subset of these are the so called transmembrane proteins, which have domains that span the membrane and have amino acids both in the extracellular and intracellular aqueous regions, as it is the case of ion channels [23].

Therefore, following its synthesis, ion channels have to travel to the membrane where it has to be located in a very particular conformation so that it forms a pore, the channels studied in this thesis belong to the KCNQ and KCNN families, which form the pore by the resulting of the tetramerization of four identical subunits, so before even addressing the functioning of ion channels it is outstanding the complexity of the process in which they are synthesized and inserted into the membrane.

### Function of ionic channels

The cytosol or intracellular fluid is mainly composed of potassium ions and aminoacids and also to a lesser extent, by sodium and chloride ions solvated in water. On the other hand, the extracellular media has a high concentration of sodium and chloride ions.

The difference in ion concentration has, as a result, the appearance of a difference in electrical potential between the two regions. Almost all cell membranes have an electrical potential across them, being the inside usually more negative than the outside, this potential has usually two basic functions, the first one is to provide power to the cell allowing it to behave as a battery, the second one is to allow some cells to transmit signals, like neurons or muscle cells.

When a cell is driven out of its relaxed state, it has two mechanisms to restore its resting potential, ion channels and ion pumps, the main difference between them is that ion pumps use energy of the cell whereas ion channels allow diffusion in favour of the ion concentration gradients. In general, ion pores are opened or closed in response to some event, this is referred to as gating, both KCNQ and KCNN channels are sensible to intracellular concentration of  $\text{Ca}^{2+}$  ions, which are recognized by **calmodulin** (CaM), a protein known to modulate calcium in several signaling events and that is bound to the channels. Upon gating, these channels will only allow potassium atoms, as they are very selective with the ion type.

Regarding the purpose of KCNN and KCNQ channels cells, they play a role in restoring the resting potential after a polarization event has been fired, channels of both families can be found in neurons and some KCNQ channels also in cardiac tissues, therefore they have a relevant clinical significance as their malfunction can lead to encephalopathies or arrhythmias when the neurons or cardiac tissues need longer times or even fail to reach their resting conditions, a more detailed description of the structures of the channels under study will be provided later on.

## Potassium channels

---

# Chapter 3

## Analysis of W344R mutation

### 3.1 Introduction

This chapter refers to a collaboration that led to this article [26], together with subsequent work done in the same research line both to better understand the published calculations and to continue exploring the hypothesis that will be presented here. The results of this chapter consist in molecular dynamics simulations (MD), which have been introduced in the previous chapter, and some calculations using a bioinformatics software called rosetta (<https://www.rosettacommons.org/>). To begin with, we will make a biological introduction of the questions addressed in this work and also a small introduction of this software, afterwards we will show and analyze the results obtained.

#### 3.1.1 Biological introduction

The potassium channel studied in this work is the Kv7.2, encoded in the KCNQ2 gene, it is mainly expressed in the nervous system and it is responsible for the so called M-current, which is critical for restoring the neuronal resting potential. They react to voltage changes in the cytosol by means of  $\text{Ca}^{2+}$  concentration, which is regulated by the CALcium MODulating proteIN or calmodulin (CaM) [27], a small protein consisting in two globular domains (N and C lobe) that are able to carry up to two calcium ions each. Variants on this gene that suppress its function are related to several medical conditions, like epileptic encephalopathy [28] or Benign Familial Neonatal Epilepsy [29].

Channel Kv7.2 is a tetramer, which means that is the result of the assembly of four identical subunits (Fig. 3.1 A, B). Each subunit consists of a transmembrane domain, spanning helices S1 to S6 and a intracellular C Terminal domain (CTD), formed by helices hA, hB, hC and hD (Fig. 3.1 D, hD is not showed), inside the transmembrane domain there are two distinguished regions according to their functions; helices S1 to S4 comprise the so-called voltage sensor domain (VSD) (Fig. 3.1

## Introduction

---

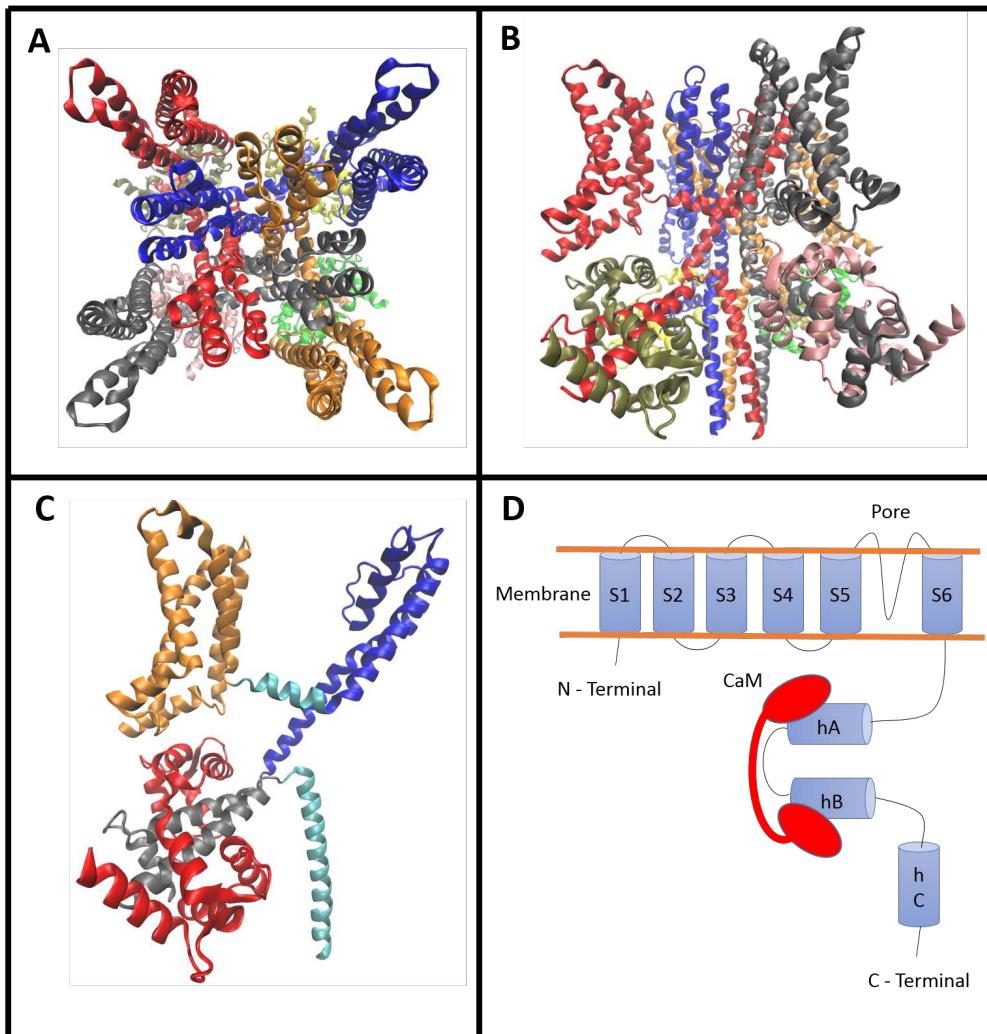
C, orange area), as there are four arginine residues <sup>1</sup> in helix S4 that are capable of sensing voltage differences, when triggered, the VSD will induce a structural change that will result in the opening of the pore. On the other hand, helices S5 and S6 constitutes the pore domain (Fig. 3.1 C, dark blue area), which contains a GYG signature (glycine-tirosine-glycine) that makes the channel very selective to potassium ions; in the intracellular CTD stands out the calcium responsive domain (CRD), an antiparallel hairpin formed by helices hA and hB (Fig. 3.1 C, grey area), which are recognized by CaM (Fig. 3.1 C, red area) providing calcium sensitivity to the channel, it is also known as the calmodulin binding domain (CaMBD). Finally both helices hC and hD allow tetramerization. Helix hD and a large intrinsic disordered domain located between helices hA and hB do not appear in the Cryo-EM structure, due to their mobility, so they are not showed in (Fig. 3.1).

This chapter is dedicated mainly to the study of the mutation W344R of the channel KV7.2, let's review now the features that make this particular variant interesting. To begin with, it has been found in patients with hereditary epilepsy [32, 33] providing clinical significance to this variant, also, its location lies in the hA helix of the CRD, and more precisely, in the so-called IQ calmodulin binding motif, a sequence of amino acids that appears often in CaM targets [34], the sequence of this motif is IQxxΘR, where x varies from one target to other and Θ is an hydrophobic residue, as a matter of fact, W344 would be the one named Θ previously, making this location tightly related to CaM binding and subsequently with the function of the channel. One of the most shocking results regarding W344R variant was that it abolished the function of the channel without affecting the binding to CaM in the process [32], this was remarkable as other mutations in the hA helix showed a nice correlation between channel function and CaM binding [35]; while the behavior of most variants suggested that the function of hA would be to serve as anchor allowing the binding to CaM, the discrepancy in W344R mutation raised suspicions regarding the purpose of hA in the channel, could it be that part of its functioning was missing?. In other words, what could be so deleterious to the channel function happening in hA that did not involve calmodulin? this question was the biological starting point of this work. All *in silico* calculations that will be presented in this chapter were made with the purpose of provide insight in the mechanisms happening upon W344R mutation at the atomic level, indeed, we were able to formulate an hypothesis that could explain the experimental results obtained, but before that let's review the key experimental evidences that were presented together with the computational results so that the reader can have a global perspective on this work, more details on the experiments can be found in [26].

Channel function upon W344R mutation was evaluated and found to be non functional, which was in agreement with previous results [32], this was carried out by electrophysiology experiments, in sort, bacterial cells are genetically modified to express the channel with the W344R mutation, afterwards the resulting membrane crowded with the channels is modeled as a capacitor so that membrane potential and current

---

<sup>1</sup>Arginines are amino acids which have a long hydrophobic tail with an extra negative charge in its edge.



**Figure 3.1:** Details on the structure of channel Kv7.2, made with VMD software [30], from the experimental structure PDB ID: 7CR3 [31]. **A** and **B**, top and side view of the channel respectively, where each monomer is colored differently. **C**, detail of one monomer where the voltage sensor domain or helices S1 to S4 (VSD) are colored in orange, the pore domain or helices S5 - pore helix - S6 are in dark blue; calcium responsive domain (CRD) or helices hA - hTW - hB are in grey and calmodulin is in red, linker S4-S5 and helix hC are left in cyan. **D**, cartoon representation of the structure of the channel, for clarity only main helices are depicted, smaller ones like linker S4-S5, pore helix (between S5 and S6) and hTW (between hA and hB) are not illustrated.

densities of membranes can be measured.

## Introduction

---

Furthermore, CRD experienced a mayor CaM binding reduction upon W344R mutation, this was showed by means of *in cellulo* Förster resonance energy transfer (FRET) experiments, the main idea is that cells are genetically modified to express the channel with two fluorophores linked to them, one of them will be the donor as it emits an electromagnetic radiation (blue light) and the other is the acceptor, since it also emits electromagnetic radiation (yellow light) but only if receives the one transmitted by the donor, so the measurement of acceptor light is a sensor that indicates that the two fluorophores are closed in space, these are big barrel-like structures (PDB IDs: 5OX8 [36], 3ED8 [37]), so in order to avoid folding distortions they have to be attached at the N terminal or C terminal of the constructs; therefore, if we carefully design our system so that the acceptor is attached to the CaM and the donor to a construct of helices hA and hB, the signals obtained will be related to the CaM binding as the acceptor will only emit light if the donor is sufficiently close, considering that CaM-hAB WT construct is stable<sup>2</sup> (PDB ID: 6FEG [38]); the output of these experiments was a FRET reduction when introducing the mutation.

Therefore, it was obtained that W344R mutation makes the channel non functional and disrupts its binding with CaM, in contrast to previous work, which showed that in *in vitro* experiments this mutation does not disturb the CaM-CRD binding [32], the functional characterization is also done *in cellulo*, so experiments stated that the mutation disrupts CaM binding (and subsequent channel function) *in vivo* but does not *in vitro*.

One important difference between these kind of experiments is how the folding occurs, in *in cellulo* or *in vivo* experiments, the whole channel or the designed construct is synthesized by cells, therefore the folding takes place vectorially and it has the physical restraint of the ribosomal tunnel, in *in vitro* experiments however, the construct is expressed in bacteria, but to recover it these cells need to be eliminated and the construct has to be solubilized in urea, a chaotropic agent, therefore when urea is replaced by water the folding and CaM recognition is done at once without the constraints imposed by the ribosome. The results suggest that the folding pathway followed by the channel in the two experiments are different, but to test that hypothesis we have to be sure that the ribosome has an effect on the folding, which would happen if the folding starts during translation, this is, in the ribosomal vestibule. To directly asses the effect of the mutation in the cotranslational folding, FRET experiments combined with SecM arresting peptide were carried out, SecM is a peptide sequence that is known to interact with the ribosome, detaining the protein synthesis unless some external force acts on it, pulling it out and allowing the restoration of the synthesis [39, 40], this force can be caused by the cotranslational protein folding [39]. Therefore, SecM peptide can act as a force sensor since the folding only occurs when it undergoes an external force, presumably exerted by the folding of the nascent chain and steric clashes with the ribosome, so a FRET experiment can be designed with a construct of the CRD that is known to fold, the sequence used was [FRET\_donor-hA-hTW-hB-linker-SecM-FRET\_acceptor], the linker length between the CRD and the SecM that will work best is not known beforehand so different lengths have to be tested. It is

---

<sup>2</sup>It was also checked that the emitted light came from CaM-CRD like structures and it was not spurious signal from protein aggregation.

believed that the CRD will start to fold in the vestibule of the ribosomal tunnel, which is around 100 Å away from the peptidyl transferase center; if the linker is too short the CRD will not be able to reach the vestibule and cotranslational folding can not take place, conversely, if the peptide is too long the CRD will be completely outside the ribosome and the force felt by the arresting peptide will be lower, there is a sweet spot then in which the experienced forced is maximal. The linker length used varied from 18 amino acids to 50, for each linker length the quantity of proteins folded was measured by collecting the FRET signal, finding that 22 amino acids was the most efficient one. Finally, the results showed that the WT construct in presence of CaM presented a peak, conversely, mutant constructs or without CaM did not showed any peak, indicating that cotranslational folding takes place, but W344R mutation avoids it. Other interesting conclusion that can be extracted from these results is that CaM is needed for the formation of the hA hB hairpin of the CRD, since in absence of CaM for the native sequence the peak was also not present.

Summing up, the experiments carried out indicated that the effect of W344R mutation was to disturb the cotranslational folding preventing the binding to CaM and leading to a non functional channel, while presenting a wild type (WT) like behavior in *in vitro* experiments. This evidences disagree with the usual conception of protein folding in which the potential landscape of a protein is a funnel like structure with the folded (native) state at its bottom [41], since it looks like something happens (presumably) cotranslationally that impedes the variant to acquire its lower state conformation, leaving the mutated channel in a non productive conformation that is related ultimately to pathologies.

### 3.1.2 Rosetta software

In parallel to the previously introduced experiments, computational effort was made to try to rationalize the results from an atomic perspective. Some of the results were obtained with the Rosetta software, a suite for macromolecular modeling and protein structure analysis. The scope of the software is wide, evidenced by the variety of applications of the suite, ranging from the prediction of protein structures with the rosetta server (<http://new.rosetta.org/>) [42] to the modeling of several systems, like protein-protein [43, 44], nucleic acid-protein [45], antibodies-immune system proteins[46], membrane proteins [47] or designing new proteins [48], among other applications.

The approach followed by this software differs from the molecular dynamics one, ruled by computing the time evolution of the system; instead, it focusses the effort in evaluating the ensemble average (Eq. 2.13), as stated before, this integral is not feasible, what is done in practice is to sample the conformational space by a predefined rosetta protocol to evaluate the physical plausibility of the sampled structures. Generally, a typical rosetta protocol can be summarize as follows [49]: the input is a biomolecule in a specific conformation (or pose), that will be transformed by the application of a mover, yielding an output structure, movers can be deterministic or based in stochastic Monte Carlo methods and usually ruled by the Metropolis criterion.

## Introduction

---

The Metropolis criterion is an algorithm that computes acceptance probability  $P$  of the new structure by energetical comparison of both new and original structures, if the energy of the new one is lower it automatically accepts it and assigns a probability otherwise [50]:

$$P = \begin{cases} 1, & \text{if } E_{\text{new}} < E_{\text{orig}} \\ e^{-(E_{\text{new}} - E_{\text{orig}})/k_B T}, & \text{otherwise.} \end{cases} \quad (3.1)$$

The variety of movers or alterations that can be applied and the effort made in improving the accuracy of the score functions used to estimate the energies of the conformations [51] makes the software as one of the most used in bioinformatics calculations.

In this chapter of the thesis we used the so-called FlexddG package [52], with the aim of computing the binding affinities of some mutations in the channel. In this case the protocol used by rosetta to sample the conformational space is complex (Fig 3.2), it starts by minimizing the input structure (step 1), then it performs the backrub protocol (step 2), which consists in introducing displacements randomly in the backbone angles in a shell of 8 Å around the mutation site, where this displacements are accepted or rejected based on the Metropolis criterion with an energy of  $1.2 k_B T$  in (Eq. 3.1) for improve the sampling, this will generate a sampling ensemble of 50,000 backrub moves. Afterwards the protocol is separated for the mutant and WT; for the WT, side chains of residues that were previously moved are optimized with the packing protocol, this is done for each structure of the ensemble obtained with the backrub protocol (step 3, pack), for the mutant, before optimizing the structures, the mutation is carried out (step 3, mutate and pack). For each structure of both WT and mutant optimized ensembles, the free energy difference of the bound complex and unbound one is computed,  $\Delta G = G_{\text{complex}} - G_{\text{unbound}}$  by means of the Rosetta energy function 2015 (REF2015) [53], and finally the comparison of the free energy differences of wild type and mutant yield the binding affinity  $\Delta\Delta G = \Delta G_{\text{Mutation}} - \Delta G_{\text{WT}}$  (step 5), where the final value of the binding affinity will be the average of all the 50,000 structure ensemble, the whole protocol is repeated 50 times for improve the sampling.

In the previously explained protocol, side chain optimization is performed with the packing protocol, the idea here is that rosetta has a list of discrete possible side chain orientations for each amino acid, known as rotamers, and the packing is the searching of the combination of rotamers of the different amino acids in the sequence that gives the lowest-energy solution, this sampling of rotamer combinations is also done stochastically. Summarizing, FlexddG protocol provides a sampling of optimized structures that have a variety of small changes in the surrounding area of both backbone and side chains of the mutated site, so that the mutated residue can fit in a plausible conformation for the posterior score.

This protocol successfully predicted most of the loss of function variants of the KCNQ1 channel [54] and partially correlate pathogenic mutations with a destabilizing value of  $\Delta\Delta G$  for the KCNQ2 channel [55], in both cases the structure of the channels

were modeled and they used previous versions of the score function.

After the rosetta FlexddG calculations have been introduced, let's make a few comments on the meaning of the computed final magnitude,  $\Delta\Delta G$ . It accounts for the impact of the mutation on the CaM-CRD binding but also on the nearby residues of the hA of the CRD, as it is built upon the comparison of  $\Delta G$  of the wild type and mutant, and each  $\Delta G$  has information of the energetics of the bound system as well as the unbounded one.

A positive value of  $\Delta\Delta G$  means that the mutant shows a weaker binding affinity than the wild type and the mutation is considered destabilizing, conversely when the  $\Delta\Delta G$  yields a negative value the mutation is considered stabilizing since its binding affinity is bigger than the one of the wild type.

This computed binding affinity has to be consider as a qualitative estimation of the real one for a variety of reasons, the first one is that it has been found that the error associated with this technique is around 0.5 kcal/mol [56], therefore if the absolute value of  $\Delta\Delta G$  is below this quantity we will say that the mutation is neutral, and it has only been after the score function of 2015 that rosetta results have been able to compare to energy units of kcal/mol, since rosetta score functions are mixes of physical and knowledge-based terms<sup>3</sup> [53, 57], the other source of error that must be considered is that we are making a quite strong assumption when making this calculations, it consists in assuming that the mutation will not affect the folding or the posterior assembly of the system, in this case the calmodulin binding domain of the channel and CaM, posterior tetramerization or membrane trafficking, as we are forcing the mutation to fit in an ensemble of wild type like conformations, allowing only a small changes in the nearby area, even though mutations located in critical locations could affect one of the many steps needed for a protein to adopt its final wild type like conformation. Nonetheless, this calculations provide interesting insight of the mutation as they give an ensemble of possible mutant conformations that we know that could appear at least in the *in vitro* experiments, in which CaM is known to recognize W344R variant.

---

<sup>3</sup>These terms penalize situations that are statistically infrequent, for example the probability of each amino acid to have certain torsion and improper dihedrals [51].

## Results

---

### 3.2 Results

In this section, we will present the results appearing in reference [26] together with parallel work that keeps exploring the relevance and role of the position 344 in the cotranslational folding. To begin with, in order to asses the stability of the CaM-CRD complex using a computational approach we calculated  $\Delta\Delta G$  using the package Flex ddG of rosetta, the motivation of making these calculations was twofold, first of all it would provide us with an estimation of the binding affinity of the mutation of interest, but also with an ensemble of structures of both wild type and mutant. The computed value of the  $\Delta\Delta G$  for W344R mutation was 5.1 kcal/mol, which is a typical value of a destabilizing mutation [52, 53], suggesting that the binding to CaM should not take place. To evaluate the capability of rosetta to make sensible predictions we evaluate other variants in position 344 whose impact on the channel function had been previously evaluated, namely W344E, W344F, W344G and W344Q, exhibiting a marked correlation between a high  $\Delta\Delta G$  and the loosing of channel function, results are depicted in Fig (3.3, A). As stated before, previous *in vitro* experiments showed that the mutation is capable of binding to CaM with a slightly better affinity than the WT [32], in contrast with the destabilizing output of rosetta's  $\Delta\Delta G$ , this made us think that there is a stable bound conformation for W344R variant that the conformational sampling of the rosetta protocol was failing to reach or that was being underestimated by the rosetta energy function. In its [github page of materials and methods](#) all python scripts needed for launching the calculations are provided together with the input and outputs of rosetta [58], the ensemble of structures is not provided due to storage limitations on github.

When visualizing the structures provided by rosetta, we saw that there were some populations (1.6% for WT and 3.4% for W344R) in which the side chain became tilted (T) towards the rest of the hA helix, instead of targeting the CaM C lobe, as it appears in the native (N) WT structure (Fig. 3.3, B and C), the differentiating characteristic between these two conformations is the angles as they are depicted in (Fig. 3.3, B and C). As the binding affinity is the result of the average of the ensemble generated by rosetta, this difference in the conformational ensemble obtained could be important.

To have a better understanding of the energetic interactions of N and T conformations of W344R mutation with CaM, three all atom MD simulations were set up, the systems were built from the rosetta output by only keeping the helix hA of the CRD and the C lobe of the calmoduline, a wild type like configuration was built (Fig. 3.4, A) to be compared with the N and T configuration of the W344R mutation (W344R-N and W344R-T), as it is showed in (Fig. 3.4, B). Afterwards, the systems were solvated with TIP3P water [59] in a cubic box assuring that the protein was at least 1.2 nm away from the sides, also ion concentrations of 120 mM of KCl and 5 mM of NaCl were added to mimic neuron physiology. All simulations were done in NAMD 2.13 software [12] with CHARMM36 force field [18], the workflow followed for the three constructs consisted in 5000 steepest descent steps followed by two thermalization steps, firstly 0.5 ns at the NVT ensemble at 298K were simulated, and then another 0.5 ns at NPT ensemble were carried out, in both cases temperature and pressure were kept constant using Langevin dynamics [60, 61]; finally, 100 ns were collected again at the NPT ensemble. The length of the simulations is considered to be enough since we do not want to observe any conformational change, only a nice statistic of the interactions occurring at position 344 is sought.

Further information about these simulations can be found in its section of the [github page of the materials and methods](#) [62], in particular all input files with binary files with final snapshots of the systems at each step of the workflow. Also some magnitudes computed to control the simulation are provided, firstly, to assure that there were no major conformational changes the RMSD of all protein atoms<sup>4</sup> and the RMSF of each amino acid of the systems<sup>5</sup> are provided, together with the angles formed by the position 344<sup>6</sup> as defined in (Fig. 3.3, B and C) so that it is clear that the studied conformations of position 344 hold throughout the simulation.

In (Fig. 3.4, C), it is showed the mean interaction energy between residues of the simulated system with W344 (Blue), with W344R-N (Black) and with W344R-T (Red), where the dashed fill is the standard deviation of the energies throughout the simulated time, the interactions with CaM residues showed started from Ile-100 since the previous ones did not interact with position 344 in any of the three runs. Interaction energies were computed with the NAMD energy plugin [21] based on CHARMM36 force field [18]. From (Fig. 3.4, C: blue line), we can describe the energetic fingerprint of Trp-344, it consists in some attractive interactions with the nearby residues

---

<sup>4</sup>Follow the links for RMSD of [WT](#), [W344R-T](#), [W344R-N](#) runs.

<sup>5</sup>Follow the links for RMSF of [WT](#), [W344R-T](#), [W344R-N](#) runs.

<sup>6</sup>Follow the links for the angles of [WT](#), [W344R-T](#), [W344R-N](#) runs.

## Results

---

belonging to hA, and with two residues of CaM, namely Glu-120 and Met-124, that would serve as the anchor points of Trp-344 to CaM, finally the mean interaction with the solvent is attractive. However, W344R-N configuration shows much more strong interactions with Glu-120, Glu-123 and specially with Glu-127, where the first two interactions have big standard deviations, indicating that these interactions do not hold throughout the whole simulation, in favor of Glu-127. Notice that glutamic acids and arginine are negatively and positively charged residues, respectively, so their interaction compensates the charges, presumably resulting in the strong interactions observed but also in inducing a hydrophobic behavior in the arginine; once its extra charge is compensated what is left is a long hydrophobic side chain that interacts less with water, as it can be seen in (Fig. 3.4, C: black), suggesting that the hydrophobic nature of the WT TRP-344 is conserved or even enhanced by W344R mutation when being in a N-like conformation, which would rationalize the *in vitro* binding experiments that stated that the mutation showed even better CaM binding [32]. Remember that 344 position is located at the hydrophobic residue  $\Theta$  of the IQxx $\Theta$ R calmodulin binding motif so a hydrophobic nature is expected. Finally, the interaction of W344R-T with CaM are drastically reduced when compared to W344R-N (Fig. 3.4, B: red line), in fact, the lack of these interactions leaves the extra positive charge of W344R-T uncompensated making it very hydrophilic, as it has a very strong interaction with water, in contrast to the smaller ones showed by W344R-N or wild type. We speculate that the hydrophilicity observed in the W344R-T conformation would prevent the binding with CaM. It also stands out the interaction of W344R-T with GLN-341 (which is also a polar amino acid), which could be the one promoting the appearance of the T configuration.

This descriptions of W344R-N and W344R-T can only be understood as a plausible explanation of how the appearance of W344R-T would be deleterious to CaM binding, as the previous comments rely on the estimation of the hidrophobicity of the amino acids by only looking its interaction with water, this approach is naive as to properly asses hidrophobicity the entropy of the displaced water has to be considered, which is not trivial at all and is beyond the scope of this work.

