Universidad del País Vasco Euskal Herriko Unibertsitatea

Tesis Doctoral

Structural characterization of helices A and B of $K_v7.2$ channel bound to Calmodulin. The calcium effect.

Memoria presentada por

D. Ganeko Bernardo Seisdedos

dirigida por

Dr. Alvaro Villarroel Muñoz

Dr. Oscar Millet Aguilar-Galindo

para optar al título de

Doctor por la Universidad del País Vasco (UPV/EHU)

Unidad de Biofísica (CSIC-UPV/EHU)

Departamento de Bioquímica y Biología Molecular

Facultad de Ciencia y Tecnología

Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

Table of Contents

1. Introduction....1

1.1 Ion Channels....3

1.1.1 How The Spark Of The Communication Works....4

1.2 Voltage-Dependent Ion Channels....7

1.3 Voltage-Gated Potassium Channels (VGPC)....10

1.4 Kv7 Channels And M-Current....17

1.4.1Kv7 family members....18

1.4.2 Regulation Of The M-Current in Kv7.2 And Kv7.3....20

1.4.3 Channalopathies Related To Kv7.2 / Kv7.3....27

1.5 Calcium Signalling....29

1.5.1 EF-Hand Protein Family....31

1.5.2 Calmodulin....32

1.5.2.1 Architecture....34

1.5.2.2 Calcium binding and affinity....35

1.5.2.3 Calmodulin target recognition....36

1.5.2.4 CaM and Disease....40

Objectives....43

2. Improving Recombinant protein expression of Kv7 intracellular segment in Escherichia coli....45

2.1 Introduction....47

2.1.1 Lack Of Information About Kv7 Channels: The Challenge Of Working With Membrane Proteins....47

2.1.2 Literature Reading And In Silico Prediction Based On Sequence…49

2.1.3 Strategies for over-expression at host level....50

2.1.4 Strategies For Over-Expression Engineering Target DNA....53

2.1.5 Screening And Selection Of Recombinant Protein....60

2.2 Materials And Methods....62

2.2.1 Molecular Biology Techniques....62

2.2.1.1 PCR technique....63
2.2.1.2 Fragment digestion, purification and ligation....64
2.2.1.3 Bacterial transformation....65
2.2.1.4 DNA extraction and quantification....66
2.2.2 *In Silico* Analysis....67
2.2.3 Protein Expression and Solubility Analysis (Pilot Experiments)....68
2.2.4 Protein Purification....69
2.2.4.1 CaM purification....69
2.2.4.2 CaM dansylation....69
2.2.4.3 His fusion proteins....69
2.2.4.4 His-CaM, His-MBP, His-NUS and His-Trx fusion proteins....70
2.2.4.5 GST fusion proteins....71
2.2.5 Spectroscopy....72
2.2.5.1 Dynamic Light Scattering....72
2.2.5.2 Fluorescence Spectroscopy....72
2.3 Results....74
2.3.1 *In Silico* Analysis Of The C-Terminus of Kv7.2 Channels Reveals Putative Soluble Protein....74
2.3.2 C-Terminal Constructs Are Intrinsically Insoluble When They Are Over-Expressed....76
2.3.3 GST-Tagged Proteins Were Soluble After The Resolubilisation Of Inclusion Bodies....77
2.3.4 Humanized Strains Are A More Convenient Alternative To The Codon-Usage Optimization Of The Target DNA Sequence....78
2.3.5 Unlike Other Tags, CaM Fusion Proteins Are Soluble And Properly Folded....80
2.3.6 None Of The Selected Kv7 Channel Orthologous Shows Improved Solubility Profile....84
2.4 Discussion....85
2.5 Appendix....88