To elaborate an hypothesis on the differential effect observed on the mutation upon CaM recognition, we computed which orientation is the most stable one in absence of CaM, since in the end this would be the one with more probability to be encountered by CaM during its recognition. To do so, three runs of 100 ns long of all atom molecular dynamics of the CRD (helices hA-TW-hB, Fig. 3.5, A) of both wild type and mutant were performed, the system was built by keeping the CRD of PDB ID: 6FEG [38], as before it was placed in a cubic box ensuring that the minimun distance between the protein and the box was half of the cutoff (0.61 nm) so that short range interactions between periodic units are prevented to happen, also ion concentrations of 120 mM of KCl and 5 mM of NaCl was introduced to mimic neuron physiology. The workflow followed for this set of simulations was repeated for each run, input structures were minimized by 5000 steps of the steepest descent algorithm, this was followed by two 0.5 ns long thermalization steps in the NVT and NPT ensemble, in which temperatures were kept at 298K using Langevin dynamics [61] and pressure was kept at 1 atm also

with a Langevin piston [60], afterwards 100 ns at NPT ensemble were collected for three independent replicas as production, all simulations were carried out with NAMD 2.13 [12] and with CHARMM36 force field [18].

Further information can be found in its [materials and methods github page](#) [63], there input data and snapshots of each simulation steps can be found, stability of the simulations through the simulated time is analyzed with the RMSD<sup>7</sup> of all atoms of the CRD and with RMSF<sup>8</sup> of each amino acid, in which it can be seen that RMSD of around 0.7 nm were observed, which is rather big deviations, however when looking at the RMSF amino acids belonging to helices entailed less RMSF and visualizations of the last snapshot guarantee that the secondary structure of the systems is always maintained, even though the systems highly fluctuate.

Angles as defined in (Fig. 3.3, B and C) were collected throughout all simulations <sup>9</sup> and stored to compute the free energy profile using the Weighted histogram Analysis Method (WHAM) [64, 65] (Fig. 3.5 B), an example of a time series of a simulation can be seen in (Fig. 3.5 B, left inset). The projection of the free energy on the angles revealed that WT has a potential well around the N orientation and the T configuration seems to be banned as no transitions were observed, in fact, simulations with the same workflow<sup>10</sup>, but starting from an output structure of rosetta protocol in which W344 was at T configuration, to start from T it quickly returned to N (Fig. 3.5 B, right inset), this was repeated three times with the same outcome<sup>11</sup>, indicating that the N configuration is the most stable one for W344. Conversely, W344R has access to a wider range of angles as the N configuration is not energetically banned, but it finds a more stable conformation around the T position. Regarding the total amount of simulated time collected in this set of simulations, it is clear that the 3 runs of 100 ns considered is long enough to observe several transitions between W344R-N and W344R-T, since the time scale associated to its energetic barrier is lower than the simulated time.

Gathering all the previous information, we can express our hypothesis that could explain the molecular mechanism according to which W344R is able to bind to CaM *in vitro* but not *in vivo*. When the folding of the CRD-CaM complex takes place at once after urea denaturation, the arginine of W344R variant can adopt its lower state bounded conformation, that would be the N orientation, in which strong electrostatic interactions between the arginine and glutamic acids take place and the arginine exhibits an hydrophobic nature that would enhance the binding. Conversely, in the ribosomal tunnel CaM has to recognize the channel and we speculate that the cotranslational appearance of the T conformation for the arginine of W344R mutation would prevent its proper binding to CaM, as we have showed that T orientation is the most stable one in absence of CaM. The metastable state of the T orientation bounded to

<sup>7</sup>Follow the links for RMSD of [WT-1](#), [WT-2](#), [WT-3](#), [W344R-1](#), [W344R-2](#) and [W344R-3](#).

<sup>8</sup>Follow the links for RMSF of [WT-1](#), [WT-2](#), [WT-3](#), [W344R-1](#), [W344R-2](#) and [W344R-3](#).

<sup>9</sup>Follow the links for the angles of [WT-1](#), [WT-2](#), [WT-3](#), [W344R-1](#), [W344R-2](#) and [W344R-3](#).

<sup>10</sup>Follow the links for RMSD of [WT-T1](#), [WT-T2](#), [WT-T3](#) and for RMSF of [WT-T1](#), [WT-T2](#), [WT-T3](#). Same comments regarding stability can be made as in previous WT and W344R simulations.

<sup>11</sup>Follow the links for angles of [WT-T1](#), [WT-T2](#), [WT-T3](#)

## **Results**

---

CaM has been computationally modeled, with the result of the loosing of interactions with the glutamic acids of CaM that left the arginine as a highly hydrophilic residue, where position 344 is known to be the hydrophobic residue of the IQ like binding motif.

Part of the previous results were published in [26], now we will present further experimental and computational efforts that were made to test our hypothesis. Looking again at (Fig. 3.4 C), tilted arginine made an attractive interaction with GLN-341, furthermore, the Pearson correlation coefficient of the angle of the arginine an the interaction energy between ARG-344 and GLN-341 yielded a value of 0.52:

$$\text{corr}(\theta, E) \frac{\text{COV}(\theta, E)}{\sigma_\theta \sigma_E} = \frac{\langle (\theta - \bar{\theta})(E - \bar{E}) \rangle}{\sigma_\theta \sigma_E} = 0.52 \quad (3.2)$$

Which was computed for values of time between 80 and 90 ns of the plot in Fig (3.5 B, left inset), this is near one transition between N and T, this constraining on the time used to compute the covariance was needed because otherwise it was hided by the usual thermal fluctuation of the MD simulation. A value of 0.52 of the correlation means that there is a partial<sup>12</sup> positive correlation between angles an energies, so an increase in angle is associated with a positive increase in energy, in other words, the smaller angles of the tilted configuration are associated with more negative interaction energies, meaning that GLN-341 would help to stabilize the tilted conformation, also when visualizing the T structures it looked like the arginine always lied close to that glutamine (Fig. 3.4, B), which could make sense as the extra charge of the arginine could interact with the polar side chain of the glutamine. Our idea to continue working on this hypothesis of the N and T configurations was the following, mutating the glutamine could entail an increase in the N orientation probability of the mutated arginine, as we would be breaking their attractive interaction, recovering a WT like behavior and restoring the posterior CaM binding.

Therefore, to successfully test this hypothesis we needed a correlation between the appearance of the N configuration with experimental binding to CaM, namely, mutations in GLN-341 that results in more *in silico* probability associated with the N orientation would increase the experimental binding to CaM, where the latter would be experimentally verified by using an experimental *in cellulo* FRET construct that was able to detect CaM binding in the past [26], it consists in placing the fluorophores in the N terminal and C terminal of the CRD, resulting in a sequence of [Fluorophore-hA-hTW-hB-Fluorophore], which is co-expressed with CaM, since it is known to arise FRET signal only if CaM recognizes the CRD, since it is not able to fold by itself but it is known to form a stable complex (PDB ID: 6FEG) [38].

Trying to anticipate energetic destabilization upon mutations on GLN-341 in CRD CaM complex, rosetta calculations of  $\Delta\Delta G$  for a few mutations were carried out, the preselection of amino acids was based on the idea that we needed amino acids whose side chain was not very big to properly fit in the hA of the CRD, aside of that a variety of them were tested. Results are in Table 3.1, rosetta predicted that all constructs will be destabilizing, but for some of the double mutants like Q341L-W344R or Q341M-W344R the global binding affinity was slightly smaller than the reference single mutant W344R, indicating a partial stabilization provoked by the mutation in

---

<sup>12</sup>Notice that the coefficient is normalized between -1 and 1 as it is divided by the standard deviations.

## Results

---

position 341, further information about input and output files can be found at its [supplementary information page](#) [66].

| Variant     | $\Delta\Delta G$ (kcal/mol) | $P_N$ |
|-------------|-----------------------------|-------|
| Q341-W344R  | 5.1                         | 0.28  |
| Q341A-W344R | 8                           | 0.24  |
| Q341C-W344R | 6                           | 0.27  |
| Q341I-W344R | 5.1                         | 0.25  |
| Q341L-W344R | 4.1                         | 0.48  |
| Q341M-W344R | 4.78                        | 0.46  |
| Q341N-W344R | 9.36                        | 0.38  |
| Q341P-W344R | 8.65                        | 0.46  |
| Q341V-W344R | 5.75                        | 0.39  |

**Table 3.1:** Results of the double mutants, binding affinities computed with rosetta in kcal/mol and probabilities of the arginine to be in a N orientation.

All atom simulations of the CRD were performed for all double mutants of Table 3.1 together with single mutant W344R; to build the systems, mutations were systematically introduced to the system of (Fig. 3.5, A), and the same procedure as before was followed, a cubic box of TIP3P water solvent [59] was generated so that the minimum distance to the protein and the simulation box was 0.61 nm, also ion concentrations of 120 mM of KCl and 5 mM of NaCl were added to reproduce neuronal physiological conditions, simulations were carried out with NAMD 2.13 [12] using CHARMM36 force field [18]. For this set of simulations the previous workflow was followed, to avoid atomic clashes coming from the introduction of the mutations, 5000 steepest descent steps were carried out, afterwards two thermalization steps of 0.5 ns at NVT followed by 1.5 ns at NPT were performed, temperatures were kept at 298K using Langevin dynamics [61] and pressure was kept at 1 atm also with a Langevin piston [60], finally three or four replicas for each construct at NPT ensemble were performed to collect 450 ns as production. This time more time was simulated for each of the variants since the differences between different double mutants would presumably be more subtle than in the previous case, in which the comparison was between the W344R single mutant and the WT, more simulated time is translated into a more resolution in the energy barrier separating the states W344R-N and W344R-T as more transitions are obtained. Similarly to the previous sets of simulations, there is supplementary information available at its [materials and method github page](#) [67] with the usual supplementary information for each replica of each double mutant; inputs and snapshots of each step of the workflow and for the production runs RMSD of the protein, RMSF of each amino acid and the angles of position 344 as defined in (Fig. 3.3, C)<sup>13</sup>.

<sup>13</sup>Instead of linking each picture, the reader is referred to the folders: Q341A-W344R([1,2,3](#)), Q341C-W344R([1,2,3](#)), Q341I-W344R([1,2,3](#)), Q341L-W344R([1,2,3](#)), Q341M-W344R([1,2,3](#)), Q341N-W344R([1,2,3](#)), Q341P-W344R([1,2,3](#)), Q341V-W344R([1,2,3](#)), Q341-W344R([1,2](#))

Instead of computing the free energy profile like in (Fig. 3.5, B), angle distribution probabilities of the arginine W344R where obtained using WHAM equations [64, 65], in order to be fit to a two Gaussian function (Fig. 3.6, black lines), each one representing one of the two possible orientations, integrating the Gaussians separately will yield the probability of the arginine of being at W344R-N and W344R-T, this methodology in the definition of the conformations was introduced to avoid arbitrariness in the computed probability associated to each state. Results are depicted in (Fig. 3.6) and the probabilities of being at the N configuration  $P_N$  are collected in Table 3.1, where it can be seen that there were some mutation that left the probability of being N ( $P_N$ ) similar, like Q341A, Q341C or Q341I whereas others indeed made this probability higher, like Q341L, Q341M, Q341N, Q341P or Q341V, which seemed encouraging to test them experimentally. However, experiments showed that all double mutants had a similar behavior as the single W344R mutant, as no FRET signal was obtained, which means that CaM binding was not recovered, proving that this approach for testing the two orientation mechanism for the CaM recognition of the Kv7.2 channel was not productive.

One issue with these kind of FRET experiments is that the FRET signal might not have been generated for a variety of reasons, making it difficult to extract information out of it, GLN-341 is indeed the Q of the IQ motif, so its alteration could be deleterious for the CaM binding, for example upon structural changes on the CRD that prevents CaM recognition or maybe by altering the binding interactions that are required for CaM binding, but even if position 341 is known to be important, mutating it was worth a shot as CaM is known to have many targets with many different binding motifs [34]. Nevertheless, the double mutant research line was a milestone in our lab as it was the first time that computational results were used to formulate an hypothesis and design experiments to test it, as we were used to work the other way around, with experiments leading the way.

## Conclusions

---

### 3.3 Conclusions

In this chapter, new mechanistic insight has been provided regarding the W344R mutation on the  $K_V7.2$  potassium channel, the problem arose from the discrepancy of the outcomes of CaM binding of this variant in *in vitro* and in *in vivo* experiments, the experiments carried out suggested that something happened cotranslationally that prevented CaM binding upon W344R mutation, however, since the experiments are indirect, they failed in providing mechanistic insight in what was happening. In order to assess this problem from an atomistic perspective, MD simulations together with rosetta calculations were performed, in fact, both methods follow different approaches to sample the configurational space, MD simulations rely on the time evolution while rosetta determines the plausibility of an ensemble of conformations; however, it has to be underlined that neither of them considered explicitly the cotranslational folding of the channel, since it is extremely challenging. It is noteworthy that in general, binding affinities computed through rosetta FlexddG package and FRET experiments agreed, as it is showed in 3.3 and in the double mutants experiments, for which rosetta predicted a destabilizing behavior and FRET experiments showed that there were no CaM binding. The exception was the *in vitro* assay of W344R mutant, in which a destabilizing binding affinity did not correlate with the CaM binding, we think that this could be because rosetta was energetically underestimating the W344R-N conformation.

The hypothesis formulated from the obtained data is that the cotranslational effect of the W344R would be to remain at the W344R-T, which would prevent CaM binding, conversely, on *in vitro* experiments the W344R-N would be the dominant one, even enhancing CaM binding. This hypothesis lies on the fact that when the CRD is immerse in water, the W344R variant has a higher probability of being at W344R-T than W344R-N while the wild type always stays at the N configuration, also, it has been showed that the strong interactions of W344R-T with water would infer high hydrophilicity in a pocket that is known to be hydrophobic, whereas charge compensation upon binding of W344R-N with glutamic acids of CaM might infer an hydrophobic behavior on W344R-N.

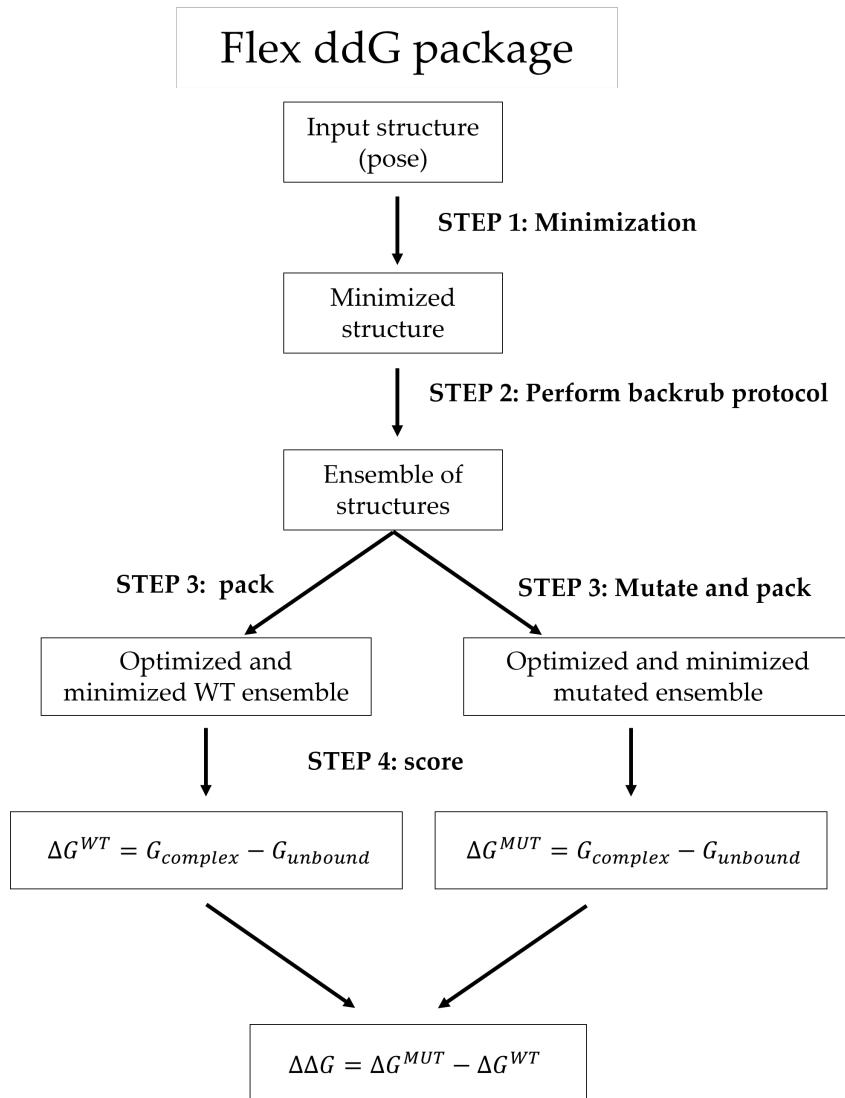
Further work on this hypothesis resulted to be unproductive; from the simulations it was clear that W344R-T was stabilized by amino acid Q341, but its mutation to other amino acids did not result in a improvement in the experimental CaM binding, even though some of the mutations on Q341 promoted the appearance of the simulated W344R-N configuration, therefore it was not possible to establish any relation between the appearance of simulated W344R-N configuration and experimental CaM; however, it could be possible that the alteration of amino acid Q341 could be very harmful for the binding, since it is the Q of the so called IQ motif. Nevertheless, this chapter contains one of the most important projects for this thesis since these were the first results obtained; we were able to elaborate an hypothesis that could explain the experimental results and after that even design a set of experiments to further test this hypothesis.

To convert this hypothesis in actual results there is still a long way to go, since there are many holes in the emerging picture coming from our results. First of all, it is

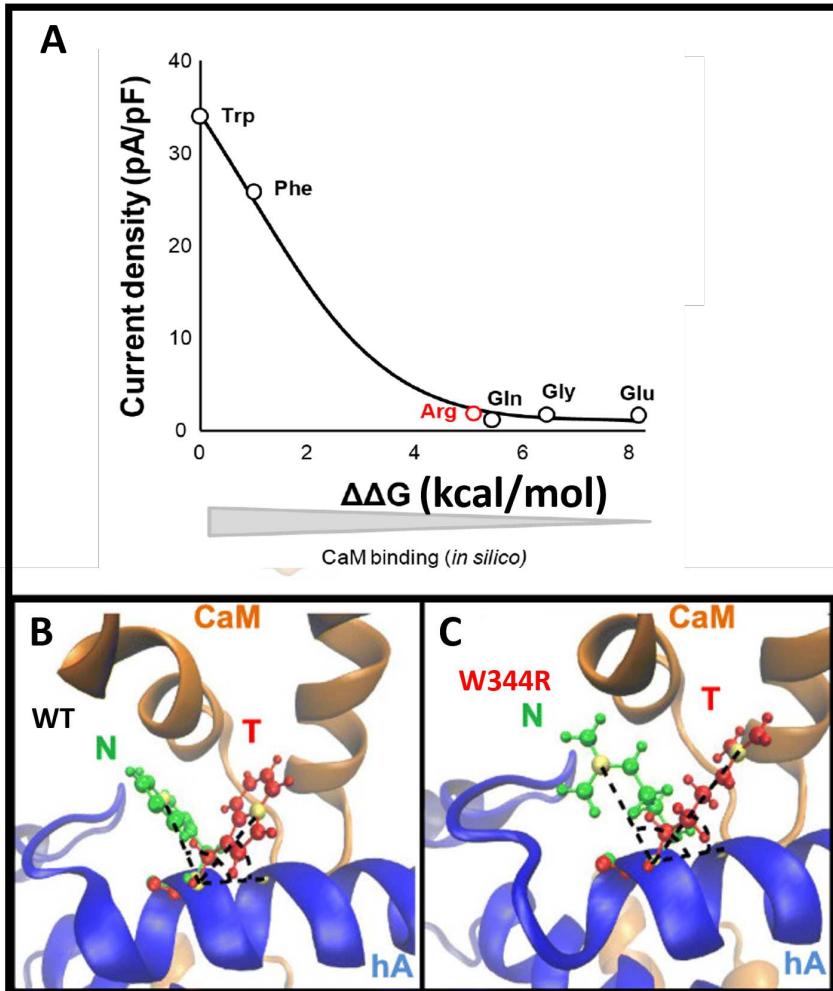
not clear how CaM recognizes the channel, it is believed to happen cotranslationally since the hairpin formed by helices hA and hB is not stable in absence of CaM [26], but it is not known if helix hA is formed when CaM recognizes the channel or if is the recognition of CaM what induces the formation of the helix, so the definitions of W344R-N and W344R-T in the absence of CaM should be taken with a pinch of salt, nonetheless, even if the two conformations are not supported by experimental information, or if it results that it is not the angle formed by amino acid 344 the parameter that prevents or enhances CaM binding and it is something else, the different interactions made by the mutated arginine with CaM for the two conformations can be enlightening, as they explain two behaviors of this arginine that would give a possible explanation for the contradictory outcomes of *in vivo* and *in vitro* experiments. The following chapter will leave aside the W344R mutation and the channel Kv7.2, in favor of the small conductance calcium-activated potassium channel SK2, with the aim of further elucidate how this recognition process of CaM takes place, since SK2 channel also has a hA like helix that serves as anchoring point for CaM.

## Conclusions

---

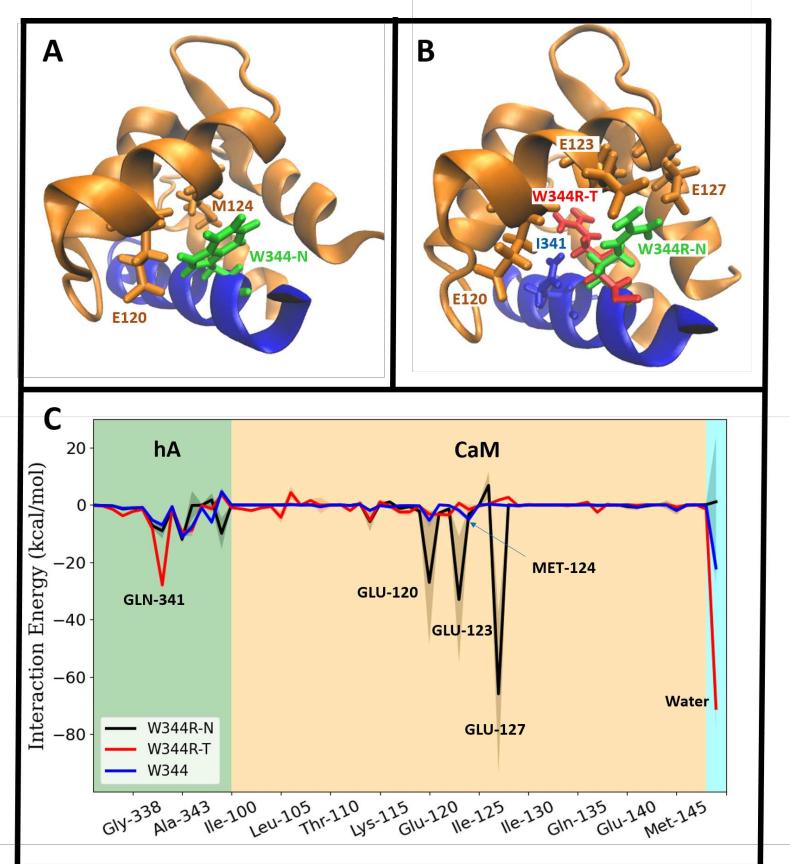


**Figure 3.2:** Chart flow with the protocol followed by the rosetta Flex ddG package.

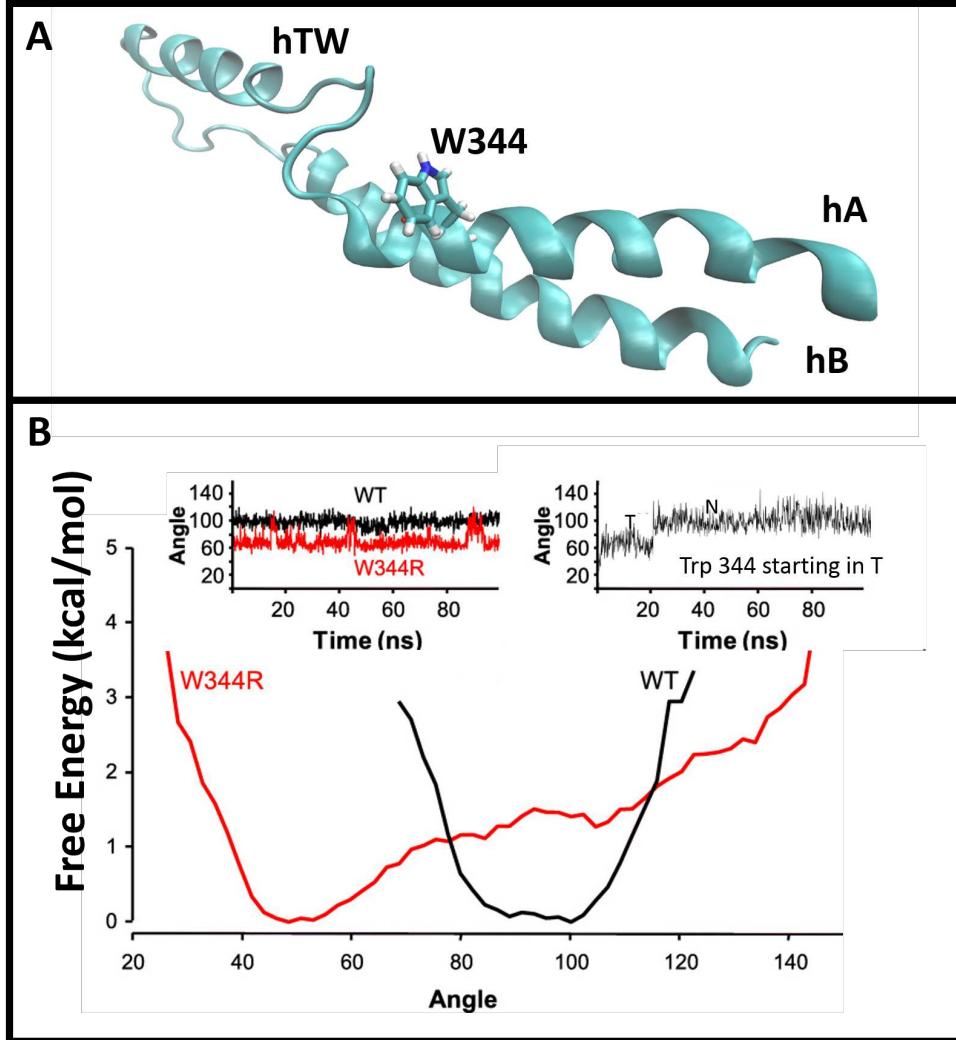


**Figure 3.3:** A, relationship between current densities and the computed binding energies of several variants of  $K_V 7.2$  channel, carrying the indicated mutations at position 344, from [26]. B-C, overlaid visualization of the native (N, green) and tilted (T, red) configurations that arose in the FlexddG calculation for WT (B) and W344R (C), with the nearby residues of CaM and helix hA. In both panels, the angle used to differentiate the conformations is plotted, the exact atoms are depicted in yellow and they are the  $C_\alpha$  of Q341,  $C_\alpha$  of position 344 and CD2 in the case of W344 and CE in the case of W344R. Made with visualization software VMD [30].

## Conclusions



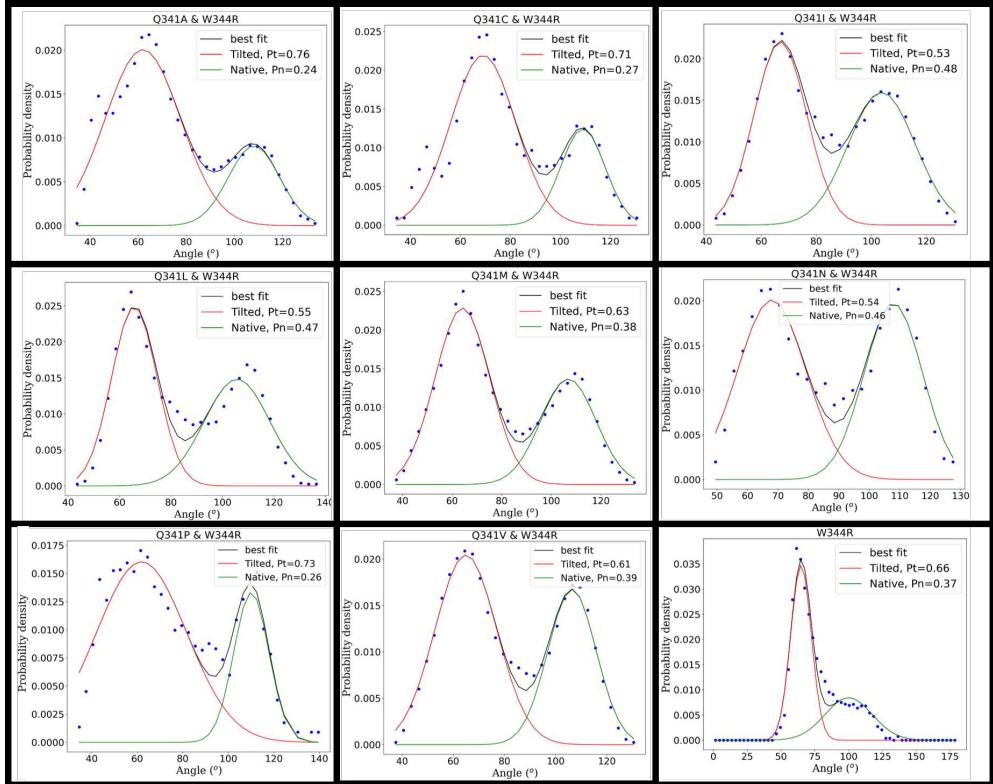
**Figure 3.4:** A, Simulated system for WT helix hA (blue), bound to CaM C lobe (orange), where W344 exhibits a N conformation (green), some important residues of the system are emphasized. B, Overlaid representation of the simulated systems containing W344R-N (green) and W344R-T (red) conformations in helix hA (blue) when bind to CaM (orange), again some important residues of CaM and hA are depicted. A and B panels were made with VMD [30]. C, interaction energies of residue in position 344 for WT (blue line), W344R-N (black line) and W344R-T (red line) conformations with the rest of the amino acids in the sequence. For ease the visualization background has been colored in green for residues of hA, orange for those of CaM C lobe and cyan for water, residues 329 to 332 in hA and 81 to 99 in CaM C lobe were also simulated but are not included in the plot as they did not interact with the position 344, lines link the mean energies obtained and errors are dashed filled areas with the standard deviation of the interaction energies throughout the simulation.



**Figure 3.5:** **A**, simulated system, consisting in the CRD and showing the location of W344, made with VMD [30]. **B**, Data collected from simulation containing helices hA, TW, hB, free energy profile projected on the angle between the helix hA and the W344 (black) and W344R mutation (red) as defined in 3.3. Left inset refers to one of the three simulations that make up the energy profile showed, right inset is an example of the simulations where the W344 was forced to start in a T orientation, follow the links in the main text for the angles of the other two runs. Adapted from [26].

## Conclusions

---



**Figure 3.6:** Probability densities of the angle formed by the arginine in position 344 for different mutations in position 341, each plot is titled with the mutations included in the set of simulations, bottom right plot is the W344R single mutant that serves as reference. Probability distributions (blue circles) have been fitted to two Gaussian distribution functions (black lines), where each Gaussian separately refers to one of the two possible states, W344R-N (green) and W344R-T (red), integration of these gaussians yield the probability of being at W344R-T orientation,  $P_T$ , and at W344R-N orientation,  $P_N$ , these values are written in the legend and also in Table 3.1.

# Chapter 4

## chapter3: $\alpha$ -helix induction in SK2 by Calmoduin

### 4.1 Introduction

In the previous chapter, when addressing the effect of the W344R mutation in the channel KV7.2, the experimental data suggested that this mutation affects the folding pathway of the channel. One of the most revealing experiments in that matter were the force profile analysis [26], in which an arresting peptide needs to experience a force in order to reestablish the synthesis of the channel, in these experiments the magnitude measured was a FRET signal that could only take place if the channel was properly synthesized and folded, the data showed that upon W344R mutation there was not FRET signal whatsoever, in the case of the wild type however, only FRET signal was observed if calmodulin (CaM) was present, indicating that CaM plays an active role in the folding of the channel at the very early stages of its folding. As it is discussed in [26], it is not clear if CaM recognizes an already folded emerging segment of the channel, namely the part of the hA that has the IQ binding motif, or if the CaM induces the adoption of the  $\alpha$ -helix exhibited in the wild type structure, which is crucial for understanding how the folding takes place and the role of CaM in it. The following chapter is a computational effort on the study of that unknown step of the synthesis, this is, how calmodulin recognizes its targets.

#### 4.1.1 Biological introduction

Calmodulin (CaM) is a small sized protein formed by 148 residues separated into two globular domains, named as N-lobe and C-lobe, each of them can carry up to two calcium ions in two helix-loop-helix motifs known as EF hands. CaM is the main calcium sensor in eukaryotic cells, one example of its importance in our organism is that it is one of the most conserved proteins in evolution, being identical in all vertebrates, mutations in CaM are usually pathogenic and have arised the name of

## Introduction

---

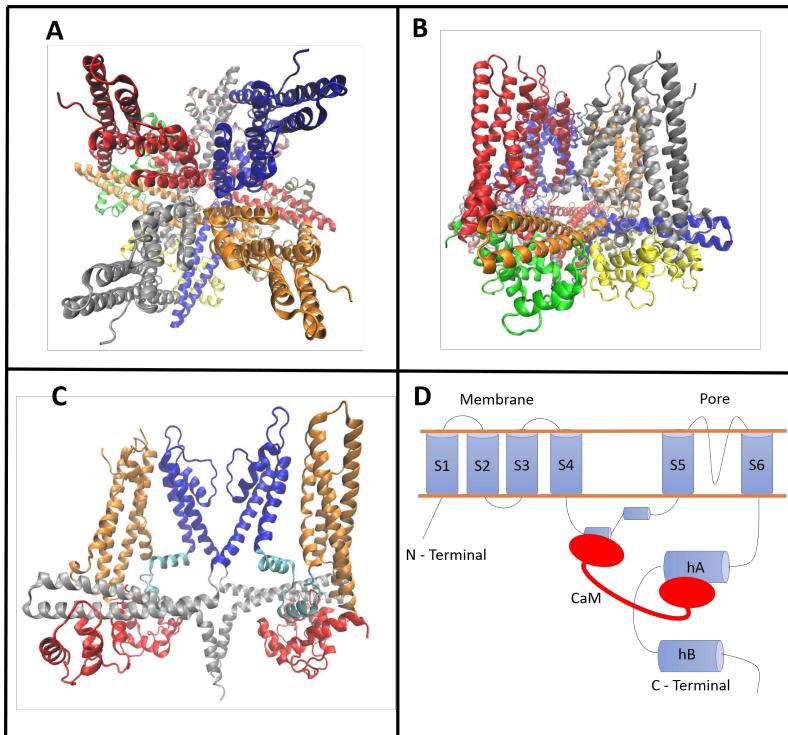
calmodulinopathies [68].

Another interesting particularity of CaM is its ability to bind to a large variety of targets, most of them display two  $\alpha$ -helical anchor points used by CaM for binding; even though their amino acid sequences differ substantially, many of them display some motifs that are repeated often [69]. As previously introduced, there are many uncertainties in CaM's recognition process due to its unspecificity towards its targets, a mixture of induced fit and conformational selection has been proposed both theoretically and experimentally [70, 71]. Conformational selection mechanism is the one by which CaM only recognizes  $\alpha$ -helices, whereas in the induced fit mechanism, CaM recognizes its binding motifs and induces in them the secondary structure.

Among the variety of CaM targets, two of them are the previously introduced KCNN and KCNQ ion channel families, in which CaM is crucial for its proper functioning, in this chapter we will focus on the SK2 channel. Like Kv7.2 channel (Fig. 3.1), SK2 is a tetramer (Fig. 4.1, A-B), where each subunit is composed by six trans-membrane helices S1-S6 and a C Terminal cytosolic domain, composed by helices hA, hB, hC and hD (Fig. 4.1, D; hC and hD are not showed). Transmembrane helices S1 to S4 form the voltage insensitive domain (ViSD) (Fig. 4.1: C, orange), it is analogous to the Kv7.2 VSD but it has been showed that it does not respond to voltage differences [27], helices S5 and S6 form the pore domain (Fig. 4.1: C, blue), which is connected to the ViSD by two small helical linkers (S4-S5a and S4-S5b) (Fig. 4.1: C, cyan), where S4-S5a is crucial as it lies in the cytosolic domain and serves as an anchor to CaM (Fig. 4.1: C, red). In the cytosolic domain, we have helices hA and hB, which form an hairpine recognized by CaM as hA has an IQ like binding motif, and finally helices hC and hD would be responsible for tetramerization, but they do not appear in the full channel cryo-EM structure [72]<sup>1</sup>. It is very impressive that the gating mechanism can be elucidated from the cryo-EM structures, in the case that each CaM has 3  $\text{Ca}^{2+}$  ions, two in N-lobe and the other in the C-lobe, the two lobes are respectively bound to the linker S4-S5a and to hA, (PDB ID: 6CNN [72]), conversely, if CaM loses the  $\text{Ca}^{2+}$  on the C-lobe, the N-lobe does not appear in the cryo-EM structure, suggesting that it is not longer bound to the S4-S5a linker and does not have an stable conformation (PDB ID: 6CNM [72]). Therefore, gathering the previous information, the hA helix would be the anchor point of CaM and the S4-S5a linker would serve as a lever that is activated by CaM for gating.

The motivation of this chapter came from a work published a couple of decades ago, in which NMR experiments showed an ensemble of structures of a small region of the SK2 channel (residues 396–487) spanning the hA and its CaM binding domain (Fig. 4.2, A and B), showing a stable helical turn [75]. As the peptide was immerse in water without any other cofactors, this results suggest that in this case a pre-helix is required for CaM binding and the rest of the helix would be induced afterwards, since this fragment is folded in the final structure (Fig. 4.2, B). An in detail analysis of the binding domain of the SK2 channel reveals that it resembles an IQ motif [34],

<sup>1</sup>The available full channel structure is the one of the family member SK4, which is very similar to SK2, in fact all the domains explained are conserved.



**Figure 4.1:** Homology model of SK2 channel, made with Swiss model (<https://swissmodel.expasy.org/>) and represented with VMD [30, 73, 74]. **A** and **B**, top and side view of the full channel, respectively, where each monomer and calmodulin protein is coloured differently. **C**, details of two opposing monomers, in which helices S1-S4 or Voltage insensitive Domain (ViSD) are in orange, linkers S4-S5a and S4-S5b are in cyan, pore domain or helices S5-S6 are in blue, helices hA and hB or Calcium Responsive Domain (CRD) are in grey and calmodulins in red, notice that in the left monomer it is visible the binding of CaM and helix hA whereas in the right one its binding to S4-S5a. **D**, cartoon representation of the structure of the channel.

## Introduction

---

although its sequence, LRxxWL, is different to the standard IQ, IQxx $\Theta$ R, where  $\Theta$  is an hydrophobic residue.

In this chapter we will study how part of the calmodulin binding domain (CaMBD) of SK2 channel acquire its secondary structure, we will focus on the linker 421-NAAANVLRETWLIYKNT-437 (Fig. 4.2 C top), which for now on will be referred to as SK2 CaMBD, and we will compare its simulated secondary structure when immerse in water and bound to the CaM C lobe, the role of the CaM C lobe in those simulations will be static, meaning that the recognition process will not be simulated, as it is out of the scope of this work, so we will focus only on the behavior of the SK2 CaMBD. We have made some strong approximations, this is reducing the system to half of CaM and a small linker of the SK2 channel, and furthermore, simulations that have CaM present will start from a bound conformation, even upon these approximations, simple molecular dynamics are not able to capture the process of secondary structure formation as we would need to simulate unaffordable time scales, therefore we will make use of the so called enhance sampling techniques, in the following we will make a short introduction of the ones used in this chapter.

### 4.1.2 enhanced sampling techniques

As stated in the introduction of this thesis, molecular dynamics (MD) simulations usually consist in performing ensemble averages in order to compute mean values of quantities, or capturing a dynamical transition between two states by computing the free energy barrier of the event, in both cases it is crucial to have a good sampling of the conformational space to obtain reliable results, in other words, we need our system to be ergodic in an affordable time scale. In general, this is very challenging for biological systems, even for the most simple ones, as they have very complicated potential landscapes with several local minimums separated by high energy barriers. To help the system to overcome this energetic barriers more easily and achieve a more efficient sampling, several methods have been proposed under the name of enhance sampling techniques [76], in this chapter we will make use of two of them, namely the Replica Exchange with Solute Scaling (REST2) [77] and metadynamics [78]. We will make a short introduction to the basic features of both methods before continuing to the results section.

### REST2

REST2 technique lies in the more general framework of the replica exchange method (REM), instead of having one simulation at one temperature in the canonical (NVT) or isothermal-isobaric (NPT) ensemble as usual, the simulation will be performed in a generalized ensemble ( $X$ ), which will consist in  $M$  non-interacting replicas ( $x_1, \dots, x_M$ ) where each  $x_i$  is a set of coordinates and momenta of the  $N$  interacting particles of one canonical or isothermal-isobaric ensemble ( $q_i^1, \dots, q_i^N, p_i^1, \dots, p_i^N$ ), weighted by its own

Boltzmann weight factor:

$$e^{-\beta_i H_i(q_i^1 \dots q_i^N, p_i^1 \dots p_i^N)} \quad (4.1)$$

where  $\beta_i = 1/k_B T_i$ , since the replicas are not interacting, the Hamiltonian of the generalized ensemble is separable, making the weight factor of the generalized ensemble ( $W_{GE}(X)$ ) also separable:

$$W_{GE}(X) = e^{-\beta H(q_1^1, \dots, q_M^N, p_1^1, \dots, p_M^N)} = \exp \left\{ - \sum_{i=1}^M \beta_i H_i(q_i^1, \dots, q_i^N, p_i^1, \dots, p_i^N) \right\} \quad (4.2)$$

And each Hamiltonian is defined as the sum of the kinetic energy and the potential,  $H_i(q_i^1 \dots q_i^N, p_i^1 \dots p_i^N) = K(p_i^1 \dots p_i^N) + V(q_i^1 \dots q_i^N)$ . We want each replica to perform a random walk in this generalized ensemble, and this is achieved by exchanging replicas occasionally. When this exchanges occur, for example between replicas  $i$  and  $j$ , the exchange can be written as follows:

$$X = (x_1, \dots, x_i, \dots, x_j, \dots, x_M) \rightarrow X' = (x_1, \dots, x_j, \dots, x_i, \dots, x_M) \quad (4.3)$$

Which means that coordinates and momenta of the particles of replicas  $i$  and  $j$  are swapped, this exchanges will take place with a probability  $w(X \rightarrow X')$ , we will impose that the average number of accepted trial moves that result in the system leaving state  $X$  to  $X'$  is the same that results in leaving state  $X'$  to  $X$ , so that the system is not biased towards a specific state, this is called the detailed balance condition and has the following mathematical form:

$$W_{GE}(X)w(X \rightarrow X') = W_{GE}(X')w(X' \rightarrow X) \quad (4.4)$$

A system that satisfies this condition is said to be a reversible Markov chain and performs a random walk in the generalized ensemble. Rearranging the terms we can compute the ratio between  $w(X \rightarrow X')$  and  $w(X' \rightarrow X)$ , using (4.2):

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = \frac{W_{GE}(X')}{W_{GE}(X)} = \frac{\cancel{e^{-\beta_1 H_1}} \dots e^{-\beta_i H_i} \dots e^{-\beta_j H_j} \dots \cancel{e^{-\beta_M H_M}}}{\cancel{e^{-\beta_1 H_1}} \dots e^{-\beta_i H_j} \dots e^{-\beta_j H_i} \dots \cancel{e^{-\beta_M H_M}}} \quad (4.5)$$

All the exponential terms that do not belong to  $i$  and  $j$  and the kinetic terms also cancel out, yielding:

$$\frac{w(X \rightarrow X')}{w(X' \rightarrow X)} = e^{[\beta_j - \beta_i](E(q_i^1 \dots q_i^N) - E(q_j^1 \dots q_j^N))} = e^{-\Delta} \quad (4.6)$$

To define the exchange probability between two replicas we have the ratio between the two transition probabilities plus the obvious condition that probabilities must be less than one, this two constraints are not enough to define  $w(X' \rightarrow X)$  and many choices can be made, one of them is the Metropolis criterium [50]:

$$w(X' \rightarrow X) = \begin{cases} 1, & \Delta \leq 0 \\ e^{-\Delta}, & \Delta > 0 \end{cases} \quad (4.7)$$

## Introduction

---

Different enhance sampling methods lie under this general framework, each of them with subtle differences, for example in Temperature Replica Exchange Method (TREMD) [79] the replicas are simulated at different temperatures  $\beta_i$ , in this thesis the method followed is Replica Exchange with Solute Scaling (REST2), in which the replicas are simulated with different Hamiltonians, methods that exchange Hamiltonians are usually called HREX and have the advantage that Hamiltonian functions are very flexible when selecting which terms are interchangeable. Specifically, REST2 simulations maintain all replicas at the lower temperature  $\beta_0$ , but the potential terms of each Hamiltonian are scaled differently, by a factor less than one, to reproduce a variety of effective temperatures  $\beta_i$ , the advantage with temperature replica exchange is that the potential is further separated into contributions coming from the solute (p) and the solvent (w) so that the solvent is left unscaled:

$$V_i^{REST2} = \frac{\beta_i}{\beta_0} V_i^{pp} + \sqrt{\frac{\beta_i}{\beta_0}} V_i^{pw} + V_i^{ww} \quad (4.8)$$

Where the square root in the interface between solute and solvent terms arise naturally from the Lennard-Jones combination rules. Applying the detail balanced condition to this scaled potential yields:

$$\Delta_{ij}^{REST2} = (\beta_i - \beta_j) \left[ (V_j^{pp} - V_i^{pp}) + \frac{\sqrt{\beta_0}}{\sqrt{\beta_i} + \sqrt{\beta_j}} (V_j^{pw} - V_i^{pw}) \right] \quad (4.9)$$

Where we recall that the exponential of this quantity rules the exchanges by the Metropolis criterium, which is therefore very sensitive to potential energy differences of solutes and their interface with the solvent for the different replicas, notice that the solvent, which has the majority of the particles and would introduce bigger energy fluctuations, does not contribute. Nonetheless, only exchanges with neighboring replicas is allowed as in practice non-adjacent replicas will have very low exchange probability.

REST2 method has been implemented to go one step beyond as it allows the user to further divide the solute into a "hot" and a "cold" region, so that only the hot region and its interface with the cold one or the solvent will contribute to the exchanges between replicas, this is done to substantially decrease the number of replicas needed to have a nice exchange probability in the effective temperature range of interest, since the number of replicas needed to obtain efficient sampling scales with the square root of the total degrees of freedom of the terms contributing to the exchanges [77].

Since only the first replica, the one that has a Hamiltonian that is not scaled, will yield physical results, only the dynamical information of the system at this first replica will be collected, however this will have conformations that have travel through all the variety of Hamiltonians, where the ones weighted by smaller values will have less energy barriers between conformations so it will be much more probable that these structures present higher energy conformations, considerably enhancing the conformational sampling, as the appearance of these structures in the usual NVT or NPT ensemble is much lower since the systems will tend to get stuck in local minimums.

## Metadynamics

The approach followed in metadynamics simulations is different [80], in this case we will project the free energy into a collective variable  $S$ , that is a function of the coordinates of the system  $S = S(\mathbf{R})$ , in such a manner that this variable fully captures the dynamic process under study, for simplicity, the collective variable will be one dimensional and therefore the free energy profile will be projected into a 2D plot. Once the collective variable is set, a history depending bias potential  $V_G(S, t)$  will be added, normally this potential is constructed by adding Gaussian functions centered around the values of the collective variable that the system has at a given time  $t'$ ,  $S(\mathbf{R}(t'))$ :

$$V_G(S, t) = \int_0^t \omega e^{-\left(\frac{S(\mathbf{R}) - S(\mathbf{R}(t'))}{2\sigma}\right)^2} dt' \quad (4.10)$$

where the constant  $\omega$  is the height of the Gaussian function used and  $\sigma$  its width. The bias potential  $V_G(S, t)$  is added to the Hamiltonian of our system, preventing the system to stay at the value of the collective variable  $S(\mathbf{R})$ . These Gaussians will be continually filling the minimums of the energetic landscape projected on  $S(\mathbf{R})$ , making the free energy barriers smaller until at some point all the local minima are full, yielding that:

$$V_G(S, t \rightarrow \infty) = -F(S) + C \quad (4.11)$$

Where  $C$  is an additive constant and  $F(S)$  is the projection of the free energy of the unbiased system on the collective variable. Under these circumstances, the effective potential felt by the particle will be a constant and therefore the probability distribution of the states alongside the collective variable will also be a constant, and the characterization of the unbiased free energy profile using (Eq. 4.11) will be converged. However, when running a simulation it is difficult to know when convergence is achieved to stop our simulation, in practice, it might happen that we continue adding Gaussian functions that introduce errors, making the correct value of  $F(S)$  fluctuate with fluctuation values depending on the size of the added Gaussians. To alleviate this, Well-Tempered metadynamics method was introduced [81], in which the height of the added Gaussians is scaled by  $w$ :

$$w = \omega e^{-V_G(S,t)/\Delta T} \tau_G \quad (4.12)$$

Where  $\omega$  is the bias deposition rate,  $V_G(S, t)$  is the value of the bias potential at the value of the collective variable  $S$  at time  $t$ ,  $\Delta T$  is a temperature-like value that can be tuned together with  $\omega$  to control how fast the height of the added Gaussians decrease and  $\tau_G$  is the Gaussians deposition rate. With this scaling scheme the final value of the free energy has to be scaled by a constant depending on the temperature of the simulation and on  $\Delta T$ :

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \quad (4.13)$$

Note that at the beginning of the simulation,  $V_G(S, t)$  will have bigger values allowing the bias potential to quickly fill the minimums of the potential and enhancing

## Introduction

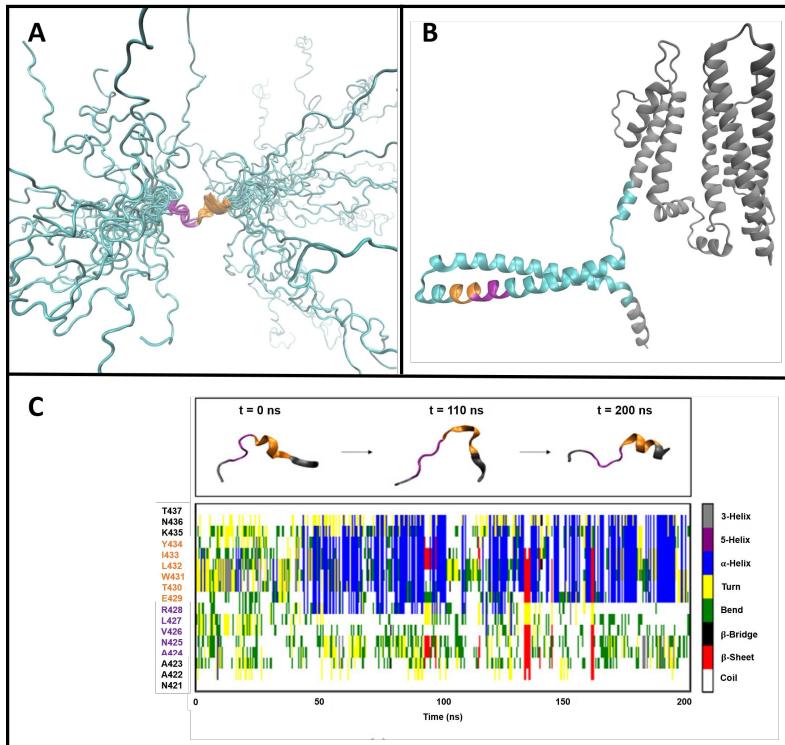
---

the conformational space sampling, but as the filling goes on the scaling will make the introduced potential decrease and eventually converge to an accurate value of the free energy alongside the collective variable.

The choice of the collective variable is crucial when setting up a metadynamics simulation, indeed, guidance and recommendations on how to choose them has been extensively reviewed [80, 82]. Essentially, metadynamics simulations are only recommended when a set of collective variable can describe a process unambiguously. Let us consider an illustrative example, if we want to measure the free energy of the dissociation of two molecules, which would play the role of a lock and a key, we could think that the distance between the center of masses could be a good collective variable and that we could project the free energy on the axes of the lock to nicely describe the dissociation process, however, for a given distance between the two center of masses the relative orientation of the molecules can differ, so that for small distances we would have a very stable configuration in which lock and key would be aligned and several others in which a rotated key is clashing with the lock, making the convergence of this simulation very cumbersome.

## 4.2 results

This chapter has two goals, the first one would be to reproduce the experimental results in which a helical turn of the SK2 CaMBD was observed by NMR experiments in aqueous solution [75], as it would be the type of conformation that CaM would encounter when recognizing the channel, and then study how this part of the channel would obtain its final  $\alpha$ -helical conformation after it has been recognized by CaM.