3 Structural characterization of CaM/K_v7.2_AB complex.....89
3.1 Introduction.....91
3.1.1 Atomic Protein Models Definition Techniques.....91
3.1.2 Protein Structure Determination by NMR.....94
3.1.3 NMR Sample Preparation.....110
3.2 Materials And Methods.....114
3.2.1 Molecular biology techniques.....114
3.2.2 Construct Selection.....117
3.2.2.1 Thermostability Assay Based on Venus Fluorescence.....117
3.2.2.2 Q2ABc Δ6L Alternative Constructs With Fused CaM and Venus.....118
3.2.2.3 Q2ABc Δ6L Alternative Constructs Co-Expressing CaM.....119
3.2.2.4 Electrophysiological recordings.....119
3.2.3 Buffer Screening.....120
3.2.4 Protein Complex Purification for NMR.....121
3.2.4.1 Whole complex labelling.....121
3.2.4.2 Whole protein purification.....122
3.2.4.3 Independent labelling of each protein from the complex.....123
3.2.4.4 Complex reconstitution on-column.....124
3.2.4.4.1 Calmodulin purification.....124
3.2.4.4.2Kv7.2 alternative construct resolubilisation from inclusion bodies.....124
3.2.4.4.3 Complex formation on-column.....125
3.2.4.5 Anisotropic media for RDCs recording.....125
3.2.5 Side-Chain and NOE assignments.....127
3.2.6 Structure Calculation.....127
3.2.7 Protein Docking.....127
3.2.8 NMR Acquisition and Processing.....128
3.3 Results.....130

3.3.1 “GB116” Kv7.2 helices A-B Alternative Construct is Selected for Structural Analysis.....130

3.3.1.1 Deletion ΔR374_K493 (Δ6L) between helices AB has high thermostability.....130

3.3.1.2 ΔR374_T501 deletion between helices AB is compatible with channel function in HEK293T cells.....131

3.3.1.3 Low ionic strength, basic or acid pHs and moderate buffering agent concentrations are optimal for buffering Kv7.2 C-terminus.....132

3.3.1.4 “GB116” alternative construct with deletions ΔF316_R325, Δ6L and ΔP533_H546 is the most convenient for further NMR analysis.....134

3.3.215 ^{15}N-HSQC Spectrum Encourages For Further Structural Analysis.....136

3.3.3 ^{15}N-HSQC Spectrum Of Individually Labelled Proteins Confirms Correct Complex Formation On-Column.....137