**Figure 4.2:** **A**, NMR ensemble of structures from PDB ID: 1KKD [75], residues forming an  $\alpha$ -helix are colored in orange and ordered structures but not helical in purple. **B**, localization of residues in **A** in a monomer of the channel, with the same color code. **C**, three snapshots of the REST2 simulations of the N421 to T437 fragment immerse in water (top) and secondary structure of the lower temperature replica throughout the simulation (bottom), according to the DSSP secondary structure definition [83]. Adapted from [34]

To asses the first goal, we tested the stability of the pre-helix in water (Fig. 4.2, A and B), for that purpose we ran REST2 simulations of the fragment 421-NAAANVLRETWLIYKNT-437 from (PDB ID: 1KKD [75]), which belongs to the calmodulin binding domain (CaMBD) and contains the pre-helix immerse in water [34], all simulations were performed in GROMACS 2016.4 [13] patched with PLUMED

## results

---

2.4.2 [84] and using the force field CHARMM36 [18]. To build the system, we solvated a dodecahedral box with TIP3P water molecules [59], with a minimum distance between protein and walls of 1.5 nm, also  $\text{K}^+$  and  $\text{Cl}^-$  ions were added to achieve a physiological concentration of KCl of 0.15 M. For the REST2 simulations, 15 replicas were set spanning a range of effective temperatures of 298K to 373K, where the temperature of replica  $i$  ( $i$  from 0 to 14) is given by the following expression  $T_i = T_0 \left( \frac{T_{14}}{T_0} \right)^{i/14}$ , which smoothly increases temperature difference between replicas with  $i$ . The hot subsystem, in which the Hamiltonian is scaled, is constituted by the peptide, and the cold one, in which the same Hamiltonian is used for all replicas, by the solvent. The workflow followed in REST2 simulations for each replica is the following, to begin with, steepest descent energy minimization was done until the maximum force was less than 100.0 kJ/(mol·nm), afterwards, 100 ps of NVT followed by another 100 ps of NPT were carried out with restrictions in the heavy atom, temperatures were kept at 298 K using a velocity-rescaling thermostat [85] and pressure to 1 bar using the Parrinello-Rahman barostat [86] in the NPT thermalization, finally 200 ns at NPT ensemble were collected as production for each replica, with exchanges attempted each 100 ps. Input parameters and snapshot of the first replica of each time step can be found at its [material and methods github page](#), in the case of the production run, time series of the effective temperature of the [first replica](#) and of [all at once](#) are also given, so that it can be seen that the first replica visited all temperatures several times in the trajectory, proving that the 200 ns long trajectory is enough to sample all temperatures, the average [exchange probabilities](#) were around 0.4, which is a sign that the temperature range was not too wide for the number of replicas used.

Secondary structure analysis of the outcome of the simulation for the first replica can be found in (Fig. 4.2, c), it reveals that there is a stable  $\alpha$ -helix between residues E429 and N436, consistent with the experiments but spanning two helical turns, whereas the experimental NMR structures only spanned one. The persistence of the appearance of this helix in the simulations is remarkable as the trajectory contains many high energy structures that could lead to other local minima lowering the total amount of helix obtained.

As previously stated in the introduction of this chapter, simulating the recognition process of the peptide by CaM is beyond the scope of this work, properly sample the total conformational space of the two proteins would be extremely challenging even with methadynamics or other enhance sampling techniques. Therefore, we skipped that step and assumed that what happens is that the C lobe of CaM recognizes the fragment containing the prehelix and bind to it; however, we were not sure if the most probable conformation for the SK2 CaMBD would be the experimental one from (PDB ID: 1KKD) or one snapshot of our simulated trajectory, since the experimental structure has only one helical turn and the simulated one two, finally we chose to run REST2 simulations for both and in the following will be referred to as 'one turn' or 'two turns', respectively. To build the systems we considered that the C lobe of CaM would recognize the fragment in the same place as it appears in the final structure, so we fitted the peptide to the experimental structure (PDB ID: 3SJQ [87]),

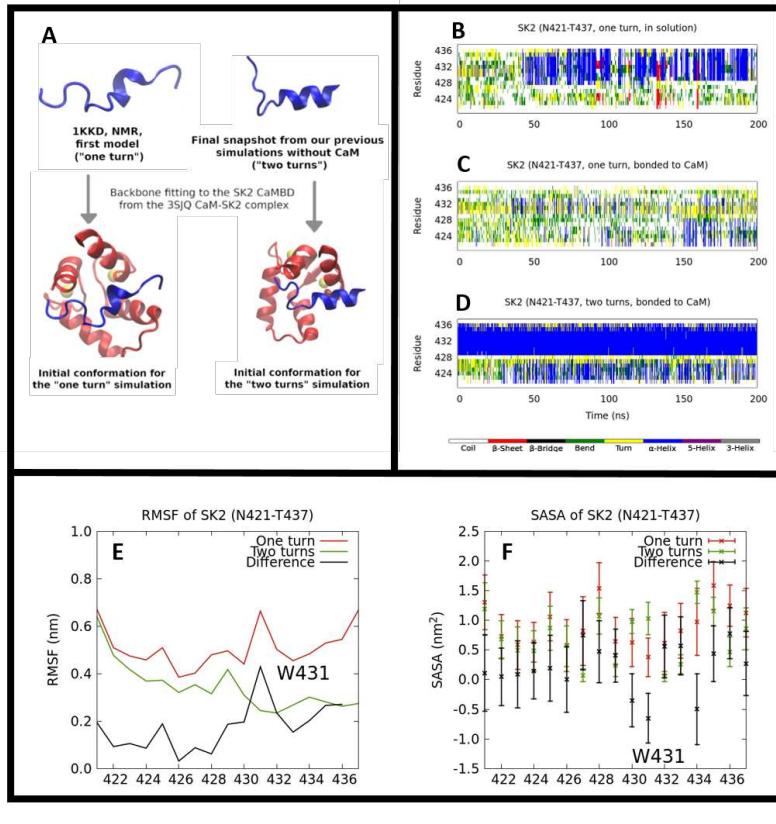
as it is depicted in (Fig. 4.3, A). As in the previous set of simulations, we solvated a dodecahedral box with TIP3P water molecules [59], assuring that the minimum distance between protein and walls were 1.5 nm, K<sup>+</sup> and Cl<sup>-</sup> ions were also added to achieve a physiological concentration of KCl of 0.15 M. The effective temperature range for the REST2 simulations was the same as before, namely 298K to 373K, in this case however, CaM C lobe and the solvent will constitute the "cold" subsystem, whereas the SK2 CaMBD peptide will be the "hot" one, notice that the number of replicas needed is related to the extension of the hot subsystem and its interactions with the cold one, which will be similar to the previous simulations in absence of CaM-Clobe, the counterpart is that the conformational space of CaM-Clobe will be substantially less sampled, so meaningful conformational changes obtained will only refer to the SK2 CaMBD.

For this set of simulations, a combination of CHARMM36 force field [18] with GROMACS 2021.3 [20] patched with PLUMED 2.7.2 [84] was used. The workflow followed is the same of the set of simulation related to (Fig. 4.3), also it has a [materials and method page](#) for this simulations, in which input files for each step can be found together with the time series of the effective temperature for the first replica ('one turn', 'two turns') and for all replicas at once ('one turn', 'two turns') [88], where it can be seen that in both cases the first replica properly explores all Hamiltonians so that both high and low energy states have been sampled. It is noteworthy than in this case the exchange probability has increased from around 0.4 in the simulation of the CaM SK2 immerse in water to above 0.75 both in 'one turn' and 'two turns' cases, where the difference lies in the interfaces between the "hot" and "cold" subsystems, indicating that the interface with water causes more fluctuations in energy than the interface with CaM C lobe plus water, that subsequently impact in the exchange probability.

Results can be seen in the right panel of (Fig. 4.3, B-D), there we have the comparison of the secondary structure of the channel in absence of CaM (Fig.4.3, A) with the same peptide starting in the 'one turn' (Fig. 4.3, B) and in the 'two turns' configuration (Fig. 4.3, C), where the last two are in presence of CaM C lobe, there it can be seen that the rest of the secondary structure can be induced in presence of CaM C lobe for the 'two turn' run, in contrast to what happened in absence of CaM, in which secondary structure in the segment N421-R428 seemed to be banned. These results differ from the ones obtained in the 'one turn' simulation, where the appearance of the helical content is not banned as it appears in different locations throughout the simulation but there is no longer a stable helix in the segment E429-T437 and there are not any frames in which the  $\alpha$ -helical content spans the whole SK2 CaMBD, this does not mean that the induction of the whole helix is completely banned, but since high energy states have been sampled it suggests that it is very unlikely so probability associated to it would be very low, making this state less productive in terms of folding.

To elucidate the mechanisms that is trapping the SK2 segment in this unfolded state we computed for each amino acid of the SK2 CaMBD the RMSF and the solvent-accessible surface area (SASA) (Fig. 4.3, bottom panel), being the latter the external area that is not buried by other solute atoms. SASA was calculated because it is

## results



**Figure 4.3:** **A**, workflow followed and depiction of the constructs build of the SK2 CaMBD and CaM C lobe complex, for the 'one turn' and 'two turns' conformations. **B-D**, secondary structure of the first replica according to DSSP definition of the SK2 immerse in water in absence of CaM (**B**) compared with the 'one turn' conformation of SK2 CaMBD bound to CaM (**C**) and the 'two turns' conformation of SK2 CaMBD bound to CaM (**D**). **E-F**, analysis of the 'one turn' (red), 'two turns' (green) simulations and their difference (black) for the amino acids belonging to the SK2 CaMBD, highlighting the amino acid W341, computed magnitudes are RMSF (**E**) and SASA (**F**).

known that CaM creates an hydrophobic pocket with its targets, so changes in SASA and RMSF will be indicators of conformational changes. It stands out that W341 exhibits higher mobility (Fig. 4.3, E) and less solvent exposure (Fig. 4.3, F) in the 'one turn' simulation when compared with the 'two turns' one, visual inspection of the trajectories proves that SASA reduction of the 'one turn' simulation is due to the fact that W341 is facing the CaM C lobe and the RMSF increase because it does not have secondary structure, in contrast to the 'two turns' case, in which W341 is facing the solvent and belongs to an  $\alpha$ -helix. Therefore the lack of secondary structure enhances mobility of the tryptophan and it is kept in the hydrophobic pocket of CaM

lowering its SASA. Besides, it coincides that some CaM C lobe residues that interact with W341 in the 'one turn' simulation (this is, when facing CaM-Clobe) interacts with L427 in the 'two turns' one , namely F92, M109, L112 and M145, suggesting that in the 'one turn' simulation, W341 has competed with L427's binding pocket leading to a less productive conformation.

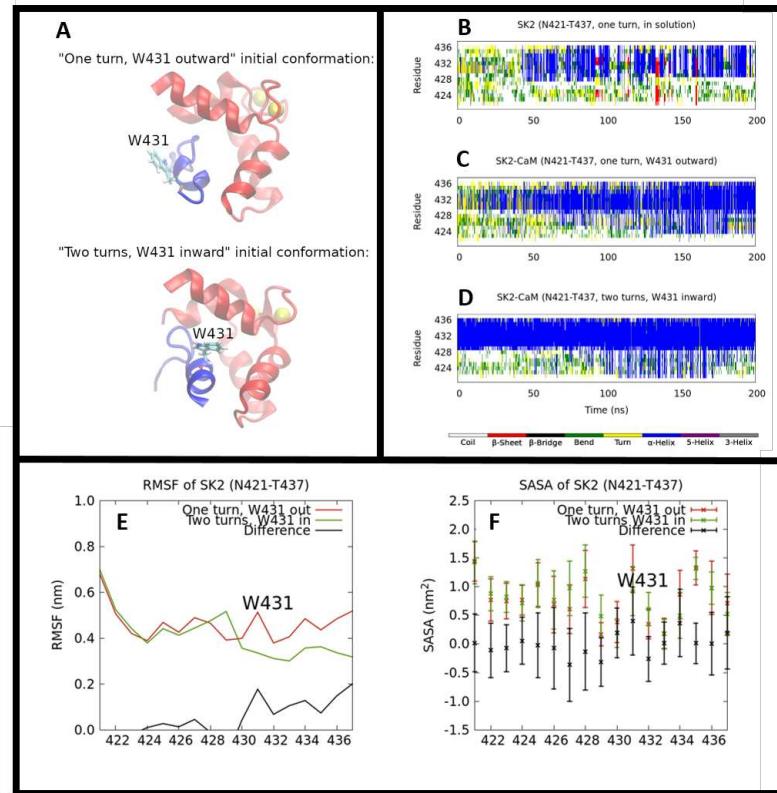
The previously discussed competence for the CaM C lobe residues appear in the in the 'one turn' simulation, in which W341 is facing CaM, therefore we wanted to see if preventing this competence the rest of the helix could be induced, so we rotated W341 180 degrees forcing it to point outwards, this is, towards the solvent (Fig. 4.3, A-top). The fact that we obtained a less productive state with the W341 facing CaM was shocking as experimental conformations of CaM-SK2 complex shows that W341 indeed points to CaM (PDB: 3SJQ, [87]), so we also rotated the 'two turn' configuration to point inwards to the CaM to see if the  $\alpha$ -helix induction holds also for this configuration (Fig. 4.4, A-bottom).

REST2 simulations were done for this two new "swapped" configurations, the building of the system, software and workflow is the same as the previous set of simulations, as usual, supplementary information can be found at its [materials and methods github page](#) [89], namely input files of all steps of the workflow with snapshots and the time series of the effective temperature for the first replica ('one turn', 'two turns') and for all replicas at once ('one turn', 'two turns').

Outcome of theses simulations are presented in (Fig. 4.4, C-D), and are also compared with the channel segment immerse in water in absense of CaM (Fig. 4.4, B). Differently to the previous case, the 'one turn' simulation produces the whole helix when W341's side chain is pointing outwards, furthermore, the helical content starts forming around this tryptophan (Fig. 4.4, C). For the 'two turns' simulation with the W341 pointing inwards (Fig. 4.4, D), the secondary structure obtained is similar to the previous case in which the side chain was pointing outward, the helical content of E429-N436 segment is maintained and the rest of the helix can be induced. In this second set of simulations, RMSF and SASA was also computed (Fig. 4.4, bottom panel), indicating that the tryptophan has now similar mobility when pointing inwards and outwards (Fig. 4.4, E) and the SASA difference is now not conclusive due to its big error bars (Fig. 4.4, F).

Putting the results together, we have seen that the W341 pointing inwards starting from the 'two turns' configuration becomes productive, this is translated into a descent in the mobility of the tryptophan from  $\approx 0.7\text{nm}$  in the 'one turn, inward' (Fig.4.3, E) to  $\approx 0.5\text{nm}$  in the 'two turns, inward' (Fig.4.4, E), since the tryptophan is constrained into a helical conformation. It is also noteworthy that when looking at the two trajectories that had W341 facing to CaM, ther is an increase of SASA from  $0.4 \pm 0.3\text{nm}^2$  in the 'one turn, inward' (Fig.4.3, F) to  $0.9 \pm 0.5\text{nm}^2$  in the 'two turns, inward' (Fig.4.4, F), meaning that W344 was better buried in the 'one turn' simulation, this SASA difference can be related to the contacts made with CaM in both situations, in the 'one turn, inward' we had that the residues of CaM A88, V91, F92, M109, L112, F141, and M145 made a contact with W341, but in the 'two turns,

## results



**Figure 4.4:** **A**, starting structures of the second set of simulations, in which the SK2 CaMBD was rotated by 180°. **B-D**, secondary structure of the first replica according to DSSP definition of the SK2 immerse in water in absence of CaM (**B**) compared with the 'one turn' conformation of SK2 CaMBD bound to CaM with the W341 pointing outwards, this is towards the solvent (**C**) and the 'two turns' conformation of SK2 CaMBD bound to CaM with the W341 pointing inwards, this is towards CaM C lobe (**D**). **E-F**, analysis of the 'one turn' (red), 'two turns' (green) simulations and their difference (black) for the amino acids belonging to the SK2 CaMBD, highlighting the amino acid W341, computed magnitudes are RMSF (**E**) and SASA (**F**).

inward' these contacts are reduced to M145 alone, it is the residue V426 the one that now makes contacts with CaM in the frames that the full helix is induced, being these E120, E123, M124, and E127.

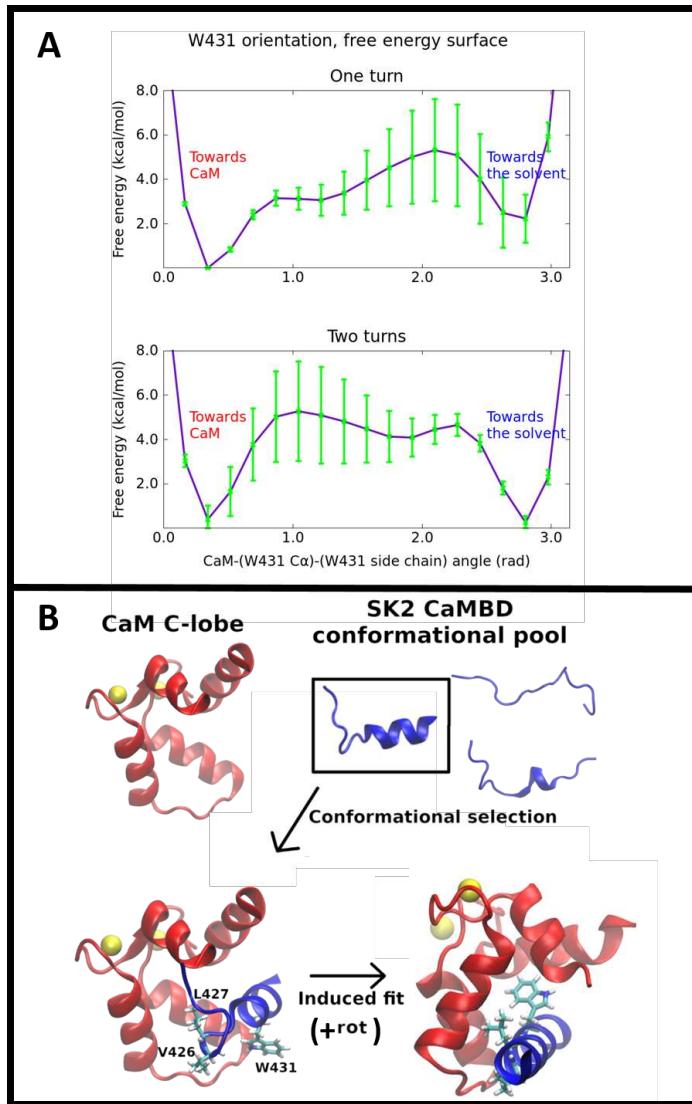
Altogether, it has to be highlighted that the previous REST2 simulations in presence of CaM C lobe did not achieve its final purpose, this is, efficiently sample the conformational space, since simulations starting from different conformations raised different results, more precisely, orientation changes in the orientation of the tryptophan were not observed, even though it seems to be important in the folding process,

likely, the energy barrier that prevents this spin to happen is bigger than the thermal energy of the highest temperature that we could afford (373K). To overcome this limitation and estimate the energy barrier separating those states we needed a different approach and metadynamics simulations turn out to be appropriate.

The systems used where the previous analyzed 'one turn' and 'two turns', which were already minimized and thermalized, as before, simulations were ran with CHARMM36 force field [18] in GROMACS 2021.3 [20] patched with PLUMED 2.7.2 [84]. For defining the collective variable for the addition of the Gaussians, the angle formed by the center of mass of CaM C lobe,  $C_\alpha$  carbon of residue W341 and the center of mass of the hexagonal aromatic ring of W341 was selected, Gaussian functions to the Hamiltonian were added each picosecond, being these centered around the value of the angle at that simulated time and with a height of  $2.0\text{ kJ/mol}$  and a width of 0.5 rad. We observed that sometimes the effect of the addition of these Gaussians provoked the dissociation of the CaM C lobe and the channel fragment, therefore we had to add a funnel like potential preventing it. 1000 ns long well-tempered metadynamics simulations at the NPT ensemble showed that there are two differentiated minima, one near  $0\text{ rad}$  of the previously described angle, this is pointing towards CaM, and another near  $\pi\text{ rad}$ , pointing towards the solvent, for both 'one turn' and 'two turns' starting conformations (Fig. 4.5, B). In both cases, the energy barrier obtained was  $5.3 \pm 2.3\text{ kcal/mol}$ , where the error is computed by doing the standard deviation of the barriers obtained in the last 200 ns of the trajectory. The obtained height of the energy barrier is similar to the one obtained for the tryptophan of peptide M13 bound to CaM with a similar methodology as ours [90]. More information of this simulations can be found in its [materials and methods github page](#) [91], in particular input files needed by gromacs to perform the MD and also the input for PLUMED plugin that is responsible of the metadynamics, together with the [time series and convergence](#) of the collective variable for the two simulations.

With the information of the stability upon orientation angle and the secondary structure analysis we can sketch the recognition mechanism of the SK2 channel by CaM C lobe (Fig. 4.5). Conformational pool of the SK2 CaMBD channel is mainly formed by the 'one turn', 'two turns' or coiled like conformations, according to the NMR experimental results (Fig. 4.2, A) and our REST2 simulations (Fig. 4.2, C), from these conformations CaM would select the productive conformations for SK2 CaMBD, namely both 'two turns' conformations and the 'one turn' that has the tryptophan pointing towards the solvent. As showed in the REST2 simulations, the rest of the  $\alpha$ -helix would be induced afterwards, and in the case that the tryptophan is facing the solvent, it would have to overcome the energy barrier of  $5.3 \pm 2.3\text{ kcal/mol}$  since experimental structures solved have this tryptophan facing the CaM C lobe [87].

Gathering the results from the simulations, it seems that the three most important residues of the channel for inducing the helix are W431, V426, and L427; in fact, X-ray studies have determined L427 and W431 as the most interacting residues with the hydrophobic pocket of CaM C lobe [87] and NMR experiments showed that mutating V426 and L427 to glycines prevented the appearance of the pre-helix of the



**Figure 4.5:** **A**, depiction of the CaM C lobe recognition mechanism of SK2 CaMBD that arises from our simulations, which is a combination of the selection of the conformations of SK2 CaMBD that have two helical turns and then the induction of the rest of the helical content together with a rotation if the selected conformation has the W341 pointing the solvent. **B**, free energy profile projected on the angle between the CaM C lobe, the  $C_\alpha$  of W341 and its aromatic ring, in  $\text{rads}$ , computed by means of metadynamics simulations for both 'one turn' and 'two turns' conformations, alongside 1  $\mu\text{s}$  long simulations and the errors represent the standard deviation of the free energies during the last 200ns.

SK2 CaMBD [75]. In our simulations, the role of these three residues can be sketched as it seems that V426 and L427 need to make contacts with the hydrophobic pocket of CaM, but there can be situations in which W341 can compete for these contacts making the state less productive.

Finally, with the purpose of generalizing the interesting results obtained for SK2 channel to other CaM targets, REST2 simulations of several CaMBDs (see Table 4.1) immerse in water were carried out, following the same methodology as in the simulation referring to (Fig. 4.2, C) [34], input parameters of each step of the workflow for each simulation can be found in its [materials and methods page](#) [92]. The first step to generalize our results was to include the rest of the members of KCNN family, this is to [KCNN1](#), [KCNN3](#) and [KCNN4](#) channels. A further generalization was to include some constructs that contain the so called IQ like binding motif, like [KCNQ1](#), [KCNQ2](#), [MYH7](#), [CAC1C](#), [INVS](#) or [IQGA2](#). Further information of the analyzed sequences can be found in Table 4.1, in the table residues V426, L427 and W431 are highlighted for all sequences including those residues or analogous, note that KCNN1 and KCNN3 have the same sequence so only one KCNN1 was simulated. In order to assure that all effective temperatures were sampled in all systems, the time series of the temperature indexes of the first replicas are provided<sup>2</sup>, also exchange probabilities are provided<sup>3</sup>. Secondary structure analysis of these targets are in (Fig. 4.6, A-H), it is interesting that residues V426, L427 and W431 are conserved through all KCNN family and the appearance of a robust pre-helix is also maintained, on the other hand, structures that only have W431 like KCNQ1 and KCNQ2 or do not have any of the three residues, like MYH7, CAC1C, INVS, or IQGA2, the helical content appears but it is not as persistent as in the KCNN family. This results suggest that the previously described mechanism can be generalized to the entire KCNN family since they exhibit similar behaviour (Fig. 4.2, C; Fig. 4.6 A-B), and that could also be present in the IQ motifs, since the ones analyzed here are also able to form pre-helices when immerse in water (Fig. 4.6, C-H), even though they are not as persistent as those of KCNN family.

---

<sup>2</sup>Follow the links for exchange time series of [KCNN1](#), [KCNN4](#) [KCNQ1](#), [KCNQ2](#), [MYH7](#), [CAC1C](#), [INVS](#) or [IQGA2](#)

<sup>3</sup>Follow the links for exchange probabilities of [KCNN1](#), [KCNN4](#) [KCNQ1](#), [KCNQ2](#), [MYH7](#), [CAC1C](#), [INVS](#) or [IQGA2](#)

## Conclusions

---

| Entry name  | ID                     | Sequence           | Nº residues |
|-------------|------------------------|--------------------|-------------|
| KCNN1_HUMAN | <a href="#">Q92952</a> | NAAANVLRETWLIYKHT  | 17          |
| KCNN2_HUMAN | <a href="#">Q9H2S1</a> | NAAANVLRETWLIYKNT  | 17          |
| KCNN3_HUMAN | <a href="#">Q9UGI6</a> | NAAANVLRETWLIYKHT  | 17          |
| KCNN4_HUMAN | <a href="#">O15554</a> | ESAARVLQEAWMFYKHT  | 17          |
| KCNQ1_HUMAN | <a href="#">P51787</a> | PAAASLIQTAWRCYAAE  | 17          |
| KCNQ2_HUMAN | <a href="#">O43526</a> | NPAAGLIRQSAWRFYATN | 17          |
| MYH7_HUMAN  | <a href="#">P12883</a> | SRIITRIQAQSREGVLR  | 17          |
| CAC1C_HUMAN | <a href="#">O13936</a> | FYATFLIQEYFRKFKKR  | 17          |
| INVS_HUMAN  | <a href="#">Q9Y283</a> | DIAAFKIQAVYKGYKVR  | 17          |
| IQGA2_HUMAN | <a href="#">Q13576</a> | EENVVKIQAFWKGYKOR  | 17          |

**Table 4.1:** Simulated constructs, referred to (Fig. 4.6). Information provided contains UniProt entry name, UniProt entry ID, simulated sequence and number of residues of each construct, information extracted from UniProt database (<https://www.uniprot.org/>), Uniprot entry IDs are linked to their data base page.

## 4.3 Conclusions

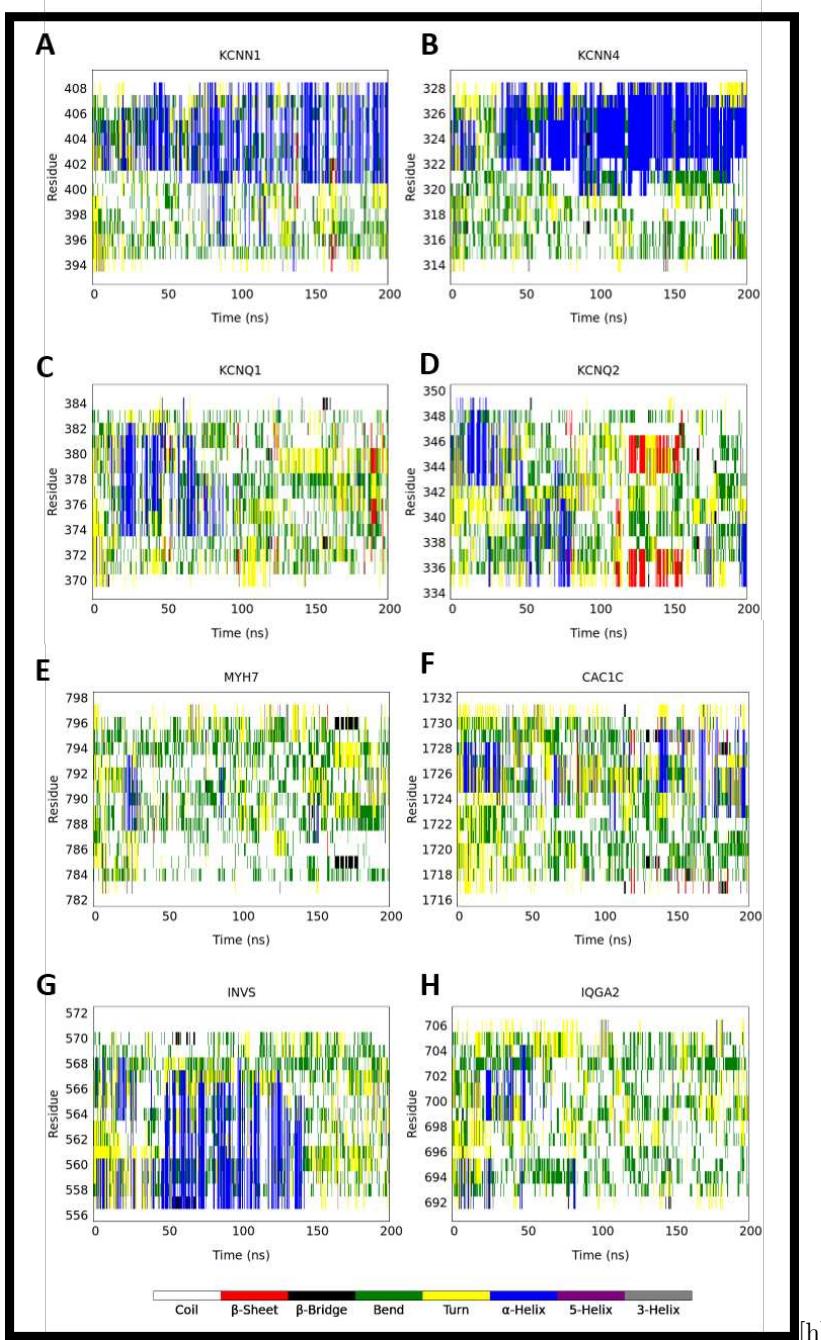
In this chapter, a computational effort has been done in order to gain insight in how CaM recognizes its targets, focusing in the channel KCNN2. Even though we have assumed the starting structure of the system composed by the CaM C lobe and SK2 CaMBD, enhanced sampling techniques had to be used in order to observe secondary structure formation, for that manner HREX REST2 simulations were carried out. In the meanwhile, we realized that the orientation of W341 was crucial for helix induction in the 'one turns' simulations, however, angle transitions were not observe throughout the simulations and we had to estimate the energy barrier through metadynamics simulations.

Our results indicate that the recognition process takes place following a mixture of the two established theoretical frameworks, which are conformational selection and induced fit. SK2 CaMBD exhibits some helical content when immerse in water experimentally (Fig. 4.2, A) [75] and computationally (Fig. 4.2, C), suggesting that this helical turns are needed for posterior recognition (conformational selection), but the apparent inability of forming more helical content would imply that CaM is needed to achieve the full secondary structure observed in experimental structures [72, 87]. In fact, how this induction takes place has been observed in our HREX REST2 simulations (Figs. 4.3, 4.4), also arising conformations that are not productive in terms of secondary structure formation, like the 'one turn' pointing towards CaM. Aiming to generalize our results, other CaM targets immersed in water were simulated to see if they were able to form some helical content in absence of CaM (fig. 4.6), from the results it was clear that the other members of the KCNN family have a similar behavior, and presumably their secondary structure will be induced in a similar manner, in contrast, other CaM targets that contain an IQ motif exhibited a less persistent secondary structure, but it has to be emphasized that helical content was found for all

targets, making us think that the previously described process for target recognition would hold for them, and the difference in secondary structure persistence could be related with the time extent of the recognition process.

## Conclusions

---



**Figure 4.6:** A-H, secondary structure according to DSSP definition of some CaM targets immerse in water and in absence of CaM.

# Chapter 5

# Chapter 4: Full channel simulations

## 5.1 Introduction

The following chapter will be the last one of this thesis and it is separated from the previous ones, that refer mainly to cotranslational folding and secondary structure adoption. This chapter is a computational effort to try to elucidate the effect of some drugs in the full channel structures of KCNQ2 and KCNN4 ion channels. Lately, as a consequence of the revolution that the cryo-EM field is experiencing, a wide variety of full ion channel structures have been reported with an outstanding resolution, in particular KCNQ1, KCNQ2 [10, 31] and small conductance calcium activated SK channel KCNN4 (also known as SK4) [72], furthermore, these structures have been obtained in presence of cofactors, like the lipid PIP<sub>2</sub>, which plays a role in many signaling events, or even drugs, leaving the possibility to study their effect from a computational perspective, recall that cryo-EM structures only contain static information of a stable conformation, meaning that some dynamical information can be obtained through simulations to boost the understanding of ion channels one step beyond.

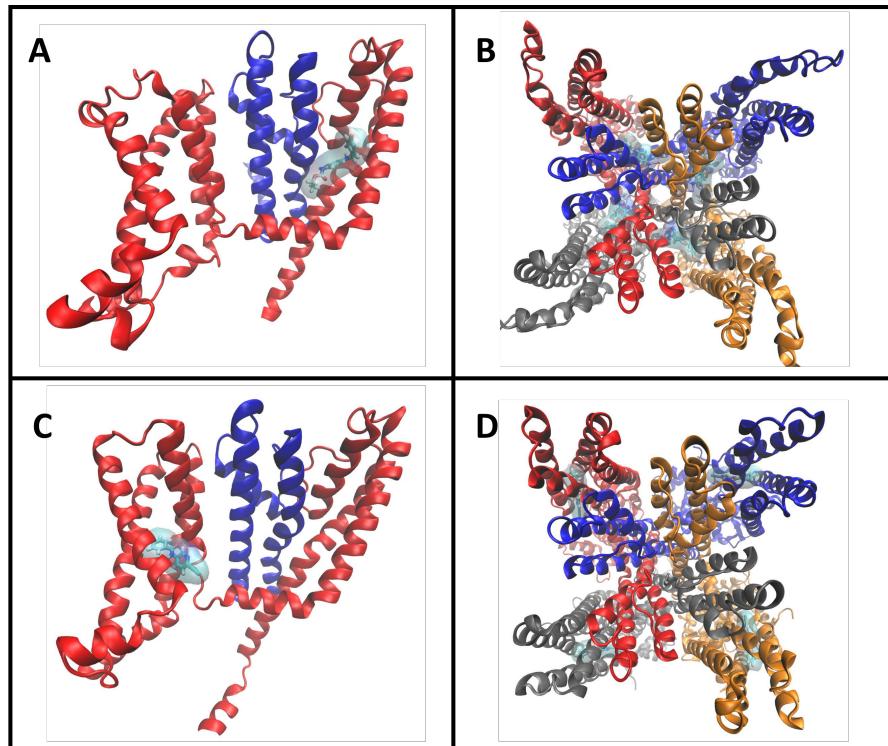
Simultaneously, the increment of accessibility to computational power has made it possible to simulate these full channel structures for periods of time of hundreds of nanoseconds [93], or models built by homology when only similar structures to the target one were available [55]. In our case, it was possible to acquire the computational power needed thanks to the Spanish high performance computing network (RES), otherwise we would not have been able to perform them.

In the case of the KV7.2 channel, the two available drugs were Retigabine (RTG, PDB ID:7CR7) [31] and ZTZ240 (ZTZ, PDB ID: 7CR4) [31], which have been reported as activators of the channels [31], furthermore, they have different effects on the channel, on the one hand ZTZ240 increases the maximum current amplitude and enhances the voltage sensitivity so that less voltage differences are needed for gating,

## Introduction

---

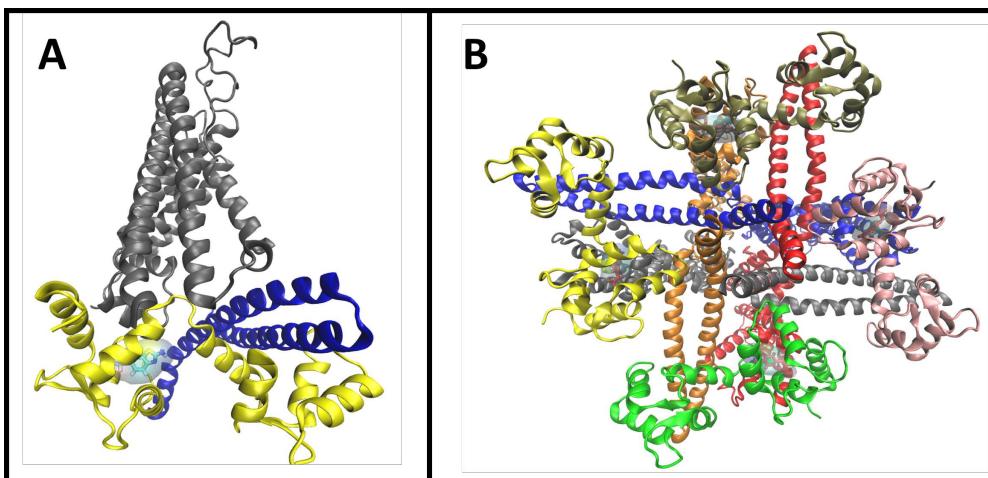
whereas RTG only enhances voltage sensitivity. The two drugs presumably alters the channel functioning differently, to support the idea that the mechanisms underneath these drugs may differ, there is the fact that their binding pockets lie in different domains; RTG locates near the pore (Fig. 5.1, A: cyan surface), between helices S5, pore helix, S6 (red) and helix S6' of the adjacent monomer (blue), it maintains the tetramersical arrangement of the channels as up to four RTG can be found in the experimental structure (Fig. 5.1, B). In contrast, ZTZ240 is situated in the voltage sensor domain of the channel (Fig. 5.1, C: cyan surface) and the drug also appears in the four monomers (Fig. 5.1, D).



**Figure 5.1:** Graphical representation of the binding pocket of the drugs Retigabine and ZTZ240 in channel Kv7.2. **A**, insight of the binding pocket of Retigabine (cyan surface), between helices S5, pore helix, S6 of the voltage sensor domain (red), and the helix S6' of the adjacent monomer (blue). **B**, top view of the channel to show the four Retigabines bound to the channel. **C**, insight of the binding pocket of ZTZ240, in the voltage sensor domain (red), for consistency, adjacent monomer is also showed (blue). **D**, top view of the channel bound to ZTZ240, with the drug appearing in the four monomers.

To complement the variety of the drugs under study, we added Riluzole to the project, it is known to be the first FDA-approved treatment for amyotrophic lateral

sclerosis (ALS), and current evidence indicates that this drug works by allosteric activation of KCNN channels[94], as its binding pocket has been experimentally resolved in the SK2 channel lying in the cytosolic domain, and more specifically in the calmodulin binding domain [94]. It was possible then to build a model of the full channel structure of SK4 [72] with the riluzole in the binding pocket defined by the reduced SK2 structure [94], since the full channel structure of SK2 channel is not available. As it is depicted in (Fig. 5.2, A: cyan surface), Riluzole finds its binding pocket between the N lobe of CaM (yellow) and a small helix belonging to the voltage sensor domain of the channel (grey), following the observed behaviour of the drugs in the  $K_V7.2$  channel, we placed one riluzole in each monomer (Fig. 5.2, B), in this picture, for better visualization of the pocket, the bottom view of the channel is showed since Riluzole is located in the cytosolic domain.



**Figure 5.2:** Graphical representation of the homology model built for the binding pocket of Riluzole in channel SK4. **A**, insight on the binding pocket of Riluzole (cyan surface), between the N lobe of Calmodulin (yellow) and the voltage sensor domain of the channel (grey). **B**, bottom view of the homology model of the channel SK4 in presence of Riluzole, built from the experimental binding pocket of the channel SK2, for which the full channel structure is not available.

The question that we wanted to address in this project is that if it is possible to characterize the effect of these drugs through simulations, we have to keep in mind that MD simulations have limitations, for example we are restricted to time scales in the order of hundreds of nanoseconds, this is not enough for such big structures to experience major conformational changes that could result in the experimentally observed effects for the drugs, moreover, the amount of data obtained from such big simulations can be really cumbersome, as the possible features to be analyzed is enormous considering that the drugs have to work allosterically, considering that the binding pockets of the drugs lie far from the pore but they actually influence its opening. In

## simulation details and preparation

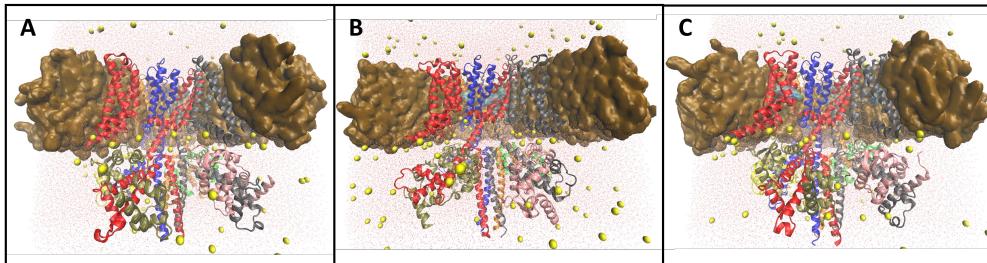
---

order to deal with these two limitations, a computational graph-based approach called “network analysis” [93] was created.

In the following chapter, we will describe the methodology followed to build the systems and run the simulations, as some bioinformatic tools were used. Then we will report a characterization of the binding pocket of the riluzole, as there were some dynamical properties that suggest a possible mechanism for the riluzole. Finally the data obtained from performing network analysis will be presented together with a thorough description of the technique, which is given as an appendix.

## 5.2 simulation details and preparation

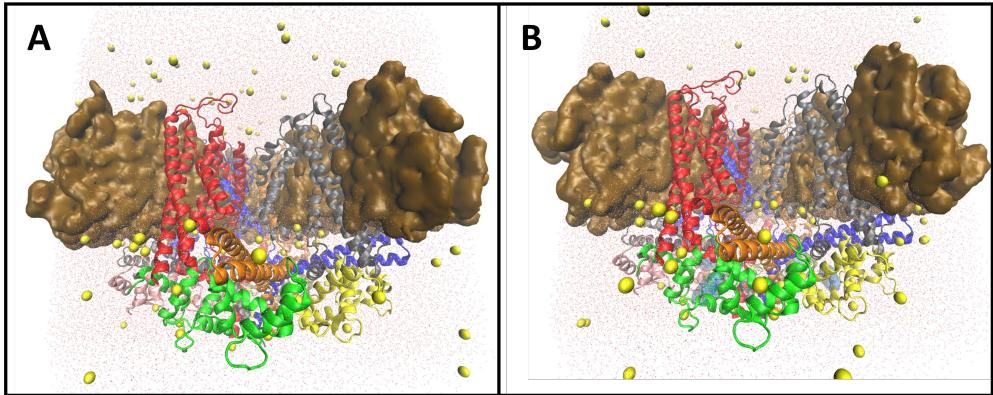
All-atom molecular dynamics simulations of the full SK4 and KCNQ2 channels were performed both in presence and in absence of the drugs, using the apo ones as control systems (Figs. 5.3 and 5.4). Each simulation will be run in triplicate, so we will obtain a total of 15 trajectories. Note that there would be two equal control systems of Kv7.2 apo (this is, without drugs) and only one will be made.



**Figure 5.3:** Graphical representation of the full channel simulations performed, lipids are represented in brown surface, water solvent as red dots, solvent ion as yellow spheres, and the complex of channel and drugs hold the format of (Fig. 5.1), for **A**, apo run of KCNQ2 **B**, KCNQ2 in presence of RTG and **C**, KCNQ in presence of ZTZ240, they will be referred to as KCNQ2\_APO, KCNQ2\_RTG and KCNQ2\_ZTZ.

Missing residues were modeled with Swiss model (<https://swissmodel.expasy.org/>), an online server that reads a sequence of amino acids and aligns it using as templates all proteins that have similar sequences [73, 74], in our case only a couple of loops were missing and therefore the templates used by the program were the full channel structures, so our structures were not altered in this step.

To continue with the system preparation, online server CHARMM-GUI (<https://www.charmm-gui.org/>) was used [95], all channels were inserted into a 130Å-long one of POPC lipids, except for the Kv7.2 channel in presence of RTG which was inserted into a 150Å-long bilayer [96], since it is in a different conformation that makes it wider in space. the systems were then solvated using TIP3P water [59], selecting the



**Figure 5.4:** Graphical representation of the full channel simulations performed, lipids are represented in brown surface, water solvent as red dots, solvent ion as yellow spheres, and the complex of channel and drugs hold the format of (Fig. 5.1), for **A**, apo run of KCNN4 and **B**, in presence of Riluzole, they will be referred to as SK4\_APO and SK4\_RIL, respectively.

option of pore water generation of step 1 of CHARMM-GUI, also 0.1 M of KCl ions were also added in this step. Finally, gromacs inputs for equilibration and production runs were also generated with CHARMM-GUI [97].

Finally, to avoid instabilities since CHARMM-GUI distributes the water molecules randomly, we erased the ones that lay within 2 Å of the protein or within 8 Å of the membrane lipids.

Parameters chosen were based on CHARMM force fields, for protein atoms CHARMM-36 was used [18] and for lipids its extension for lipids [98]. For the drugs, however, there were not parameters available and we produced them using the ligand reader and modeller of CHARMM-GUI [99].

As stated before, in the case of SK4 and riluzole, there is no structural information of the binding pocket, so we performed a docking calculation to construct the initial structure, taking the binding pocket of SK2 (PDB ID: 5V02) [94] as a starting point by means of the Schrodinger software [100].

For equilibration of the systems, the standard procedure of CHARMM-GUI was followed, it consists of 6 steps, all of them at the NPT ensemble with constraints applied to the membrane and protein, these constraints are gradually decreased in each step. Following reference [93] we increased the last one to 3ns, since the systems are quite large, they might need more time to stabilize its root mean square deviation and most importantly, we wanted to be positive that the systems were perfectly equilibrated as the simulations are computationally very costly. Afterwards, three independent replicas of 500 ns were produced for each system, where all simulations were performed in GROMACS 2016.4 [13].

For this set of simulations, supporting information is provided in its [supplementary information github page](#) [101], for each run, input files of all simulation steps are

provided, in the case of the last thermalization step, the RMSD of all protein atoms is also provided, as a proof that extending this step resulted in properly thermalized systems<sup>1</sup>. In the case of the production runs, together with the final snapshot of each replica a plot with the RMSF of each amino acid throughout the simulation can be found, in those plots the different domains have been projected on the background for ease the visualization<sup>2</sup>, also RMSD of the membrane domain, CaM and CTD can be found<sup>3</sup>. It is interesting that in general, the CTDs of the KCNQ2 channels have more mobility than the SK2 ones, mainly because swimming movement of helix hC. In RMSD and RMSF plots of the production runs the error plotted is the standard deviation from the four monomers of the channels, it has to be underlined that due to the symmetry of the channels, magnitudes will be computed for each monomer, as each system is simulated by triplicate, essentially, we will have 12 monomers simulated for 500 ns each.

### 5.3 Riluzole 's pocket characterization

A detailed characterization of the binding pocket of riluzole will be now presented, from the three drugs studied, it is the only one that it does not have a full channel structure resolved experimentally, so it is worth to make a detailed characterization of the pocket obtained from the simulations.

The first thing analyzed is that if riluzole indeed stays stable in its pocket, to do so the root mean square fluctuation (RMSF) was computed, arising a value of  $(0.19 \pm 0.03) nm$ , this is that in average, the standard deviation of the positions of all atoms belonging to the drug is  $0.19 nm^4$ , which is a typical value due to thermal fluctuations, the error of the magnitude comes from the standard deviation upon all four riluzoles that a channel can admit of the three runs performed. After the stability of drug in the pocket was settled we can define the elements of the pockets, namely the linker S4S5a and the N lobe of CaM (CaM N lobe), a snapshot of the pocket is showed in (Fig. 5.5, A) with each  $\alpha$ -helix labeled, besides, each residue is colored with the contact probability with riluzole, the numerical values of these probabilities together with their standard deviation upon monomers and runs are provided in (Table 5.1). Some of these residues of CaM that interacts with Riluzole have been stated to suffer side chain rearrangements when bound to the SK2 CaMBD [94], namely (Leu-32, Met-51, Glu-54, and Met-71).

In order to investigate if the presence of the drug alters the behavior of residues

<sup>1</sup>Follow the links for RMSD during last equilibration step for [KCNQ2\\_APO](#) [KCNQ2\\_RTG](#) [KCNQ2\\_ZTZ](#) [SK4\\_APO](#) [SK4\\_RIL](#)

<sup>2</sup>Follow the links for RMSF during production for KCNQ2\_APO([1,2,3](#)) KCNQ2\_RTG([1, 2,3](#)) KCNQ2\_ZTZ([1,2,3](#)) SK4\_APO([1, 2,3](#)) SK4\_RIL([1,2,3](#))

<sup>3</sup>Follow the links for RMSD during production for KCNQ2\_APO([1,2,3](#)) KCNQ2\_RTG([1,2,3](#)) KCNQ2\_ZTZ([1,2,3](#)) SK4\_APO([1,2,3](#)) SK4\_RIL([1,2,3](#))

<sup>4</sup>Since the channel can diffuse through the membrane, center of mass translations were substracted from the RMSF calculations.

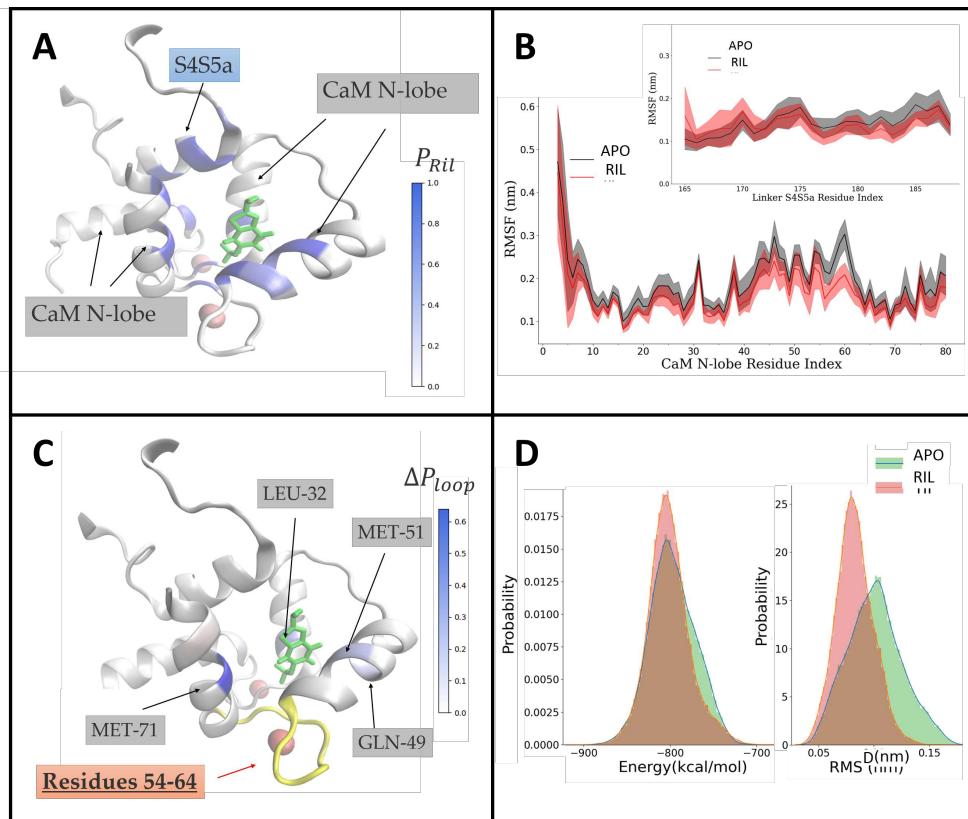
| Linker S4S5a |        |            | CaM N-lobe |       |            |
|--------------|--------|------------|------------|-------|------------|
| ResID        | P      | $\Delta P$ | ResID      | P     | $\Delta P$ |
| Ser-181      | 0.95   | 0.03       | Phe-19     | 0.99  | 0.02       |
| Ile-182      | 0.15   | 0.05       | Ile-27     | 0.9   | 0.2        |
| Ala-184      | 0.994  | 0.004      | Leu-32     | 0.9   | 0.2        |
| Leu-185      | 0.9998 | 0.0003     | Met-51     | 0.9   | 0.3        |
| Gln-187      | 0.7    | 0.2        | Ile-52     | 0.7   | 0.3        |
|              |        |            | Glu-54     | 0.6   | 0.2        |
|              |        |            | Val-55     | 0.99  | 0.02       |
|              |        |            | Ile-63     | 0.97  | 0.04       |
|              |        |            | Phe-68     | 0.99  | 0.02       |
|              |        |            | Met-71     | 0.999 | 0.008      |
|              |        |            | Met-72     | 0.5   | 0.2        |
|              |        |            | Lys-75     | 0.92  | 0.01       |

**Table 5.1:** List of residues that are in contact with riluzole more than 10% of the frames, chosen contact length is 4.5 Å

belonging to the pocket, we computed the root mean square fluctuation (RMSF) of each residue of CaM N lobe (Fig. 5.5, B) and linker S4S5a (Fig. 5.5, B - inset) both in presence and absence of the drug, as usual the error is the standard deviation between monomers and runs. For most amino acids of CaM N lobe and linker S4S5a the computed error bars of the RMSF of both situations overlap, indicating that the appearance of the drug does not carry any substantial differences, however, for residues 54 to 64 of the CaM N lobe we see an statistically significant reduction of their mobility, it turns out that this segment is very relevant as it is one of the loops that can accommodate a calcium ion in the starting structure and throughout the simulation (Fig. 5.5, C - yellow). To have an alternative description to validate this difference in mobility, we computed the contact probability of all amino acids of the pocket with the loop,  $P_{loop}$ , the difference between the probabilities of apo runs and in presence of the drug  $\Delta P_{loop} = P_{loop}^{APO} - P_{loop}^{RIL}$  is showed in (Fig. 5.5, C), there were not any residue with more contact probability with the loop in presence of Riluzole and residues Leu-32, Gln-49, Met-51 and Met-71 yielded values of  $\Delta P(Leu-32) = 0.4 \pm 0.3$ ,  $\Delta P(Gln-49) = 0.2 \pm 0.3$ ,  $\Delta P(Met-51) = 0.2 \pm 0.3$  and  $\Delta P(Met-71) = 0.6 \pm 0.2$ , meaning that in runs with the drug present the loop interacts less with these residues, which can be interpreted as the loop being less flexible to make the contacts that usually does in absence of the drug.