3.3.4 Protein-Complex BackBone Assignments.....138

3.3.4.1 Random coils detection.....138

3.3.4.2 Secondary structure detection.....140

3.3.4.3 Dihedral angles prediction.....142

3.3.5 RDCs Measurements shows Relative Orientations of Complex forming Proteins.....143

3.3.6 GB116 Protein Structure Confirms an anti-parallel Coiled-Coil Conformation of Helices AB.....144

3.3.7 Calmodulin Structure Shows a Semi-Open Conformation.....147

3.3.8 Final Complex Formation: Protein Docking.....150

3.4 Discussion.....151

4. The Calcium Effect.....155

4.1 Introduction.....157

4.1.1 Protein Complexes Beyond Static Representations.....157

4.1.2 Techniques to Study Protein Dynamics.....158

4.1.2.1 X-Ray Crystallography.....158

4.1.2.2 NMR.....158

4.1.3 Techniques to Study Protein Interactions.....160

4.1.3.1 X-Ray Crystallography.....160

4.1.3.2 Electron microscopy.....160

4.1.3.3 NMR.....161

4.1.3.4 Small Angle X-Ray Scattering (SAXS).....164

4.1.3.5 Föster Resonance Energy Transfer (FRET).....164

4.1.3.6 Surface Plasmon Resonance (SPR).....164

4.1.3.7 Isothermal titration calorimetry (ITC).....164

4.1.3.8 Other methods.....165

4.1.4 Kv7 Channels and Dynamics.....165

4.2 Materials And Methods.....166

4.2.1 Molecular Biology Techniques.....166

4.2.2 Protein Binding Characterization Fluorescence Spectroscopy.....167

4.2.3 CaM-Kv7.2 Complex Chemical Shifts Library Comparison.....169

4.2.4 Calcium Titration on 15N-CaM/Kv7.2_AB Complex.....170

4.3 Results.....172

4.3.1 D-CaM Binding Characterization Towards Helices A and B.....172

4.3.1.1 CaM complex formation with Q2ABc weakens its Ca2+ binding affinity.....172

4.3.1.2 CaM interaction with Q2ABc is dominated by N-lobe in the absence and by C-lobe in the presence of Ca2+.....173

4.3.1.3 CaM interaction with Q2ABc is dominated by Helix B in the absence and by Helix A in the presence of Ca2+.....175

4.3.2 Chemical Shifts Comparison Shows a Likely calcification-state of the N-lobe and an Apo-state of the C-lobe.....177

4.3.3 Calcium Titration Analysed by NMR.....179

4.3.3.1 Slow-Exchange dynamics uncovers three sets of signals in 15N-HSQCs.....180

4.3.3.2 Calcium is present in the N-lobe and is absent in C-lobe of the structurally characterized CaM/Kv7.2_AB complex.....181

4.3.3.3 Calcium does not provoke any CSPs in AB helices.....183

4.4 Discussion.....184

Conclusions.....191
5. Resumen.....193
Bibliography.....199
Publications.....211
Appendix.....213
Abbreviations.....213
Figure List.....214
Table List.....217

1 INTRODUCTION

1.1 ION CHANNELS

Nervous systems are based on electrical signals that enable a high-speed communication between different cells. The source of this ability is due to the ion channels located in the plasma membrane: all the cells are delimited by a lipid bilayer that is considered as a selective barrier which demarcate the inner and the outer part of the cell. When the concentration of different solutes like phosphate compounds, amino acids and ions are different on both sides of the membrane, an electrical potential arises. Along evolution, all cells have developed a plethora of mechanisms to maintain this accumulation of charges far from their equilibrium. Additionally, some types of cells have even been able to take advantage of the transmembrane electrical potential so as to store energy and to synthesize ATP. Furthermore, other specialized cells, better known as "excitable cells" – neurons, endocrine cells and muscular cells – have specialized to produce and handle electrical signals in an exquisite way, as mentioned before.

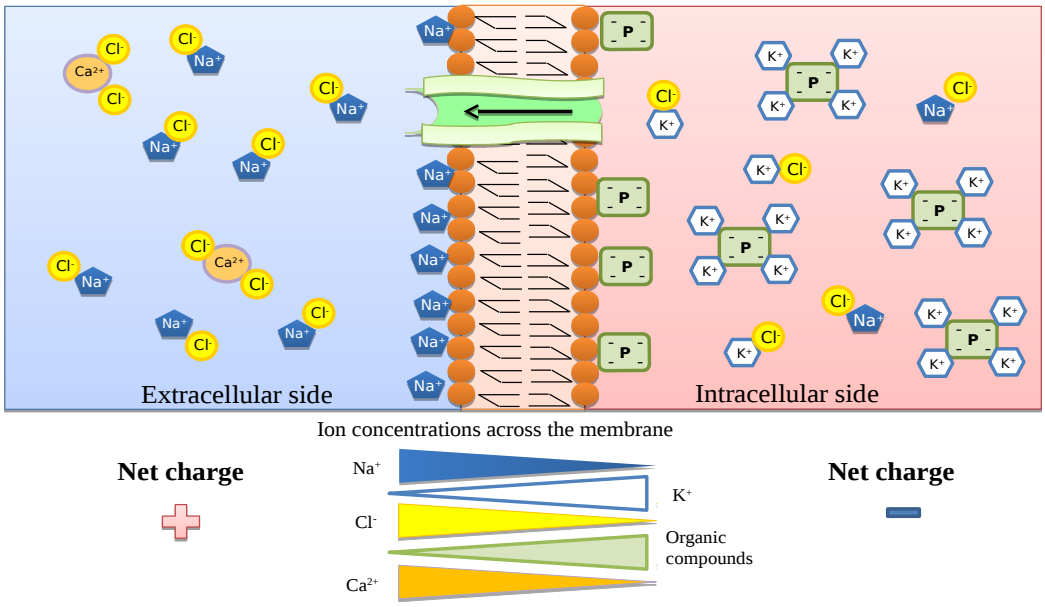
Fast electrical signalling is driven by ions. The most widely used cations, as we will see later, are sodium and potassium monovalent cations (Na^+ and K^+) and calcium divalent cation (Ca^{2+}), whereas the most prevalent anion is the chloride (Cl^-). Steady and slow homeostatic mechanisms of animal cells such as ion pumps and transporters spend around 50 % of their metabolic energy to establish the standard environment and ionic composition: outside the cell we find a high [Na^+] and [Cl^-] with a low [K^+], whereas in the intracellular region there is a high [K^+], but a low Na^+, Cl^- and Ca^{2+} concentration. In the case of the excitable cells, these ionic gradients facilitate rapid changes in membrane voltage, due to passive flow through ion channels. These proteins form a pore in the plasma membrane where ions rapidly and selectively pass by along their electrochemical gradient. Despite this efficient mechanism, gating of the channels is often regulated and coupled to the cellular processes and requirements.

Due to the small dimension of the cell, membrane voltage can rapidly change from positive to negative values only as a result of small changes in the concentration of each ion. This code of electrical signals can be translated between excitable and non-excitable cells and are ultimately responsible for the activation or inhibition of some cellular processes such us muscular contraction, hormone release, generation of the action potential or changes in the cell cycle. Given such crucial roles, a malfunction of any ion channel often derive in different diseases also known as channelopathies.

1.1.1 How The Spark Of The Communication Works

In order to explain how electrical signalling is generated, we should first define the **resting membrane potential**. Depending on the ion-pumps, transporters and exchangers (remarkably Na^+/K^+ pump) that are located in the plasma membrane, each cell type will have a different resting potential. Specifically, the involved proteins ensure different concentrations of the ions across the membrane, thus ultimately shaping the net charge observed between both sides of the membrane (Figure 1.1).

(A)



(B) Nernst equation (Eq. I1)

$$V = \frac{RT}{zF} \ln \frac{C_o}{C_i}$$

(C) Goldman-Hodgkin-Katz equation

$$V_m = \frac{RT}{zF} \ln \left(\frac{\Sigma_i^N P_{Mi}^+ [M_i^+]_{out} + \Sigma_j^N P_{Aj}^- [A_j^-]_{\text{in}}}{\Sigma_i^N P_{Mi}^+ [M_i^+]_{\text{in}} + \Sigma_j^N P_{Aj}^- [A_j^-]_{out}} \right)$$

$$V_m = \frac{RT}{zF} \ln \left(\frac{P_K [K^+]_{out} + P_{Na} [Na^+]_{out} + P_{Cl} [Cl^-]_{\text{in}}}{P_K [K^+]_{\text{in}} + P_{Na} [Na^+]_{\text{in}} + P_{Cl} [Cl^-]_{out}} \right)$$

Figure 1.1. Resting membrane potential. (A) Differences in concentration of ions in both sides of the membrane. (B) When the equilibrium is reached Nernst equation is applied to calculate the equilibrium potential for each ion where *V*, equilibrium potential in Volts; *R*, gas constant (2 cal mol^{-1} K^{-1}); *T*, absolute temperature (Kelvin); *z*, charge (valence) of the ion; *F*, Faraday's constant (2.3×10^4 cal V^{-1} mol^{-1}); *ln*, logarithm with e base; C_o *and* C_i, outside and inside concentrations of the ion. (C) Goldman equation for more than one ion where V_m, membrane potential in Volts; P_{ion}, permeability for a specific ion (meters per second); $[ion]_{out}$, extracellular concentration of the ion (mol/m^2); $[ion]_{in}$, intracellular concentration of the ion (mol/s^2). Second equation is adapted for most abundant ions in cells.