We wanted to correlate differences in flexibility with other quantities that would allow us to state a hypothesis of the effect of the drug in the channel, with that objective we computed the interacting energy of the  $Ca^{2+}$  kept in the loop under study with all protein residues and Riluzole (in the runs that is present) for all time steps of all simulations, by means of the NAMD energy plugin [12] implemented in VMD [30] with CHARMM36 force field [18], a histogram of the computed energies

## Riluzole's pocket characterization



**Figure 5.5:** Characterization of the binding pocket of Riluzole in SK4 channel. **A**, Contact probability of Riluzole (green) within its binding pocket, helices belonging to CaM N lobe or linker S4S5a are labeled. **B**, RMSF of all amino acids belonging to the binding pocket depicted in **A**, where the linker S4S5a is showed as an insight, values of apo runs are connected by black lines and Riluzole values by red lines, standard deviation between the four monomers of each of the three runs is showed as a dashed fill. **C**, difference of interaction probability of residues 54-64 (yellow) with the rest of the binding pocket,  $\Delta P_{loop} = P_{loop}^{APO} - P_{loop}^{RIL}$ , so that blue color means that the residue interacts more with the yellow loop in apo runs. **D**, Histograms of the interaction energy of  $\text{Ca}^{2+}$  with nearby protein residues or Riluzole if present (left) and the RMSD with respect to the minimum energy frame of apo and Ril runs (right), information of all monomers of all runs were used to construct the histogram.

are showed in (Fig. 5.5, D - left), it is interesting that runs in which riluzole was present the histogram presents a Gaussian like distribution with a minor shoulder around high energies ( $\sim -750 \text{ kcal/mol}$ ), this shape is expected as it is dominated by the thermal fluctuation of the system, in contrast with the apo analogous, in which a bigger shoulder is found around  $\sim -750 \text{ kcal/mol}$ , the increment of probability for higher energy terms could be revealing and would imply that the ion would be more stable in presence of riluzole. Finally, we computed the root mean square deviation

(RMSD) of the atoms that are less than 5Å away from the ion and depicted it as an histogram in (Fig. 5.5, D - right), using the less energy state from the three runs of each system to set the reference for the RMSD calculation, as before, there is a marked shoulder around high deviations for apo, to connect the two ideas that we have introduced, this is that riluzole decreases the mobility of the loop and increases energetical stability, we compute the covariance of the energy and the RMSD, defined as:

$$cov(E, RMSD) = \sum_j^{monomers} \sum_i^{frames} \frac{<(E_{ij} - \bar{E})(RMSD_{ij} - \bar{RMSD})>}{\Delta E \Delta RMSD} \quad (5.1)$$

yielding a result of 0.52 for APO and 0.32 for ril, supporting the idea that there is a connection between calcium stability and mobility of the loop, the smaller correlation for Riluzole can be another symptom that the thermal effect is more relevant in presence of the drug. This results also suggest that the molecular mechanism of the drug would be to stabilize the loop in order to enhance the stability of the calcium ion held, since Riluzole acts as a channel activator by enhancing its affinity for calcium ions [94]. This hypothesis is consistent with the fact that amino acids Gly-25, Gly-61, Gly-98, and Gly-134, each belonging to one CaM loop that holds an calcium ion, were experimentally found to be perturbed in the complex formed by [CaM-SK2 CaMBD-Riluzole] (PDB ID: 5V02) [94], making the authors think that these glycines act as Riluzole-mediated hinges that make subtle changes increasing  $\text{Ca}^{2+}$  affinity.

## 5.4 Network Analysis

### 5.4.1 Introduction

With the aim of studying allosteric regulations within proteins (i. e. long range interactions), graph theory has been drawing an increasing attention over the last two decades [102–105]. Graph theory is a mathematical theory that consists in extract information out of a graph, which is a set of nodes connected by edges, these nodes will usually have structural features of the system topology, for example they can be the  $\alpha$ -carbon of one amino acid, notice that this definition of the node implies a simplification of the system as all information corresponding to one amino acid is contained in one node, as in the case of coarse grained force fields, finally, edges will take into account the connection between the nodes of our system. These connections are translated into a square matrix, usually called the adjacency matrix  $A_{ij}$ , whose matrix elements are equal to zero if nodes  $i$  and  $j$  are not connected, and a positive value otherwise, so that  $A_{ij}$  accounts for the strength of connections between nodes but also if they are connected at all.

The main reason of introducing graph theory to the analysis of our simulations is the concept of node centrality, essentially, it assesses the importance of the nodes within the network, several ways of defining centrality have been proposed [106] but

when applied to allosteric regulation the common method is to make use of betweenness centrality, which has been found to distinguish relevant residues in allosteric signaling [107–109]. Betweenness centrality, also known as geodesic centrality, takes into account the fraction of all possible shortest paths connecting any pair of nodes of our system that crosses a given node, this magnitude provides information of which nodes are more important for general propagation of signals, a node that has low betweenness centrality will be in the outskirts of the topology and will be less involved in transmitting any signal. Therefore, centrality measures are capable of translate all the information contained in the edges of the graph into a single number for each node, reducing drastically the amount of data that will be analyzed.

There are some systems in which the start and end of the signal is known and the interest lies in how this signal is transmitted, this is the case of ionic channels, in which the pore is allosterically connected to the C lobe of the calmodulin, which acts as a  $Ca^{2+}$  sensor, or the voltage sensor domain, since the pore reacts to variations of cytosolic calcium concentration and to membrane potential variations. In this case, the predetermination of the starting and ending nodes of the signal changes the approach followed, we will be interested not only on the shortest paths but in all nodes that might have a contribution for propagating the signal, otherwise we might have instabilities when applying the method to molecular dynamics as different replicas might yield different shortest paths and there might also be fluctuations over time.

Recently, current-flow betweenness centrality (CFB), a new methodology for computing betweenness centralities has been introduced [109], in this framework the start and end of the signal need to be specified and its foundations are based on the analysis of electrical circuits [110]. Additionally, it has been proved that this measure is the asymptotic limit of a random walk based method [111], the advantages of this method with respect to geodesic shortest path is that it has been proven to be more robust, computationally less expensive and it is a cut-off free method [109], being the latter an improvement with respect to suboptimal pathway calculations [112].

Given that graph theory only requires a set of nodes and the connections between them, the abstraction of the protein into a graph can be very varied, as reviewed in [104]; for proteins, most works consider that each node is an amino acid, but for the weights of the connections there is no consensus in the definition, some make use of unweighted graphs (all edges are equal to 1), others the number of atoms that are in contact between the two nodes, the inverse of the averaged distance, the average contact energy, among others. Finally, Network analysis will be the application of graph theory to compute the current-flow betweenness centrality of the nodes of our system, in the following paragraphs we will see the exact definition of our graph and the mathematical foundations behind the calculations.

### 5.4.2 Definition of the network

For the definition of our network, we followed the methodology presented in [93, 113], each amino acid refers to one node and lipids will be separated into three (i. e. one

node for the hydrophilic head and two for the hydrophobic tails), for the adjacency matrix, the matrix elements will be the product of the contact map and the mutual information of the fluctuations of the positions of the pair of nodes, so that both structural and dynamical information of our simulated system is introduced into the network.

The contact map will have structural information of our system and will help us to discard each pair of non-interacting nodes, making the calculation of the graph easier, to determine it we will not use a binary value, for distances lower than the cutoff  $c = 4.5 \text{ \AA}$ , the probability will be set to one but for bigger distances a smoother probability distribution  $K(d)$  will be used so that the probability goes smoothly to zero, as it has been proven to be more reliable [109]:

$$K(d_{ij}) = \frac{\exp(-\frac{d_{ij}^2}{2\sigma^2})}{\exp(-\frac{c^2}{2\sigma^2})} \quad (5.2)$$

where  $d_{ij}$  is the distance between atoms  $i$  and  $j$ , this distribution is used until a upper cutoff of  $8 \text{ \AA}$ , the standard deviation of the gaussian  $\sigma$  is  $1.38 \text{ \AA}$ . All the dynamical information of the simulation is translated into the contact map as the value of  $K(d_{ij})$  is averaged over all time steps, since we are performing this average, the appropriate time scale for these technique will be around the hundreds of nanoseconds, as no major conformational changes are sought but statistical robustness is required, otherwise we might obtain unrealistic contact maps.

The contact map determines the connections that appear in our network, but we also want to add the information of the correlation between the fluctuations of each pair of nodes, this is achieved by computing the mutual information matrix of the atomic fluctuations. The mutual information of two statistical variables contains the information you can obtain from one variable by knowing the other one, in other words, how correlated are the two variables. In our case, we will compute the matrix of mutual information  $M_{ij}$ , for two nodes (or amino acids)  $i$  and  $j$ :

$$M_{ij}(x, y) = \int \rho_{ij}(x, y) \ln \frac{\rho_{ij}(x, y)}{\rho_i(x)\rho_j(y)} dx dy \quad (5.3)$$

where  $\rho_i(x)$  and  $\rho_j(y)$  are the probability density distributions of atoms  $i$  and  $j$  fluctuating by  $x$  and  $y$  around their resting positions, so using  $x$  and  $y$  is a compact notation for displacements  $\mathbf{x} - \bar{\mathbf{x}}$  and  $\mathbf{y} - \bar{\mathbf{y}}$ ; similarly  $\rho_{ij}(x, y)$  is the joint probability density, this is the probability that atom  $i$  has a fluctuation  $x$  and atom  $j$  has a fluctuation  $y$ . Mutual information is defined as a positive value, notice that in particular, if variables  $x$  and  $y$  are independent the joint probability distribution is equal to the product of the marginal distributions  $\rho_{ij}(x, y) = \rho_i(x)\rho_j(y)$ , defining the lower bound of the mutual information as zero, this is when there is not correlation between the variables. Mutual information can be written in terms of the statistical entropies ( $H$ ), of the variables :

$$M_{ij}(x, y) = H_i(x) + H_j(y) - H_{ij}(x, y) \quad (5.4)$$

Where the entropies are defined as  $H_i(x) = -\int \rho_i(x) \ln \rho_i(x) dx$ , this relation is important as in practice the mutual information will be computed in terms of the entropies, as

it is computationally less costly, its mathematical derivation is attached as part of [Appendix A](#), further information on how (Eq. 5.4) is computed in practice can be found in [113]. Finally, the matrix network will be defined as the product of the averaged contact map and the mutual information  $K(d_{ij})M_{ij}$  for nodes  $i \neq j$  and 0 for the diagonal elements.

### 5.4.3 Current flow betweenness centrality

To characterize the allosteric pathways in our system we will compute the Current Flow Betweenness centrality (CFB), it will help us illustrate which amino acids are more important to transmit a given signal. The main idea is that we are computing all possible paths connecting source and sinks, characterizing how much information flows through each node and pouring it into the centrality measure, in [Appendix A](#) a thorough derivation of its mathematical expression [110] can be found, here we just write its expression:

$$\tau_i = \frac{1}{2|S_0|} \sum_{s,j} |p_i(s) - p_j(s)| A_{ij} \quad (5.5)$$

in sort, we compute the information current passing through node  $i$ , coming from all its adjacent nodes  $j$ , and this is repeated for all sources  $s$ , so the final value is normalized by the total number of sources  $S_0$ , note that the information flow is treated as an electrical current so that the previous equation is like the Ohm's law in which the electrical current is the potential differences between nodes times the resistivity (this is, the inverse of the resistance).

As introduced before, in the particular case of ion channels, the sources will be the gating charges of the voltage sensor domain in the case of KV7.2, or the C lobe of the calmodulin for both studied channels, since SK4 channel is voltage insensitive [27]; whereas the sink will always be the pore, following reference [93]. The identification of these residues for the channel KV7.2 was carried out following [31], and [72] for SK4 channel, residues can be found at Table 5.2. In the following section we will start by reviewing how the results are provided, a more detailed explanation of the calculations and the code used can be found in [114].

**Table 5.2:** Residues used to define the sources and sinks of the studied signals.

| Sources            | KV7.2                    | SK4       |
|--------------------|--------------------------|-----------|
| VSD gating charges | 198, 201, 207, 210 (Arg) | -         |
| CaM C lobe         | 80-148                   | 80-148    |
| Sinks              | KV7.2                    | SK4       |
| Pore               | 314 (Ser)                | 282 (Val) |

## 5.5 Current flow betweenness centrality calculations

The results presented in this section consist in the calculations of the current flow betweenness centrality (CFB) for all simulations, for both Kv7.2 and SK4 channels. Before we start let's comment how this magnitude is represented in the following figures. Since channels have four identical monomers, the signal will be defined to begin at the source residues of the four monomers and to end at the four pore residues, from the definition of the supply vector (see [Appendix A](#)), each term of the sum that is made in (Eq. 5.5) comprises one source residue and all sinks, so when we are computing the signal starting in the CaM C lobe for example, we are averaging between the signals starting in all its residues of each monomer and all arriving to the four pore residues.

The calculation of a signal will then give the CFB of all amino acids of the tetramer, but for simplicity in the posterior depicting these are averaged between the monomers so that only one monomer is presented in the plots.

It also stands out that the pore residues have a zero value of CFB in all plots, as stated in the mathematical derivation of [Appendix A](#), for a CFB calculation both pore and sinks are given a value of zero, and the pore is the same in all terms of (Eq. 5.5) so its value is always zero when averaging between sources and monomers. Conversely, each term of (Eq. 5.5) refers to a signal starting in that source and we set that value to zero, but different terms referring to nearby sources will yield a CFB different from zero, so that when doing the averages the total CFB is different from zero.

Finally, we will also make the averages between the three different replicas that we have for each system. Therefore, the depicted errors as transparent areas of color will be the standard deviation between the different runs and monomers.

As we can see in figure (Fig. 5.6, A), we represented the current flow betweenness centrality of all amino acids of the channel and calmodulin, ordered by sequence, for clarity, we have joined the CFBs of the different residues with solid lines, and also we have illustrated the different domains by coloring the background, adopting the following color code: transmembrane helices S1-S6 are in blue, intracellular helices hA-hB-hC are in yellow and calmodulin lobes are in green. Before we comment the results let's make some comments on the general features of these graphs for the sake of clarity, taking advantage that the mathematical derivation of the magnitude is presented in [Appendix A](#).

### 5.5.1 General features

CFB can be seen as the propagation of an information current, where this current has been modeled using circuit theory, therefore we used current conservation laws in the mathematical derivation, meaning that the current that flows into one node is equal to one flowing out of it, so that we have to take into account that nodes located near an important node will benefit from it and necessarily will have greater centralities, in other words, the magnitude of the centrality also accounts for the location of the node. One example of this is that in all plots, residues near the sources or sinks have in general high centrality, this is because the information flows from sources to sinks and is conservative, so it will spread from the sources and will gather around the sinks,

## Current flow betweenness centrality calculations

---

increasing the centrality of the nodes nearby, this introduces an overestimation near the sources and sinks and therefore it has to be taken into account, unfortunately, it is very difficult to estimate a cut-off from which this effect is negligible.

A common feature of the graphs is that in general, the CFB is higher in amino acids belonging to helices, this can be explained as the secondary structure makes the motion of the amino acids to be more correlate, making them excellent transmitters of information flow, conversely, non structured linkers have more freedom to fluctuate resulting (in general) in a lower mutual information with the surrounding residues, blocking the information flow signal.

Another interesting aspect is that in helices the graphs tend to have a saw-tooth like shape, if we look closely it can be noticed that it has the same periodicity of the helical turns, as in this case the information follows two paths, it travels following the sequence but also straightly jumping through the hydrogen bonds, the latter effect is the one responsible of the saw-tooth like shape.

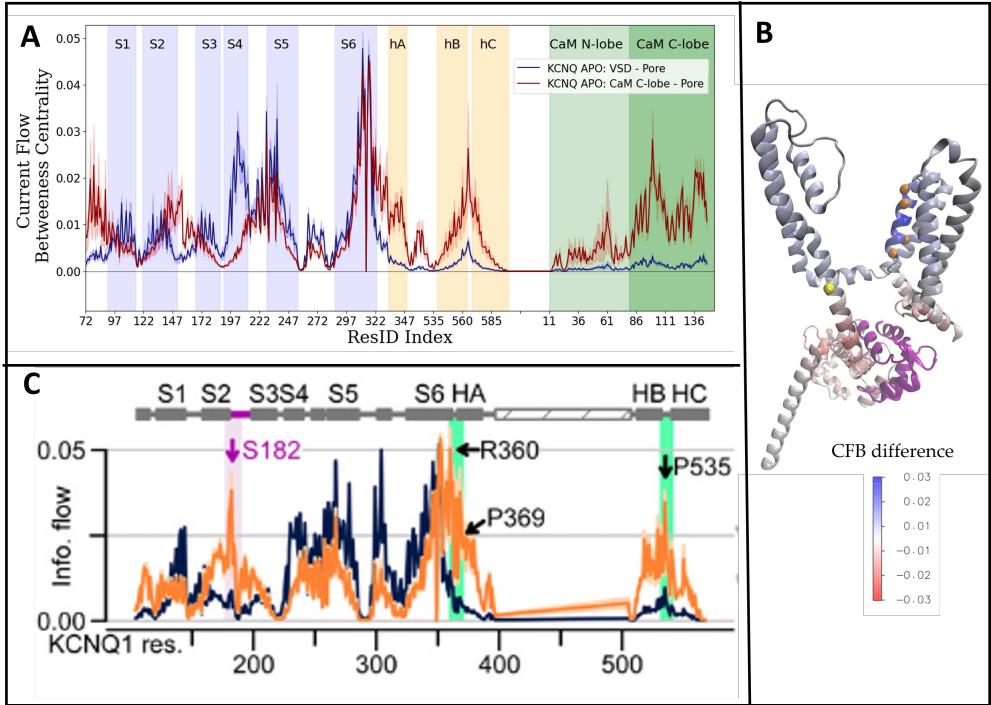
In general, we will compare the CFB obtained with the ones of the control systems, since it will be difficult to rationalize a given value of CFB and also the plots are not easy to read.

After this introduction of the calculations of the CFB using the network analysis methodology, we are in position to present the obtained results.

### 5.5.2 Kv7.2 Apo

Let's analyze now the results of the CFBs taking into account the general comments made previously. To begin with, we have represented in the case of the Kv7.2 channel in absence of drugs, the signals starting in the gating charges of the VSD (Fig. 5.6, A: blue) and in the CaM C lobe (Fig. 5.6, B: red), looking at the first one, we can have a feeling of how the signal is transmitted from the VSD gating charges to the pore; firstly, there is a great hub of centrality in helices S4 and S6, as sources are located in S4 and the sink in S6, also S5 exhibits a high centrality, and also S1, S2 and S3, but to a lesser extent, it is very important to know the three dimensional arrangement of these helices and the positions of the gating charges to conclude that, roughly speaking, the flow follows mainly three paths: through S5', where the accent is because it belongs to the contiguous monomer (arginine 198 and 201 are in contact with S5'), through S4S5 linker (arginine 207 and 210 are closer to the S4S5 linker) and with a lesser extent through S1 (arginine 201 faces also S1 helix).

In the case of the signal flowing from the CaM C lobe to the pore (Fig. 5.6, B: red), it can be noticed that the centralities differ considerably with respect to the VSD-pore signal; as expected, there are centrality hubs in the CaM C lobe (source) and near the pore (sinks), however this time the centrality of the sink increases from S6 and from hA helix, as there is part of the signal traveling through the cytosolic domain now, while the rest goes through the transmembrane domain as in the previous case. Looking at the cytosolic domain, there are two big hubs of centrality, the one would be helix A together with its linker to the pore domain, and the linker between hB and hC, which also lies near the pore.



**Figure 5.6:** Results obtained from the Network analysis for the Kv7.2 apo runs. **A**, mean CFB obtained for each residue for all apo runs, for the signal starting in the VSD gating charges and finalizing in the pore (blue) and from the CaM C lobe to the pore (Red), standard deviation upon different runs and monomers is introduced as dashed fills. The domains of the channel are colored in the background. **B**, Difference between the CFB of the VSD-pore signal minus the CaM C lobe-pore signal, the gating charges are depicted as orange balls, the pore as a yellow one and the CaM C lobe is colored in purple. **C**, data extracted from [93] in which the authors made the same calculations as in **A** but for the Kv7.1 channel, some centrality hubs are emphasized, at the top it is showed the domain to which each amino acid belongs.

To elucidate the disparity of the two signals, we can compute their difference for each amino acid; results of the CFB of the VSD-pore signal minus the CaM C lobe one are showed in (Fig. 5.6, B), in this case the results have been projected into the three dimensional structure of a monomer, where it is clear that all cytosolic domain is influenced mostly by calmodulin and the transmembrane domain by the VSD charges, there are some exceptions in the later though, like residues before S1 and the region near the S2S3 linker which lie near CaM.

Our results are similar of those published in [93] for the channel Kv7.1 (Fig. 5.6, C), whose structure is very similar to the one of Kv7.2 [10], in fact all the domains S1-

## Current flow betweenness centrality calculations

---

S6 and hA-hC are conserved, note that in (Fig. 5.6, C) the centralities of CaM residues are not showed. When comparing our centralities with theirs, it stands out that their pore helix, a small helix that lies between S5 and S6 in the sequence, has a bigger centrality, both when the signal starts in the VSD or CaM C lobe, which could be an indicator that the pore helix is more important for the channel Kv7.1; furthermore, in the case of the signal starting from the CaM C lobe, they observed a high peak in the linker S2-S3, which lie very close to CaM C lobe, its absence in the Kv7.2 and the fact that our centralities in the residues before the S1 helix are higher, which also lie near CaM C lobe, might be an indicator that the communication between CaM and the transmembrane domain could be different in the channels, however this does not seem easy to check experimentally and the cause could just be that the experimental structures of the two channels used are in slightly different conformations.

### 5.5.3 drugs

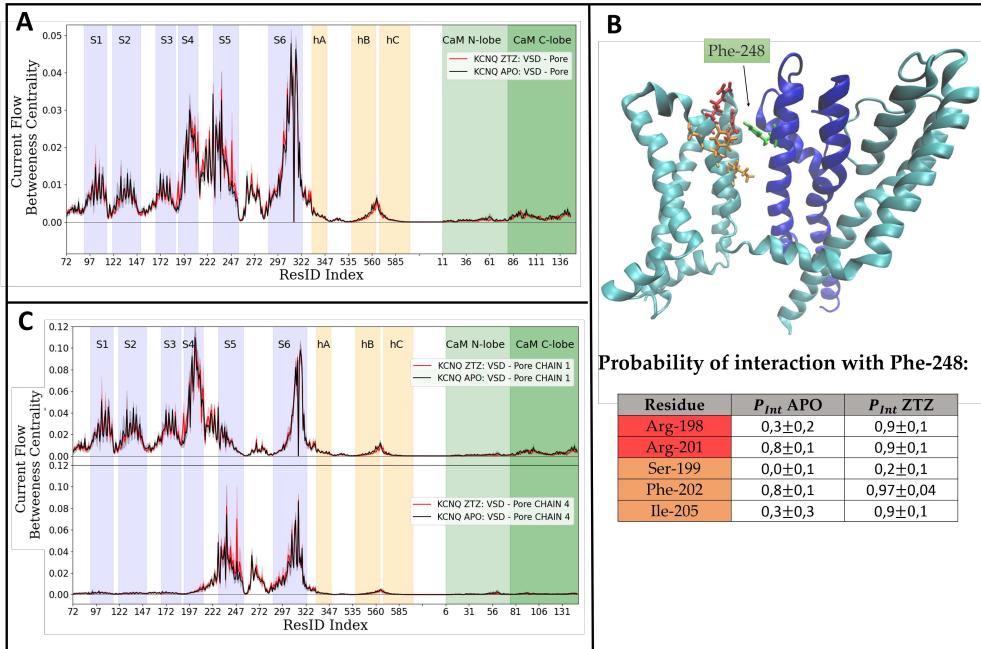
Continuing with the results obtained, we can comment how the different drugs affect the current flow betweenness centrality of the channels, as introduced before, one of the key aspects of the selected drugs is the variety of the location of the binding pockets (Fig. 5.1-5.2), so that presumably the effect of the drugs should be different. Let's analyze the effect of the different drugs separately.

#### ZTZ240

ZTZ binding pocket is situated between the four helices that constitute the VSD (Fig. 5.1, C and D). When the signal starts in the VSD gating charges, it has a CFB of  $0.016 \pm 0.004$ , which is similar as residues belonging to S1, S2 or S3 helices, all centralities for this signal are depicted in (Fig. 5.7, A), in general when comparing with the apo curve we can see that the plots are very similar, but there are many small differences in the height of some peaks, the one that we found that could be the most important is the one at Phe-248, the gain of betweenness of this residue in presence of the drug could be an indicator that the drug enhances the communication with the pore through this amino acid, to elucidate this we computed the contact probability of Phe-248 with the nearby residues, observing a noticeable increment with several residues in the VSD in presence of the drug when compared to apo, they are depicted in (Fig. 5.7, B), it stands out that there is an increment with the Arg-198 and Arg-201, which are two of the four gating charges, but also with three more residues of S4 helix, namely Phe-202 and Ile-205, this results suggest that the appearance of ZTZ implies a small rearrangement of the VSD so that the communication with S5' is enhanced through the residue Phe-248, which would serve as a bridge connecting the two domains.

The importance of the communication between S4 and S5' can be illustrated if, instead of putting the sources of the four monomers in the sum of (Eq. 5.5), we do them separately, and assign the name of chain 1 to the monomer in which the signal is originated and chain 4 the adjacent one, chains 2 and 3 will not be represented as their amino acids will almost do not exhibit CFB, except from amino acids near the

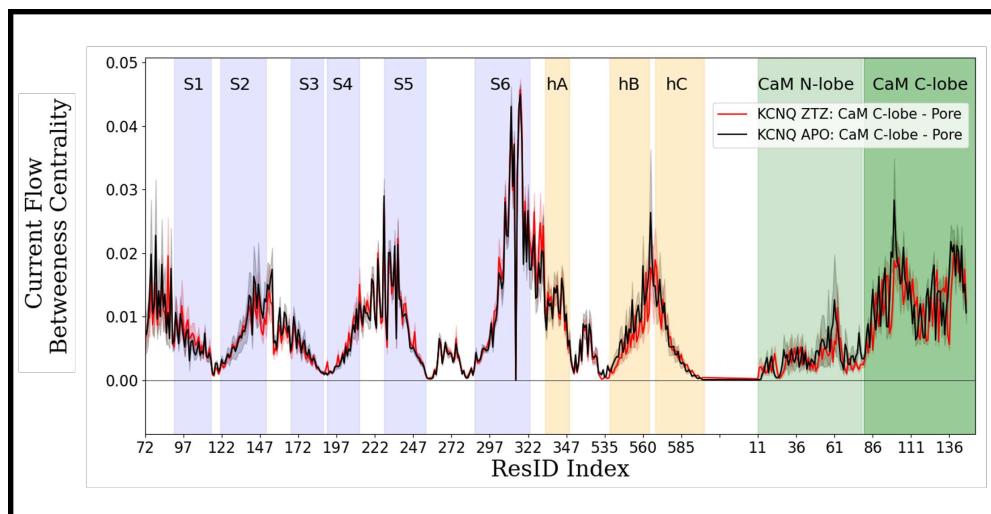
pore, as they lie far from chain 1; therefore we can have, as before, 12 signals to obtain the standard deviation but still have the information of how the signals jump between monomers. As can be seen in (Fig. 5.7, C), much information that came from the VSD of chain 1 (cyan in Fig. 5.7, B) jumps to chain 4 (dark blue in Fig. 5.7, B), this is expected since this channel exhibits a domain swapped conformation, which means that its monomers are interlaced.



**Figure 5.7:** Results obtained from the Network analysis for the Kv7.2 simulations in presence of ZTZ240 drug. **A**, results of CFB of the VSD-pore signal in presence of the drug ZTZ240 (red) compared with the apo runs (black). **B**, visualization of amino acid Phe-248 (green) belonging to S5' of the chain 4 (adjacent monomer, in blue), and some of the interacting residues of S4 of chain A (cyan), namely Ser-199, Phe-202 and Ile-205 (orange) and the gating charges Arg-198 and Arg-201 (red) with a table comparing the probability interactions of the residues with Phe-248 of S5' in apo and in presence of the drug. **C**, breakdown of the CFB by chains, for this calculation the CFB has not been averaged over the monomers, chain A and D refer to the cyan and blue monomers in **B**, respectively.