When the cell reaches the electrochemical equilibrium for a certain ion, the net flux for this ion stops. Under this conditions the **Nernst equation** can be applied to calculate the resting potential for a given ion. This equation takes into account both, the voltage and the concentration gradients across the membrane. If we keep in mind that the membrane potential is influenced by more than one ion, it is best to use **Goldman-Hogkin-Katz equation** which contemplates all multi-ionic scenarios. Actually, this equation is used nowadays to estimate the theoretical resting potential in any kind of cell.

The membrane potential of excitable cells does not remain steady. In the plasma membrane excitable cells have inserted voltage-gated channels that are responsible of triggering the **action potential** (AP) (Figure 1.2).

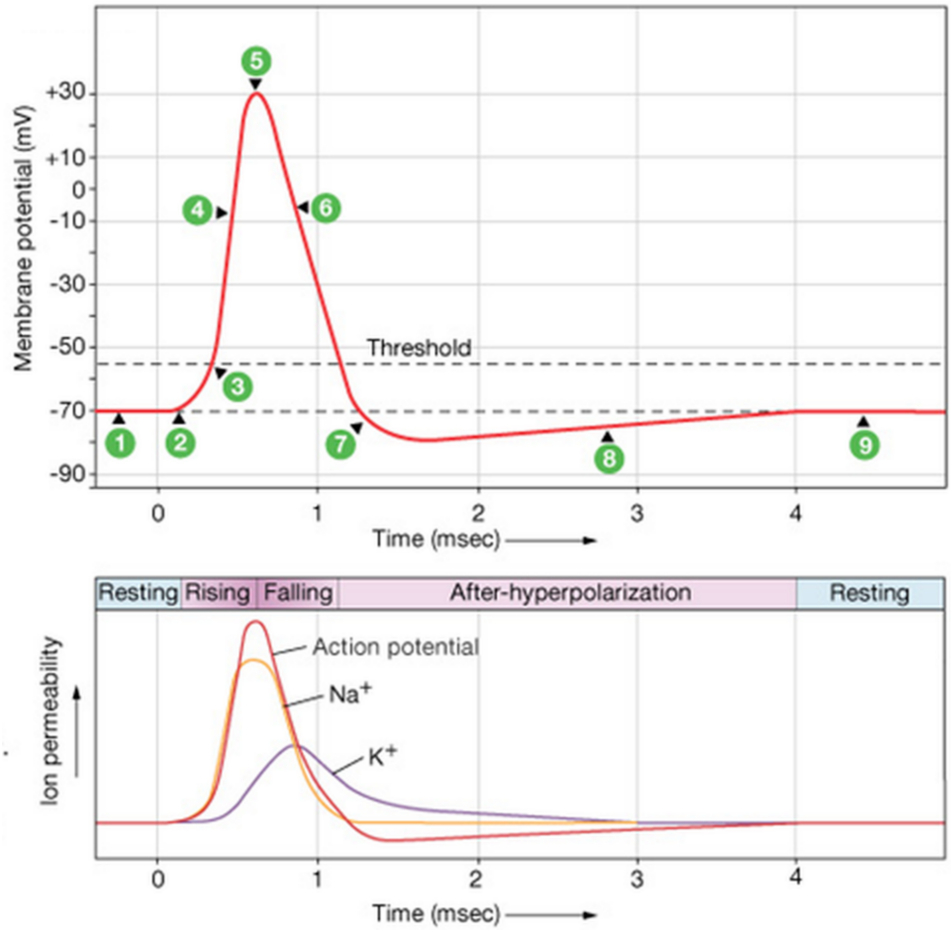


Figure 1.2. Ideal representation of an action potential with its different phases. (1) Without any stimulus, resting membrane potential. (2) Cell suffers a depolarizing stimulus. (3) If incoming stimulus overcome the threshold, voltage sensing Na_v channels open. (4) Na^+ inward flow depolarize the cell. (5) Rapid influx of Na^+ causes plasma membrane polarity reversion. Therefore Na^+ ion channels rapidly inactivate and slower K^+ channels open. (6) Na^+ ions can no longer enter the neuron and they are transported out of the cell by active sodium and potassium pumps. (7) K^+ channels remains activated, and there is an outward flow of K^+ ions. In this case plasma membrane suffers a hyperpolarization which creates a brake to new depolarizing stimulus. (8) K_v channels close and some have K^+ enters through leak channels. (9) After a while, the electrochemical gradient returns to the resting state. Adapted from Human Physiology, an integrated approach, third edition.

Whenever one of those channels opens its gates an ion flux through the membrane is initiated, letting millions of ions (10^6-10^8 ions per second) travelling passively following their electrochemical gradient. Even though several voltage gated ion channels take part in the generation and propagation of the AP, Na^+ and K^+ channels (Na_v and K_v) are to be considered as key elements in the process. For instance, in nerve cells, depolarizing stimuli which overcome the threshold are necessary to open Na_v channels, provoking a massive influx of Na^+ ions to the inner part of the cell consequently causing a depolarization (Figure 1.2.2). After a positive feedback, this phenomena will cause the simultaneous opening of more Na_v channels, resulting in a change in the resting value from -70 mV to more positive values, (close to the Na^+ Nernst equilibrium value around +50 mV), in just few milliseconds (Figure 1.2.3). To return to its resting potential two important mechanisms are required for the cell: on one hand, Na_v channels change their conformation from active to inactive automatically, a state in which a new stimulus cannot reactivate the process. A few milliseconds are needed after the membrane potential to ensure that channels are back into their resting potential. At this point Na_v channels recover their resting molecular conformation. On the other hand, K_v channels open to speed up the recovery of the original negative potential in the cell membrane.

Actually, both voltage-gated ion channels are closely coordinated and just after Na_v channels open, K_v channels begin to get active (Figure 1.2.5). In the case of K_v, enabling the free traffic of K^+ from the inside to the outside, the membrane potential will produce hyperpolarization, according to Nernst equilibrium, (Figure 1.2.6-I2.7). Additionally, K_v channels also suffer conformational changes after their activation through inactivation to resting state, but these channels show slightly slower kinetic as compared to sodium channels (Figure 1.2.8-2.9).

1.2 VOLTAGE-DEPENDENT ION CHANNELS

More than 200 genes have been described in humans that encode for ion channels, and within this group around 140 members are regulated by voltage, either inhibiting or activating them. All of them are built according to a common structure, although they show enough variations to provide adequate functional specialization. Actually, K_v, Ca_v and Na_v channels are members of a phylogenetically related superfamily (Gutman et al., 2005). According to these authors, the common ancestor of all voltage-gated ion channels could be the K_v, because it appears to be the simplest and also the most divergent group of channels (Figure 1.3).

A classical voltage-dependent ion channel is organized in multimers, composed by the assembly of four or even more α subunits that can be identical or different and they organize to establish the selective pore for one type of ion. Every single subunit has at least two transmembrane (TM) helices traversed in the lipid bilayer which ensembles the pore-forming domain. The rest of transmembrane domains or intra- or extra-cellular terminus completes the regulatory domain and will vary depending on the channel type.