Going back to the standard type of plot for the signals, in the case that it flows from the CaM C lobe to the pore we do not observe major changes with respect to APO to comment (Fig 5.8), in this case, the CFB of the drug is  $0.007 \pm 0.003$ , less than the previous situation, this indicates that the drug would be more involved in the signal that starts at the VSD, which can be related to the position of its binding pocket.

## Current flow betweenness centrality calculations



**Figure 5.8:** Results of the CFB of the amino acids of the sequence of the channel Kv7.2 for the CaM C lobe-pore signal in presence of ZTZ240 (red) compared to the apo runs (black).

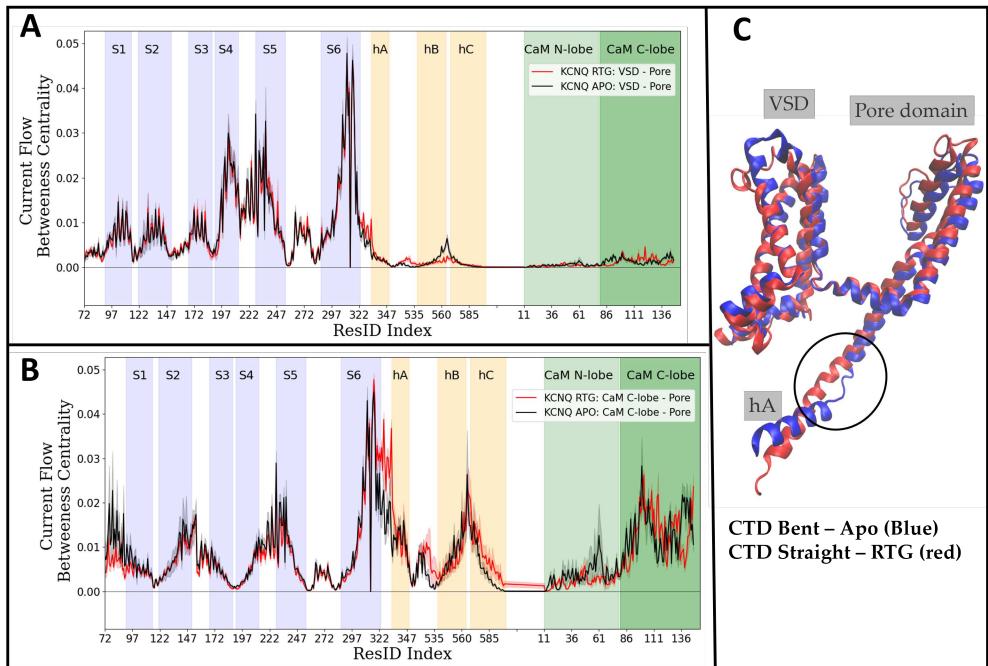
## Retigabine

The binding pocket of RTG is located next to helix S6 and facing the membrane (Fig. 5.1, A and B), looking at the results, retigabine shows a CFB of  $0.022 \pm 0.008$  for the signal starting at the VSD and  $0.015 \pm 0.005$  when starting at the CaM C lobe. In the case of the VSD-pore signal Fig 5.9 A), we do not observe any difference in the transmembrane domain when compared with the apo runs, and the cytosolic domain has low CFB indicating that in this case almost all information goes through the membrane. Conversely, there are more differences between apo and RTG drugs when the signal starts in the CaM C lobe (Fig 5.9 B), there is a CFB reduction in helix S1 and the previous residues and an enhancement in the cytosolic domain, this would indicate that there is less communication with CaM and for that reason more information would flow through the cytosolic domain.

To rationalize this results we have to comment a peculiarity of the structure of the channel in presence of RTG, this is that exhibits what is called a CTD-straight conformation, where CTD stands for C terminal domain (Fig 5.9 C: red) in which the S6-hA linker is different with respect to the apo and ZTZ structures (Fig 5.9 C: blue), in presence of RTG this linker does not longer exist as helices S6 and hA are united into one helix that no longer forms any angle (CTD straight), conversely, in apo and ZTZ structure this linker has no secondary structure and form an angle (CTD bent), this structural change has other peculiarities as the rearrangement that has to take place has as a consequence the separation of calmodulin from the VSD. This would explain then the differences obtained in (Fig 5.9 B), as CaM is further away from the VSD it does not interact with the residues before S1 lowering its centrality and the residues between hA and S6 exhibits an increase of their centrality due to their

different secondary structure.

It is difficult to obtain any insight of the effect that has this drug on the channel, aside from the conformational changes observed in the experimental structure. Taking a different perspective, we can state that the technique is able to detect conformational changes, even though in this case it was obvious from visualizing the structures.



**Figure 5.9:** Results of the network analysis for the Kv7.2 channel in presence and absence of Retigabine. **A**, CFB of the amino acids of the sequence in the apo (black curve) and RTG (red curve) run for the VSD-pore signal. **B**, CFB of the amino acids of the sequence in the apo run (black curve) and RTG run (red curve) for the signal starting in the CaM C lobe and ending in the pore. **C**, Three dimensional starting structure of the apo run (blue), which is in the CTD bent conformation, and the one of RTG (red), which is in the CTD straight one.

### Riluzole

In the case of Riluzole and SK4 channel, the modeled pocket according to the experimental information available, lies between the CaM N lobe and helicoidal linker S4S5a (Fig. 5.2), as stated before, SK4 channel have lost the sensitivity for potential differences [27], so only the CaM Clobe-Pore signal is studied, the CFB of Riluzole for this signal is  $0.03 \pm 0.02$ , which is quite big considering the centrality of the amino acids belonging to the pocket (Fig. 5.10, A), and it shows a bigger standard deviation than the previous drugs, this made us consider the robustness of the calculations upon different runs, which will be commented later on, but before we get to that let's comment

## Current flow betweenness centrality calculations

---

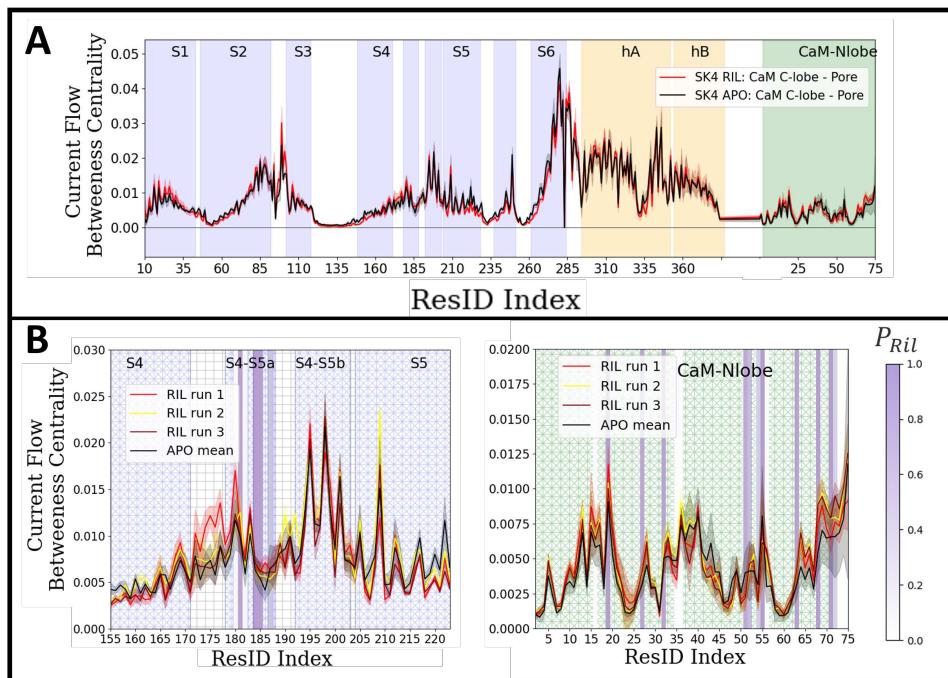
first the obtained centralities of (Fig. 5.10, A). Several residues in the pocket show a small increment of their CFB, some of them in the CaM N lobe and other two in the S4S5a and the linker connecting this helix to helix S4, but they are not statistically significant as their errors overlap with the ones of the apo runs. Additionally, there is an increment of CFB on the S2-S3 linker and in the hA helix, where the S2-S3 linker can be found in (Fig. 5.2, A) as it is the gray linker between the two helices that is very close to the dark blue one (which is hB of the monomer in front); also small reductions in all areas that lie in the transmembrane domain but closer to outside the cell can be found, for example amino acids near linkers S1-S2, S3-S4, S5-pore helix or pore helix-S6. These two characteristics could indicate that riluzole makes the information to flow following paths involving the cytosolic domain, in particular hA, and minimizing the contributions from amino acids near the outside environment, but this effect is not statistically significant.

In order to have a closer look on the amino acids belonging to the pocket and check the robustness of the calculations, we separated the calculated CFB of each of the three runs in presence of Riluzole and compare it to the mean apo value, it is projected in the background in color scale the contact probability of the residues belonging to S4S5a with Riluzole, if the probability is zero, the previous background labeling the domains of the sequence is kept, but using patterns to fill the surface. In the case of the S4S5a helix (Fig. 5.10, B: left) it can be seen that in general the Riluzole lines do not tend to overlap, meaning that the CFB computed at each run is different, in the case of the amino acids belonging to the CaM N lobe we obtained a similar result, the CFB values of different runs were generally different, in this case it is interesting that almost all amino acids that interact with Riluzole exhibit slightly higher CFB than the mean apo runs, but again it is not statistically significant.

### 5.5.4 Conclusions

In this chapter, molecular dynamics simulation were carried for the full channel structures of Kv7.2 and SK2, in presence and absence of some activators of the channels; the goal of the simulations was to see if it was possible to elucidate their atomic mechanisms that results in enhancing the channel functioning. Considering the distances of the binding pockets to the pore, it is clear that the drugs have to work allosterically, the huge amount of information obtained from a MD simulation makes it almost impossible to extract meaningful information of these mechanisms by mere visualization, unless a more sophisticated method is used. One candidate methodology for describing the pathways or signals and how these drugs affect them is network analysis, which was reviewed and applied in this chapter. The results obtained of the CFB can be regarded as the importance of a given node for a defined pair of hubs that define the sources and sinks of information flow. It has to be underlined that the calculations of the CFB takes into account all possible pathways connecting the sources and sinks and pours all this information into one value per residue, which is a mayor reduction on the amount of output data to be analyzed.

From the results obtained for the Kv7.2 in absence of drugs, the comparison with



**Figure 5.10:** Results of the network analysis for the SK4 channel in presence of Riluzole. **A**, CFB of the SK4 apo channel and in presence of Riluzole for the signal flowing from the CaM C lobe to the pore. **B**, insight of the CFB of the part of the binding pocket belonging to the channel, and belonging to the CaM N lobe; the CFB of each of the three Riluzole runs are separated and compared to the mean apo value, in the background the contact probability of the amino acids with Riluzole is showed.

the analogous channel K<sub>V</sub>7.1 raised apparent differences between their communication with CaM and in the importance of the pore helix. It is difficult however to translate this differences in CFB into measurable observables, but one possible explanation is that their experimental structures might be in slightly different conformations.

Further comparisons of our CFB of the apo channels with the ones obtained in presence of drugs were carried out. In the case of ZTZ240, the small conformational changes in the VSD provoked by the drug has as a result a higher CFB in Phe-248, also, the contact probabilities obtained suggest that this increment in CFB of this amino acids is because it serves as a bridge connecting the VSD gating charges with S5' (this is, helix S5 of the adjacent monomer), this could result in the enhancement of voltage sensitivity caused by the drug, however, further experimental should be carried out to confirm this hypothesis.

For the case of RTG, it was not possible to extract any information since the differences on the CFBs came from the different conformations of the starting structures. Since the binding pocket of RTG involves the membrane it could be that the effect

## **Current flow betweenness centrality calculations**

---

of the drug could have something to do with it, but it was not possible for us to extract any meaningful information of the CFBs of the membrane, even though they are explicitly included in the calculations.

In the case of Riluzule, the small differences obtained were not statistically significant, however, as there were small increments in several residues of the residues, we came up with the idea of computing the medium RMSF of the amino acids of the pockets and ended up with the pocket characterization of the Riluzole and its possible atomic mechanism that is presented at the begining of the chapter.

To conclude, this technique can be useful to detect small conformational changes or mechanisms from the simulations, but in our case the effects of the drugs were not clear, this is because we have to realize that even though the model used will be all-atom based, it implies several approximations and when comparing the simulations with and without drug we will essentially changing few tens of atoms (one node) out of three hundred thousands (hundreds of nodes), also maybe the time scales simulated are not enough for observing the conformational changes caused by the drugs.

# Appendix A - Mathematical derivations

## A.1 Mutual Information

This section is dedicated to prove the following relation:

$$M_{ij}(x, y) = H_i(x) + H_j(y) - H_{ij}(x, y) \quad (5.6)$$

In order to achieve that, we will make use of the definition of the joint probability distribution of two dependent variables:

$$\rho_{ij}(x, y) = \rho_i(x|y)\rho_j(y) \quad (5.7)$$

This relation is quite intuitive, and means that the joint probability of two events (for us that atom  $i$  has a fluctuation  $x$  and atom  $j$  has a fluctuation  $y$ ), is the probability of having the first event given that the second one has already happened, multiplied by the probability of this second event, if the two events are independent  $\rho_i(x|y) = \rho_i(x)$ , as the probability of the first event is independent of the second one. If we insert this relation in the definition of the mutual information (5.3) and separate the logarithms:

$$M_{ij} = \int \rho_{ij}(x, y)[\ln \rho_i(x|y) - \ln \rho_i(x)]dxdy \quad (5.8)$$

We can integrate the variable  $y$  in the second term, obtaining  $\int \rho_i \ln(\rho_i(x)) = H_i(x)$ . Inserting relation (5.7) in the first term:

$$\int \rho_{ij}(x, y) \ln(\rho_i(x|y)) = \int \rho_j(y) \rho_i(x|y) \ln(\rho_i(x|y)) dxdy = - \int \rho_j(y) dy H_i(x|y) \quad (5.9)$$

where  $\int \rho_j(y) dy = 1$ , as the probability is normalized. Therefore we have obtained that:

$$M_{ij}(x, y) = H_i(x) - H_i(x|y) \quad (5.10)$$

In relation (5.4), it appears the joint entropy of the two variables,  $H_{ij}(x, y)$ , similarly, we are going to relate this magnitude with the conditional and marginal entropies,

---

starting from the definition:

$$H_{ij}(x, y) = \int \rho_{ij}(x, y) \ln(\rho_{ij}(x, y)) dx dy = \int \rho_{ij}(x, y) \ln(\rho_i(x|y) \rho_j(y)) dx dy \quad (5.11)$$

In the last equality we have used again relation (5.7), separating the product inside the logarithm we have two terms  $\ln((\rho_i(x|y) \rho_j(y))) = \ln((\rho_i(x|y))) + \ln(\rho_j(y))$ ; where in the second we can integrate the variable  $x$  to obtain  $H_j(y)$  and in the first one we can use again (5.7):

$$\int \rho_{ij}(x, y) \ln \rho_i(x|y) dx dy = \int \rho_j(y) \rho_i(x|y) \ln(\rho_i(x|y)) dx dy = H_i(x|y) \quad (5.12)$$

Therefore, we have obtained the following relation:

$$H_{ij}(x, y) = H_j(y) + H_i(x|y) \Rightarrow H_i(x|y) = H_{ij}(x, y) - H_j(y) \quad (5.13)$$

Finally, if we substitute  $H_i(x|y)$  from (Eq. 5.13) in (Eq. 5.10), we finally obtain that:

$$M_{ij}(x, y) = H_i(x) + H_j(y) - H_{ij}(x, y) \quad (5.14)$$

## A.2 Current flow betweenness centrality

Previously, we have already defined our network as a matrix that has the information of the contacts and their mutual information, this will be the adjacency matrix of our graph  $A$ . The requirements for  $A$  is that the graph must be connected and the weights of the edges must be positive, this is our case as the mutual information and the contact map are positive magnitudes. From  $A$  we can define the Laplacian of the network:

$$L = D - A \quad (5.15)$$

Where  $D$  is the degree matrix, a diagonal matrix formed by summing the rows of  $A$ . The Laplacian of the graph has the information of the degree of the nodes in the diagonal, this tell us how well connected this vertex is, and also the connections that makes with their respective weights.

As stated before, one of the characteristics of the network analysis is that we have to define where the signal originates and where it ends before we start our calculations, this will be done mathematically using a supply vector,  $b(v)$ , that is 1 if the vertex  $v$  is a source, -1 if it is a sink and 0 otherwise.

To introduce the current flow of the graph we have to provide our edges with an orientation, this is needed since in the end we want to account for the directionality of signals or perturbations within our network, one orientated edge will be denoted as  $\vec{e}$ , with coordinates  $\vec{e} = (u, v)$ , which mean that the edge  $\vec{e}$  connects nodes  $u$  and  $v$ , its direction is set arbitrarily. The set of all oriented edges is called  $\vec{E}$ . The previous definition of the supply vector together with our adjacency matrix *univocally* determines the current flow of the graph  $x(\vec{e})$ . The current flow takes as input one orientated

edge and gives a positive value if the current goes in the direction of  $\vec{e}$  and negative if it goes against it, the value of  $x(\vec{e})$  will have the information that describes how the signal is transmitted throughout our graph. This approach is based on electric circuits, the adjacency matrix has the information of the values of the electrical conductances connecting points in the circuit and the supply vector tells us where the current starts and ends, therefore it is clear that once we have our circuit defined with the electrical resistances and the electrical source (and sink) the electrical current that flows through the circuit is known, for us the current has the information of how easily a signal is transmitted in our protein, according to the properties that we have projected in our adjacency matrix.

The current flow is conserved throughout our graph aside from the sources or sinks and there is not any external current entering the circuit aside from the one defined in the supply vector, therefore Kirchhoff's laws will be followed, these are that the sum of the current of orientated nodes forming a cycle is zero and that the sum of the currents that flow into one node and the ones that flows out of that node is defined by the supply vector, mathematically:

$$\sum_{(v,w) \in \vec{E}} x(v,w) - \sum_{(u,v) \in \vec{E}} x(u,v) = b(v) \quad (5.16)$$

where the pairs  $(v,w)$  and  $(u,v)$  are the coordinates of oriented edges belonging  $\vec{E}$ , with  $b(v) = 0$  for vertices not defined as sources or sinks. Continuing with the electrical jargon, we will use Ohm's law to define the potential difference of a vertex:

$$\hat{p}(\vec{e}) = \frac{x(\vec{e})}{c(e)} \quad (5.17)$$

where  $c(e)$  are the weights of the edges <sup>5</sup> that appear in the adjacency matrix (and therefore in the Laplacian). We can come back to the non-directional point of view to define the *absolute potentials*,  $p(v)$ , a vector that gathers the potential differences of the nodes up to a common constant, with this definition we are projecting the directional information provided by the current flow in the absolute potentials, and we are not loosing any information since the current will flow from the higher potential values to the lower ones.

The potentials are very useful as they have the same information of the current flow, furthermore they connect the Laplacian to the supply vector:

$$Lp = b \quad (5.18)$$

To prove this relation, we will start with the left part of the equation, doing matrix product of the first row of the Laplacian with the potentials, which has the information

---

<sup>5</sup>note that in the definition of the edge weights  $c(e)$ ,  $e$  is also defined by two coordinates but it is not oriented and it does not have the vector sign, as this quantities do not have directionality.

---

of the first node:

$$\sum_{i=2}^N c(1,i)p(1) - c(1,2)p(2) - \dots - c(1,N)p(N) \quad (5.19)$$

We recall that the sum is the degree of the node, which is the diagonal of the Laplacian, expanding the sum and grouping:

$$c(1,2)[p(1) - p(2)] + c(1,3)[p(1) - p(3)] + \dots + c(1,N)[p(1) - p(N)] \quad (5.20)$$

where the potential difference  $p(1) - p(2)$  is the previously called  $\hat{p}(\vec{e}) = \hat{p}(1,2)$ , using Ohm's law ( Eq. 5.20) results in the sum of all currents passing through node the first node<sup>6</sup>, and applying Kirchhoff's law of currents we get the supply vector at this node, so equation (5.18) holds, before multiplying on the left by the inverse of the Laplacian to separate the potentials, we have to make a small comment about it, it is not possible as the laplatian of the graph will not have inverse; looking at the definition, the vector  $V = (1\ 1\dots 1)^T$  is an eigenvector with eigenvalue 0, this means that this set of equations:

$$p = L^{-1}b \quad (5.21)$$

will have infinite solutions, up to a constant, one approach is exploiting the fact that the laplatian accepts Moore–Penrose generalized inverse [115], but a simpler method consists in setting the potential of the first sink to zero, turning it into our potential reference, which is fine because we only care about potential differences. Mathematically this is done by defining a reduced Laplacian  $\tilde{L}$ , a matrix defined by setting the row and column of that node to zero, for example if that node is the first one we would get:

$$p = \begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & \tilde{L}^{-1} \end{pmatrix} \cdot b \quad (5.22)$$

Once we have the potentials the current flow betweenness of node  $i$ ,  $\tau_i$ , will be easily computed, as it is defined as the sum of all currents passing through  $i$ :

$$\tau_i = \frac{1}{2|S_0|} \sum_{s,v} |x_s(i,v)| \quad (5.23)$$

Where  $S_0$  is the number of sources, the sum is done for all sources defined in our signal  $s$  and for all nodes  $v$  that are connected with the target node  $i$ . We have to take into account that each source will define a different supply vector  $b_s(i)$  which will be also translated into a different current flow  $x_s(\vec{e})$ . The definition of the betweenness centrality can also be expressed in terms of the potentials as well using Ohm's law:

$$\tau_i = \frac{1}{2|S_0|} \sum_{s,j} |p_i(s) - p_j(s)| A_{ij} \quad (5.24)$$

---

<sup>6</sup>With the notation followed, all currents have positive sign but the value can be negative as it depends on the value of the difference  $p(1) - p(2)$ .

---

## Appendix A - Mathematical derivations

which will be the equation used in practice, since the potentials can easily be computed from the Laplatian using (Eq. 5.20), the notation  $p_i(s)$  describes the potential of vertex  $i$  when the potentials are computed from the source  $s$ .



# **Appendix B - List of Abbreviations**

## **Frequently used abbreviations**

|             |                            |
|-------------|----------------------------|
| <b>RMSD</b> | Root Mean Square Deviation |
| <b>WT</b>   | Wild Type                  |

## **Chapter 1: Introduction**

|            |                           |
|------------|---------------------------|
| <b>CRD</b> | Calcium Responsive Domain |
|------------|---------------------------|

## **Chapter 2**

|            |                           |
|------------|---------------------------|
| <b>CRD</b> | Calcium Responsive Domain |
|------------|---------------------------|

## **Chapter 3**

|            |                           |
|------------|---------------------------|
| <b>CRD</b> | Calcium Responsive Domain |
|------------|---------------------------|

## **Chapter 4**

|            |                           |
|------------|---------------------------|
| <b>CRD</b> | Calcium Responsive Domain |
|------------|---------------------------|



# Bibliography

- [1] Lucie Delemotte et al. “Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations”. In: *Proceedings of the National Academy of Sciences* 108.15 (2011), pp. 6109–6114. ISSN: 0027-8424. DOI: [10.1073/pnas.1102724108](https://doi.org/10.1073/pnas.1102724108).
- [2] Morten Ø. Jensen et al. “Mechanism of Voltage Gating in Potassium Channels”. In: *Science* 336.6078 (2012), pp. 229–233. DOI: [10.1126/science.1216533](https://doi.org/10.1126/science.1216533).
- [3] Aashish Manglik et al. “Structure-based discovery of opioid analgesics with reduced side effects”. In: *Nature* 537.7619 (Sept. 2016), pp. 185–190. ISSN: 1476-4687. DOI: [10.1038/nature19112](https://doi.org/10.1038/nature19112).
- [4] Aaron Chevalier et al. “Massively parallel de novo protein design for targeted therapeutics”. In: *Nature* 550.7674 (Oct. 2017), pp. 74–79. ISSN: 1476-4687. DOI: [10.1038/nature23912](https://doi.org/10.1038/nature23912).
- [5] K Wüthrich. “Protein structure determination in solution by NMR spectroscopy.” In: *Journal of Biological Chemistry* 265.36 (1990), pp. 22059–22062. ISSN: 0021-9258. DOI: [https://doi.org/10.1016/S0021-9258\(18\)45665-7](https://doi.org/10.1016/S0021-9258(18)45665-7).
- [6] Isabel Moraes et al. “Membrane protein structure determination — The next generation”. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1838.1, Part A (2014). Structural and biophysical characterisation of membrane protein-ligand binding, pp. 78–87. ISSN: 0005-2736. DOI: <https://doi.org/10.1016/j.bbamem.2013.07.010>.
- [7] Edward H. Egelman. “The Current Revolution in Cryo-EM”. In: *Biophysical Journal* 110.5 (Mar. 2016), pp. 1008–1012. ISSN: 0006-3495. DOI: [10.1016/j.bpj.2016.02.001](https://doi.org/10.1016/j.bpj.2016.02.001).
- [8] Yukihiko Sugita et al. “Cryo-EM structure of the Ebola virus nucleoprotein–RNA complex at 3.6 Å resolution”. In: *Nature* 563.7729 (Nov. 2018), pp. 137–140. ISSN: 1476-4687. DOI: [10.1038/s41586-018-0630-0](https://doi.org/10.1038/s41586-018-0630-0).
- [9] Heena Khatter et al. “Structure of the human 80S ribosome”. In: *Nature* 520.7549 (Apr. 2015), pp. 640–645. ISSN: 1476-4687. DOI: [10.1038/nature14427](https://doi.org/10.1038/nature14427).
- [10] Ji Sun and Roderick MacKinnon. “Structural Basis of Human KCNQ1 Modulation and Gating”. In: *Cell* 180.2 (2020), 340–347.e9. ISSN: 0092-8674. DOI: <https://doi.org/10.1016/j.cell.2019.12.003>.