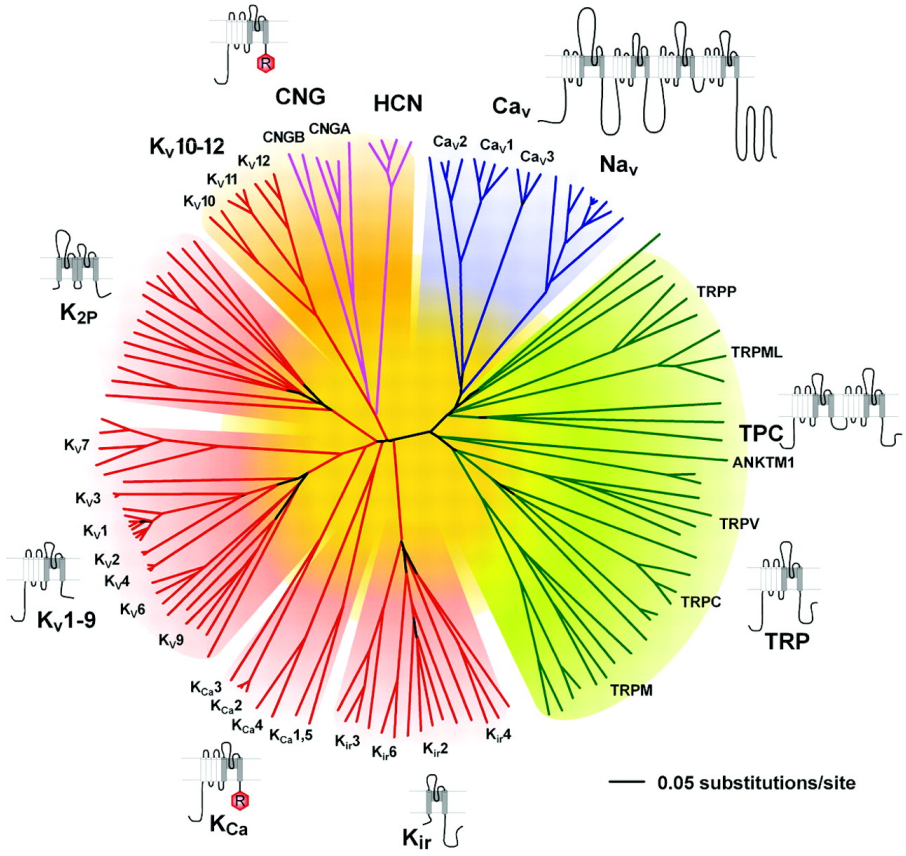


Figure 1.3. Phylogeny tree of VGPCs (A) Representation of the amino acid relationships of the minimal pore regions of the voltage-gated ion channels superfamily. Seven groups are related genetically. In blue, four-domain channels (Ca_v and Na_v); in red, potassium channels except for Kv10-12 channels that are more related to CNGs; in green, TRPs where two pore channels are included. Taken from (Yu et al., 2005).

A classical voltage-dependent ion channel shares equivalent architecture. Every single subunit has six hydrophobic segments which are inserted in the plasma membrane forming the pore (Figure 1.4). Moreover, segments S1-S4 are responsible for the voltage sensing, especially the forth segment which has several amino acids with positive charge (Long et al., 2005). They have intracellular carboxyl and amino moieties which often are important for the channels regulation because they comprise the intracellular interaction site for ligands. In fact, usually these kinds of ion channels are expressed in association with one or more auxiliary subunits which modulate and increase their functional properties.

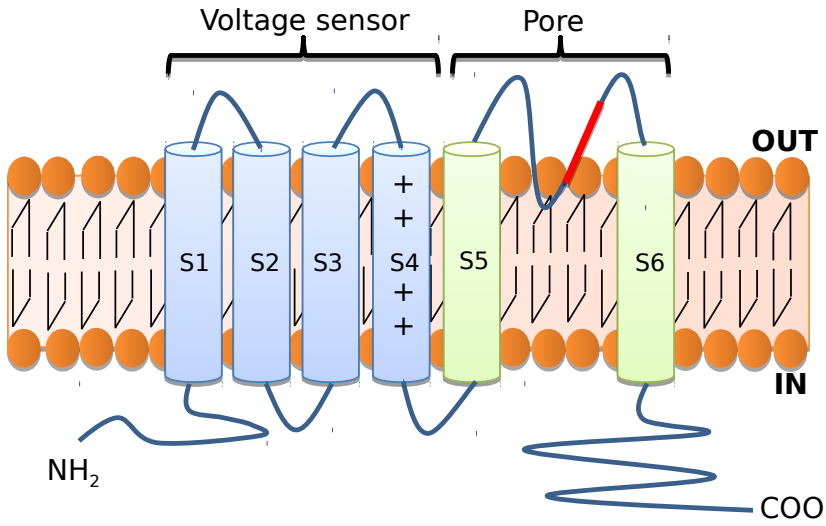


Figure 1.4. General architecture of a voltage-gated ion channel. Each subunit is made of six transmembranal helix (S1-S6) with N- and C- intracellular termini. S1-S4 form the voltage sensor part of the channel whereas S5-S6 compose the selective pore. The amino acids in between S5 and S6 (red line) are responsible for the selectivity.