## Bibliography

---

- [11] Scott A. Hollingsworth and Ron O. Dror. “Molecular Dynamics Simulation for All”. In: *Neuron* 99.6 (2018), pp. 1129–1143. ISSN: 0896-6273. DOI: <https://doi.org/10.1016/j.neuron.2018.08.011>.
- [12] James C. Phillips et al. “Scalable molecular dynamics on CPU and GPU architectures with NAMD”. In: *The Journal of Chemical Physics* 153.4 (2020), p. 044130. DOI: <10.1063/5.0014475>.
- [13] Mark James Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1-2 (2015), pp. 19–25. ISSN: 2352-7110. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>. URL: <https://www.sciencedirect.com/science/article/pii/S2352711015000059>.
- [14] Daan Frenkel and Berend Smit. “Chapter 4 - Molecular Dynamics Simulations”. In: *Understanding Molecular Simulation (Second Edition)*. Ed. by Daan Frenkel and Berend Smit. Second Edition. San Diego: Academic Press, 2002, pp. 63–107. ISBN: 978-0-12-267351-1. DOI: <https://doi.org/10.1016/B978-012267351-1/50006-7>.
- [15] Jorge Kohanoff. *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*. Cambridge University Press, 2006. DOI: <10.1017/CBO9780511755613>.
- [16] Chuan Tian et al. “ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution”. In: *Journal of Chemical Theory and Computation* 16.1 (Jan. 2020), pp. 528–552. ISSN: 1549-9618. DOI: <10.1021/acs.jctc.9b00591>.
- [17] Leela S. Dodda et al. “LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands”. In: *Nucleic Acids Research* 45.W1 (Apr. 2017), W331–W336. ISSN: 0305-1048. DOI: <10.1093/nar/gkx312>.
- [18] Jing Huang and Alexander D MacKerell Jr. “CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data”. eng. In: *Journal of computational chemistry* 34.25 (Sept. 2013). 23832629[pmid], pp. 2135–2145. ISSN: 1096-987X. DOI: <10.1002/jcc.23354>.
- [19] Alexander D. Mackerell Jr. “Empirical force fields for biological macromolecules: Overview and issues”. In: *Journal of Computational Chemistry* 25.13 (2004), pp. 1584–1604. DOI: <https://doi.org/10.1002/jcc.20082>.
- [20] Paul Bauer, Berk Hess, and Erik Lindahl. *GROMACS 2022 Manual*. Version 2022. Feb. 2022. DOI: <10.5281/zenodo.6103568>.
- [21] James C. Phillips et al. “Scalable molecular dynamics with NAMD”. In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1781–1802. DOI: <https://doi.org/10.1002/jcc.20289>.
- [22] Darryl K. Granner Peter A. Mayes Victor W. Rodwell Robert K. Murray. *Harper’s Illustrated Biochemistry*. 26th. McGraw-Hill Medical, 2003.
- [23] H. Lodish et al. *Molecular Cell Biology*. 4th. New York: W. H. Freeman, 2000.

- [24] Marija Liutkute, Ekaterina Samatova, and Marina V. Rodnina. “Cotranslational Folding of Proteins on the Ribosome”. In: *Biomolecules* 10.1 (2020). ISSN: 2218-273X. DOI: [10.3390/biom10010097](https://doi.org/10.3390/biom10010097).
- [25] Michael Thommen, Wolf Holtkamp, and Marina V Rodnina. “Co-translational protein folding: progress and methods”. In: *Current Opinion in Structural Biology* 42 (2017). Folding and binding • Proteins: Bridging theory and experiment, pp. 83–89. ISSN: 0959-440X. DOI: <https://doi.org/10.1016/j.sbi.2016.11.020>.
- [26] Janire Urrutia et al. “An epilepsy-causing mutation leads to co-translational misfolding of the Kv7.2 channel”. eng. In: *BMC biology* 19.1 (May 2021). 34020651[pmid], pp. 109–109. ISSN: 1741-7007. DOI: [10.1186/s12915-021-01040-1](https://doi.org/10.1186/s12915-021-01040-1).
- [27] Eider Núñez, Arantza Muguruza-Montero, and Alvaro Villarroel. “Atomistic Insights of Calmodulin Gating of Complete Ion Channels”. In: *International Journal of Molecular Sciences* 21.4 (2020). ISSN: 1422-0067. DOI: [10.3390/ijms21041285](https://doi.org/10.3390/ijms21041285).
- [28] Carolina Gomis-Pérez et al. “Homomeric Kv7.2 current suppression is a common feature in KCNQ2 epileptic encephalopathy”. In: *Epilepsia* 60.1 (2019), pp. 139–148. DOI: <https://doi.org/10.1111/epi.14609>.
- [29] Snezana Maljevic and Holger Lerche. “Chapter 2 - Potassium channel genes and benign familial neonatal epilepsy”. In: *Genetics of Epilepsy*. Ed. by Ortrud K. Steinlein. Vol. 213. Progress in Brain Research. Elsevier, 2014, pp. 17–53. DOI: <https://doi.org/10.1016/B978-0-444-63326-2.00002-8>.
- [30] William Humphrey, Andrew Dalke, and Klaus Schulten. “VMD – Visual Molecular Dynamics”. In: *Journal of Molecular Graphics* 14 (1996), pp. 33–38.
- [31] Xiaoxiao Li et al. “Molecular basis for ligand activation of the human KCNQ2 channel”. In: *Cell Research* 31.1 (Jan. 2021), pp. 52–61. ISSN: 1748-7838. DOI: [10.1038/s41422-020-00410-8](https://doi.org/10.1038/s41422-020-00410-8).
- [32] Paolo Ambrosino et al. “Epilepsy-causing mutations in Kv7.2 C-terminus affect binding and functional modulation by calmodulin”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1852.9 (2015), pp. 1856–1866. ISSN: 0925-4439. DOI: <https://doi.org/10.1016/j.bbadi.2015.06.012>.
- [33] Maria Virginia Soldovieri et al. “Novel KCNQ2 and KCNQ3 Mutations in a Large Cohort of Families with Benign Neonatal Epilepsy: First Evidence for an Altered Channel Regulation by Syntaxin-1A”. In: *Human Mutation* 35.3 (2014), pp. 356–367. DOI: <https://doi.org/10.1002/humu.22500>.
- [34] Arantza Muguruza-Montero et al. “Do calmodulin binding IQ motifs have built-in capping domains?” In: *Protein Science* 30.10 (2021), pp. 2029–2041. DOI: <https://doi.org/10.1002/pro.4170>.

## Bibliography

---

- [35] Alessandro Alaimo et al. “Calmodulin Activation Limits the Rate of KCNQ2 K<sup>+</sup> Channel Exit from the Endoplasmic Reticulum \*”. In: *Journal of Biological Chemistry* 284.31 (July 2009), pp. 20668–20675. ISSN: 0021-9258. DOI: [10.1074/jbc.M109.019539](https://doi.org/10.1074/jbc.M109.019539).
- [36] Guillaume Gotthard et al. “Chromophore Isomer Stabilization Is Critical to the Efficient Fluorescence of Cyan Fluorescent Proteins”. In: *Biochemistry* 56.49 (Dec. 2017), pp. 6418–6422. ISSN: 0006-2960. DOI: [10.1021/acs.biochem.7b01088](https://doi.org/10.1021/acs.biochem.7b01088).
- [37] Corinna Ottmann et al. “Applicability of superfolder YFP bimolecular fluorescence complementation in vitro”. In: 390.1 (2009), pp. 81–90. DOI: doi: [10.1515/BC.2009.008](https://doi.org/10.1515/BC.2009.008).
- [38] Ganeko Bernardo-Seisdedos et al. “Structural basis and energy landscape for the Ca<sup>2+</sup> gating and calmodulation of the Kv7.2 K<sup>+</sup> channel”. In: *Proceedings of the National Academy of Sciences* 115.10 (2018), pp. 2395–2400. DOI: [10.1073/pnas.1800235115](https://doi.org/10.1073/pnas.1800235115).
- [39] Nurzian Ismail et al. “A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration”. eng. In: *Nature structural & molecular biology* 19.10 (Oct. 2012). 23001004[pmid], pp. 1018–1022. ISSN: 1545-9985. DOI: [10.1038/nsmb.2376](https://doi.org/10.1038/nsmb.2376).
- [40] Ola B. Nilsson et al. “Cotranslational Protein Folding inside the Ribosome Exit Tunnel”. In: *Cell Reports* 12.10 (Sept. 2015), pp. 1533–1540. ISSN: 2211-1247. DOI: [10.1016/j.celrep.2015.07.065](https://doi.org/10.1016/j.celrep.2015.07.065).
- [41] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G. Wolynes. “THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective”. In: *Annual Review of Physical Chemistry* 48.1 (1997). PMID: 9348663, pp. 545–600. DOI: [10.1146/annurev.physchem.48.1.545](https://doi.org/10.1146/annurev.physchem.48.1.545).
- [42] Hahnbeom Park et al. “Automatic structure prediction of oligomeric assemblies using Robetta in CASP12”. eng. In: *Proteins* 86 Suppl 1.Suppl 1 (Mar. 2018). 28913931[pmid], pp. 283–291. ISSN: 1097-0134. DOI: [10.1002/prot.25387](https://doi.org/10.1002/prot.25387).
- [43] Nicholas A. Marze et al. “Efficient flexible backbone protein-protein docking for challenging targets”. eng. In: *Bioinformatics (Oxford, England)* 34.20 (Oct. 2018). 29718115[pmid], pp. 3461–3469. ISSN: 1367-4811. DOI: [10.1093/bioinformatics/bty355](https://doi.org/10.1093/bioinformatics/bty355).
- [44] Jens Meiler and David Baker. “ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility”. In: *Proteins: Structure, Function, and Bioinformatics* 65.3 (2006), pp. 538–548. DOI: <https://doi.org/10.1002/prot.21086>.
- [45] Kalli Kappel and Rhiju Das. “Sampling Native-like Structures of RNA-Protein Complexes through Rosetta Folding and Docking”. In: *Structure* 27.1 (2019), 140–151.e5. ISSN: 0969-2126. DOI: <https://doi.org/10.1016/j.str.2018.10.001>.

- [46] Brian D. Weitzner et al. “Modeling and docking of antibody structures with Rosetta”. In: *Nature Protocols* 12.2 (Feb. 2017), pp. 401–416. ISSN: 1750-2799. DOI: [10.1038/nprot.2016.180](https://doi.org/10.1038/nprot.2016.180).
- [47] Rebecca F. Alford et al. “An Integrated Framework Advancing Membrane Protein Modeling and Design”. In: *PLOS Computational Biology* 11.9 (Sept. 2015), pp. 1–23. DOI: [10.1371/journal.pcbi.1004398](https://doi.org/10.1371/journal.pcbi.1004398).
- [48] Valerie Vaissier Welborn and Teresa Head-Gordon. “Computational Design of Synthetic Enzymes”. In: *Chemical Reviews* 119.11 (June 2019), pp. 6613–6630. ISSN: 0009-2665. DOI: [10.1021/acs.chemrev.8b00399](https://doi.org/10.1021/acs.chemrev.8b00399).
- [49] Julia Koehler Leman et al. “Macromolecular modeling and design in Rosetta: recent methods and frameworks”. In: *Nature Methods* 17.7 (July 2020), pp. 665–680. ISSN: 1548-7105. DOI: [10.1038/s41592-020-0848-2](https://doi.org/10.1038/s41592-020-0848-2).
- [50] Nicholas Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: <https://doi.org/10.1063/1.1699114>.
- [51] Andrew Leaver-Fay et al. “Scientific benchmarks for guiding macromolecular energy function improvement”. eng. In: *Methods in enzymology* 523 (2013). 23422428[pmid], pp. 109–143. ISSN: 1557-7988. DOI: [10.1016/B978-0-12-394292-0.00006-0](https://doi.org/10.1016/B978-0-12-394292-0.00006-0).
- [52] Kyle A. Barlow et al. “Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation”. eng. In: *The journal of physical chemistry. B* 122.21 (May 2018). 29401388[pmid], pp. 5389–5399. ISSN: 1520-5207. DOI: [10.1021/acs.jpcb.7b11367](https://doi.org/10.1021/acs.jpcb.7b11367).
- [53] Rebecca F. Alford et al. “The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design”. eng. In: *Journal of chemical theory and computation* 13.6 (June 2017). 28430426[pmid], pp. 3031–3048. ISSN: 1549-9626. DOI: [10.1021/acs.jctc.7b00125](https://doi.org/10.1021/acs.jctc.7b00125).
- [54] Georg Kuenze et al. “Upgraded molecular models of the human KCNQ1 potassium channel”. In: *PLOS ONE* 14.9 (Sept. 2019), pp. 1–33. DOI: [10.1371/journal.pone.0220415](https://doi.org/10.1371/journal.pone.0220415).
- [55] Jiaren Zhang et al. “Identifying mutation hotspots reveals pathogenetic mechanisms of KCNQ2 epileptic encephalopathy”. In: *Scientific Reports* 10.1 (Mar. 2020), p. 4756. ISSN: 2045-2322. DOI: [10.1038/s41598-020-61697-6](https://doi.org/10.1038/s41598-020-61697-6).
- [56] Jainab Khatun, Sagar D. Khare, and Nikolay V. Dokholyan. ““Can contact potentials reliably predict stability of proteins?””. In: *Journal of molecular biology* 336.5 (2004), pp. 1223–1238. DOI: <https://doi.org/10.1016/j.jmb.2004.01.002>.
- [57] [https://www.rosettacommons.org/docs/latest/rosetta\\_basics/Units-in-Rosetta](https://www.rosettacommons.org/docs/latest/rosetta_basics/Units-in-Rosetta). Accessed: April of 2022.
- [58] [https://github.com/ORBallesteros/Thesis/tree/chapter2/rosetta\\_calculations/344\\_mutants](https://github.com/ORBallesteros/Thesis/tree/chapter2/rosetta_calculations/344_mutants). Accessed: September of 2022.

## Bibliography

---

- [59] Pekka Mark and Lennart Nilsson. “Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K”. In: *The Journal of Physical Chemistry A* 105.43 (Nov. 2001), pp. 9954–9960. ISSN: 1089-5639. DOI: [10.1021/jp003020w](https://doi.org/10.1021/jp003020w).
- [60] Scott E. Feller et al. “Constant pressure molecular dynamics simulation: The Langevin piston method”. In: *The Journal of Chemical Physics* 103.11 (1995), pp. 4613–4621. DOI: [10.1063/1.470648](https://doi.org/10.1063/1.470648).
- [61] Axel Brünger. “X-PLOR Version 3.1. A System for X-ray Crystallography and NMR”. In: Yale University Press, 1993. ISBN: 978-0300054026.
- [62] [https://github.com/ORBallesteros/Thesis/tree/chapter2/CaM-Clobe\\_hA](https://github.com/ORBallesteros/Thesis/tree/chapter2/CaM-Clobe_hA). Accessed: September of 2022.
- [63] [https://github.com/ORBallesteros/Thesis/tree/chapter2/channel\\_only\\_CRD](https://github.com/ORBallesteros/Thesis/tree/chapter2/channel_only_CRD). Accessed: September of 2022.
- [64] Shankar Kumar et al. “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. DOI: <https://doi.org/10.1002/jcc.540130812>.
- [65] Grossfield, Alan, \OT1\textquotedblleft WHAM:theweightedhistogramanalysismethod\OT1\textquotedblright, version 2.0.11, [http://membrane.urmc.rochester.edu/wordpress/?page\\_id=126](http://membrane.urmc.rochester.edu/wordpress/?page_id=126).
- [66] [https://github.com/ORBallesteros/Thesis/tree/chapter2/rosetta\\_calculations/double\\_mutants](https://github.com/ORBallesteros/Thesis/tree/chapter2/rosetta_calculations/double_mutants). Accessed: September of 2022.
- [67] [https://github.com/ORBallesteros/Thesis/tree/chapter2/double\\_mutants\\_CRD](https://github.com/ORBallesteros/Thesis/tree/chapter2/double_mutants_CRD). Accessed: September of 2022.
- [68] Janire Urrutia et al. “The Crossroad of Ion Channels and Calmodulin in Disease”. In: *International Journal of Molecular Sciences* 20.2 (2019). ISSN: 1422-0067. DOI: [10.3390/ijms20020400](https://doi.org/10.3390/ijms20020400).
- [69] Allen R. Rhoads and Felix Friedberg. “Sequence motifs for calmodulin recognition”. In: *The FASEB Journal* 11.5 (1997), pp. 331–340. DOI: <https://doi.org/10.1096/fasebj.11.5.9141499>.
- [70] Nicholas J. Anthis, Michaeleen Doucleff, and G. Marius Clore. “Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit”. In: *J. Am. Chem. Soc.* 133.46 (Nov. 2011), pp. 18966–18974. DOI: [10.1021/ja2082813](https://doi.org/10.1021/ja2082813).
- [71] Fei Liu et al. “Molecular Mechanism of Multispecific Recognition of Calmodulin through Conformational Changes”. In: *Proc. Natl. Acad. Sci. U. S. A.* 114.20 (2017), E3927–E3934. ISSN: 0027-8424. DOI: [10.1073/pnas.1615949114](https://doi.org/10.1073/pnas.1615949114).
- [72] Chia-Hsueh Lee and Roderick MacKinnon. “Activation mechanism of a human SK-calmodulin channel complex elucidated by cryo-EM structures”. In: *Science* 360.6388 (2018), pp. 508–513. DOI: [10.1126/science.aas9466](https://doi.org/10.1126/science.aas9466).

- [73] Andrew Waterhouse et al. “SWISS-MODEL: homology modelling of protein structures and complexes”. In: *Nucleic acids research* 46.W1 (July 2018). 29788355[pmid], W296–W303. ISSN: 1362-4962. DOI: [10.1093/nar/gky427](https://doi.org/10.1093/nar/gky427).
- [74] Stefan Bienert et al. “The SWISS-MODEL Repository-new features and functionality”. In: *Nucleic acids research* 45.D1 (Jan. 2017). 27899672[pmid], pp. D313–D319. ISSN: 1362-4962. DOI: [10.1093/nar/gkw1132](https://doi.org/10.1093/nar/gkw1132).
- [75] Ralph Wissmann et al. “A Helical Region in the C Terminus of Small-Conductance  $\text{Ca}^{2+}$ -Activated  $\text{K}^+$  Channels Controls Assembly with apo-Calmodulin”. In: *J. Biol. Chem.* 277 (2002), pp. 4558–4564. DOI: [10.1074/jbc.M109240200](https://doi.org/10.1074/jbc.M109240200).
- [76] Rafael C. Bernardi, Marcelo C.R. Melo, and Klaus Schulten. “Enhanced sampling techniques in molecular dynamics simulations of biological systems”. In: *Biochimica et Biophysica Acta (BBA) - General Subjects* 1850.5 (2015). Recent developments of molecular dynamics, pp. 872–877. ISSN: 0304-4165. DOI: <https://doi.org/10.1016/j.bbagen.2014.10.019>.
- [77] Lingle Wang, Richard A. Friesner, and B. J. Berne. “Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2)”. In: *The Journal of Physical Chemistry B* 115.30 (Aug. 2011), pp. 9431–9438. ISSN: 1520-6106. DOI: [10.1021/jp204407d](https://doi.org/10.1021/jp204407d).
- [78] Alessandro Laio and Michele Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. DOI: [10.1073/pnas.202427399](https://doi.org/10.1073/pnas.202427399).
- [79] Yuji Sugita and Yuko Okamoto. “Replica-exchange molecular dynamics method for protein folding”. In: *Chemical Physics Letters* 314.1 (1999), pp. 141–151. ISSN: 0009-2614. DOI: [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
- [80] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. “Metadynamics”. In: *WIREs Computational Molecular Science* 1.5 (2011), pp. 826–843. DOI: <https://doi.org/10.1002/wcms.31>.
- [81] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Phys. Rev. Lett.* 100 (2 Jan. 2008), p. 020603. DOI: [10.1103/PhysRevLett.100.020603](https://doi.org/10.1103/PhysRevLett.100.020603).
- [82] Giovanni Bussi and Alessandro Laio. “Using metadynamics to explore complex free-energy landscapes”. In: *Nature Reviews Physics* 2.4 (Apr. 2020), pp. 200–212. ISSN: 2522-5820. DOI: [10.1038/s42254-020-0153-0](https://doi.org/10.1038/s42254-020-0153-0).
- [83] W. Kabsch and C. Sander. “Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features”. In: *Biopolymers* 22 (1983), pp. 2577–2637. DOI: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211).
- [84] Gareth A. Tribello et al. “PLUMED2: New Feathers for an Old Bird”. In: *Comput. Phys. Commun.* 185 (2014), pp. 604–613. DOI: [10.1016/j.cpc.2013.09.018](https://doi.org/10.1016/j.cpc.2013.09.018).

## Bibliography

---

- [85] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical Sampling through Velocity Rescaling”. In: *J. Chem. Phys.* 126 (2007), p. 014101. DOI: [10.1063/1.2408420](https://doi.org/10.1063/1.2408420).
- [86] M. Parrinello and A. Rahman. “Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method”. In: *J. Appl. Phys. (Melville, NY, U. S.)* 52 (1981), pp. 7182–7190. DOI: [10.1063/1.328693](https://doi.org/10.1063/1.328693).
- [87] Miao Zhang et al. “Structural Basis for Calmodulin as a Dynamic Calcium Sensor”. In: *Structure* 20.5 (May 2012), pp. 911–923. ISSN: 0969-2126. DOI: [10.1016/j.str.2012.03.019](https://doi.org/10.1016/j.str.2012.03.019).
- [88] <https://github.com/ORBallesteros/TESIS/chapter2/cam.....COMPLETAR>. Accessed: September of 2022.
- [89] <https://github.com/ORBallesteros/TESIS/chapter2/cam.....COMPLETAR>. Accessed: September of 2022.
- [90] G. Fiorin et al. “Using Metadynamics to Understand the Mechanism of Calmodulin/Target Recognition at Atomic Detail”. In: *Biophysical Journal* 91.8 (2006), pp. 2768–2777. ISSN: 0006-3495. DOI: <https://doi.org/10.1529/biophysj.106.086611>.
- [91] <https://github.com/ORBallesteros/TESIS/chapter2/cam.....COMPLETAR>. Accessed: September of 2022.
- [92] <https://github.com/ORBallesteros/TESIS/chapter2/cam.....COMPLETAR>. Accessed: September of 2022.
- [93] Po Wei Kang et al. “Calmodulin acts as a state-dependent switch to control a cardiac potassium channel opening”. In: *Science Advances* 6.50 (2020), eabd6798. DOI: [10.1126/sciadv.abd6798](https://doi.org/10.1126/sciadv.abd6798).
- [94] Lily T.-Y. Cho et al. In: *Structure* 26.4 (Apr. 2018), 533–544.e3. ISSN: 0969-2126.
- [95] Sunhwan Jo et al. “CHARMM-GUI: A web-based graphical user interface for CHARMM”. In: *Journal of Computational Chemistry* 29.11 (2008), pp. 1859–1865. DOI: <https://doi.org/10.1002/jcc.20945>.
- [96] Sunhwan Jo, Taehoon Kim, and Wonpil Im. “Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations”. In: *PLOS ONE* 2.9 (Sept. 2007), pp. 1–9. DOI: [10.1371/journal.pone.0000830](https://doi.org/10.1371/journal.pone.0000830).
- [97] Jumin Lee et al. “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field”. In: *Journal of Chemical Theory and Computation* 12.1 (Jan. 2016), pp. 405–413. ISSN: 1549-9618. DOI: [10.1021/acs.jctc.5b00935](https://doi.org/10.1021/acs.jctc.5b00935).
- [98] Jeffery B. Klauda et al. “Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types”. In: *The Journal of Physical Chemistry B* 114.23 (June 2010), pp. 7830–7843. ISSN: 1520-6106. DOI: [10.1021/jp101759q](https://doi.org/10.1021/jp101759q).

- [99] Seonghoon Kim et al. “CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules”. In: *Journal of Computational Chemistry* 38.21 (2017), pp. 1879–1886. DOI: <https://doi.org/10.1002/jcc.24829>.
- [100] Woody Sherman et al. “Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects”. In: *Journal of Medicinal Chemistry* 49.2 (Jan. 2006), pp. 534–553. ISSN: 0022-2623. DOI: [10.1021/jm050540c](https://doi.org/10.1021/jm050540c).
- [101] <https://github.com/ORBallesteros/TESIS/chapter2/cam.....COMPLETAR>. Accessed: September of 2022.
- [102] Csaba Böde et al. “Network analysis of protein dynamics”. In: *FEBS Letters* 581.15 (2007), pp. 2776–2782. DOI: <https://doi.org/10.1016/j.febslet.2007.05.021>.
- [103] Jeffrey R. Wagner et al. “Emerging Computational Methods for the Rational Discovery of Allosteric Drugs”. eng. In: *Chemical reviews* 116.11 (June 2016). 27074285[pmid], pp. 6370–6390. ISSN: 1520-6890. DOI: [10.1021/acs.chemrev.5b00631](https://doi.org/10.1021/acs.chemrev.5b00631).
- [104] L. Di Paola et al. “Protein Contact Networks: An Emerging Paradigm in Chemistry”. In: *Chemical Reviews* 113.3 (Mar. 2013), pp. 1598–1613. ISSN: 0009-2665. DOI: [10.1021/cr3002356](https://doi.org/10.1021/cr3002356).
- [105] Mahdi Ghorbani, Bernard R. Brooks, and Jeffery B. Klauda. “Exploring dynamics and network analysis of spike glycoprotein of SARS-CoV-2”. eng. In: *Biophysical journal* 120.14 (July 2021). 33705760[pmid], pp. 2902–2913. ISSN: 1542-0086. DOI: [10.1016/j.bpj.2021.02.047](https://doi.org/10.1016/j.bpj.2021.02.047).
- [106] D. J. Klein. “Centrality measure in graphs”. In: *Journal of Mathematical Chemistry* 47.4 (May 2010), pp. 1209–1223. ISSN: 1572-8897. DOI: [10.1007/s10910-009-9635-0](https://doi.org/10.1007/s10910-009-9635-0).
- [107] Andre A. S. T. Ribeiro and Vanessa Ortiz. “Determination of Signaling Pathways in Proteins through Network Theory: Importance of the Topology”. In: *Journal of Chemical Theory and Computation* 10.4 (Apr. 2014), pp. 1762–1769. ISSN: 1549-9618. DOI: [10.1021/ct400977r](https://doi.org/10.1021/ct400977r).
- [108] Andre A. S. T. Ribeiro and Vanessa Ortiz. “Energy Propagation and Network Energetic Coupling in Proteins”. In: *The Journal of Physical Chemistry B* 119.5 (Feb. 2015), pp. 1835–1846. ISSN: 1520-6106. DOI: [10.1021/jp509906m](https://doi.org/10.1021/jp509906m).
- [109] Wesley M. Botello-Smith and Yun Luo. “Robust Determination of Protein Allosteric Signaling Pathways”. In: *Journal of Chemical Theory and Computation* 15.4 (Apr. 2019), pp. 2116–2126. ISSN: 1549-9618. DOI: [10.1021/acs.jctc.8b01197](https://doi.org/10.1021/acs.jctc.8b01197).
- [110] Ulrik Brandes and Daniel Fleischer. “Centrality Measures Based on Current Flow”. In: *STACS 2005*. Ed. by Volker Diekert and Bruno Durand. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 533–544. ISBN: 978-3-540-31856-9.

## Bibliography

---

- [111] M.E. J. Newman. “A measure of betweenness centrality based on random walks”. In: *Social Networks* 27.1 (2005), pp. 39–54. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2004.11.009>.
- [112] Adam T. Van Wart et al. “Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis”. In: *Journal of Chemical Theory and Computation* 10.2 (Feb. 2014), pp. 511–517. ISSN: 1549-9618. DOI: [10.1021/ct4008603](https://doi.org/10.1021/ct4008603).
- [113] Annie M. Westerlund et al. “Network analysis reveals how lipids and other cofactors influence membrane protein allostery”. In: *The Journal of Chemical Physics* 153.14 (2020), p. 141103. doi: [10.1063/5.0020974](https://doi.org/10.1063/5.0020974).
- [114] <https://github.com/delemottelab/allosteric-pathways>.
- [115] Enrico Bozzo and Massimo Franceschet. “Resistance distance, closeness, and betweenness”. In: *Social Networks* 35.3 (2013), pp. 460–469. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2013.05.003>.