This doctoral thesis has focused on K_v channels. However, for the sake of completeness, we are going to briefly describe some other voltage-gated ion channel families:

- **Ca_v channels**: They are involved in many physiological processes such as muscular contraction and relaxation, heart beating, intracellular signalling and secretion of neurotransmitters and hormones. Every time these channels open, Ca^{2+} ions get inside the cell in favour of their electrochemical gradient. The pore-forming unit of all known Ca^{2+} channels is expressed as a single protein composed of four linked domains that are enzymatically separated during the production process. The α1 subunit comprises the pore and the voltage sensor has 190-250 kDa and it is arranged in four homologous motifs (I-IV), each of which owns six transmembrane domains. Some of them have associated an intracellular β subunit and a transmembranal α2 β subunit. Unlike K^+ and Na^+ channels, Ca^{2+} channels show even bigger structural and functional variation as it is noticeable, for instance, in the heart contraction. Indeed, 3 subfamilies are described to take part in every step of the heart beating: Ca_v1 (type L), Ca_v2 (type P/Q, N and R) and Ca_v3 (type T).

- **Na_v channels**: Even though those channels have an extreme selectivity towards Na^+ ion, there are structural similarities with Ca^{2+} channels. There are 11 genes that encode for Na^+ channels' α subunits in humans. Similar to Ca_v channels, a big amount of Na^+ ions crosses the membrane upon channel opening. Besides, they are as well known to be responsible for many physiological processes, for example the generation and propagation of the action potential.

The next channels are not as strictly voltage-dependent as all those mentioned before, but the show large homology with the voltage-gated ion channels.

- **Transient Receptor Potential (TRP)**: They are described as a superfamily of cationic channels. Their structure is quite similar to the previously described channels: they have six transmembrane domains in which segment five and six are responsible for forming the pore. They are organized in tetramers and they play an important role in sensory and pain sensing as well as in the transmission of temperature, pressure, organic compounds, mechanic stimuli or in the inflammatory response. Nonetheless, they don't have a big sensitivity to voltage changes. In the human genome we can find 28 members, classified in six families (TRPA, TRPC, TRPML, TRPP, TRPM and TRPV).

- **Channels activated by cyclic-nucleotides (CNG)**: Once more the CNG family shares the same structural characteristics with the other channels described. They are completely essential for the olfactory and visual perception. Moreover, this superfamily is divided in two main groups: channels activated by cyclic-nucleotides and channels activated by a hyperpolarization and mediated by cyclic-nucleotides (HCN). Those channels are cationic too, but they have more sensitivity and dependence to the voltage. Nevertheless, they are only activated when their ligands are bound to their intracellular carboxyl terminus (cAMP or cGMP).

1.3 VOLTAGE-GATED POTASSIUM CHANNELS (VGPC)

After the first *shaker* potassium channel was described in *Drosophila melanogaster* in the eighties, as much as 78 genes have been discovered until now. Moreover, 50 % of them depend on the membrane voltage to operate properly (Bradding & Wulff, 2013). According to the International Union of Basic and Clinical Pharmacology (www.iuphar.org), VGPCs are classified in twelve families (K_v1-12) (Figure 1.5) where it is noticeable that two main clusters could be discerned: the first group has inside K_v1 to K_v9 channels, in which K_v7 channels are separated rapidly; the second goes from K_v10 to K_v12. However, even considering that the number of VGPC genes are low, their diversity of the family grows exponentially when taking into account what happens *in vivo*:

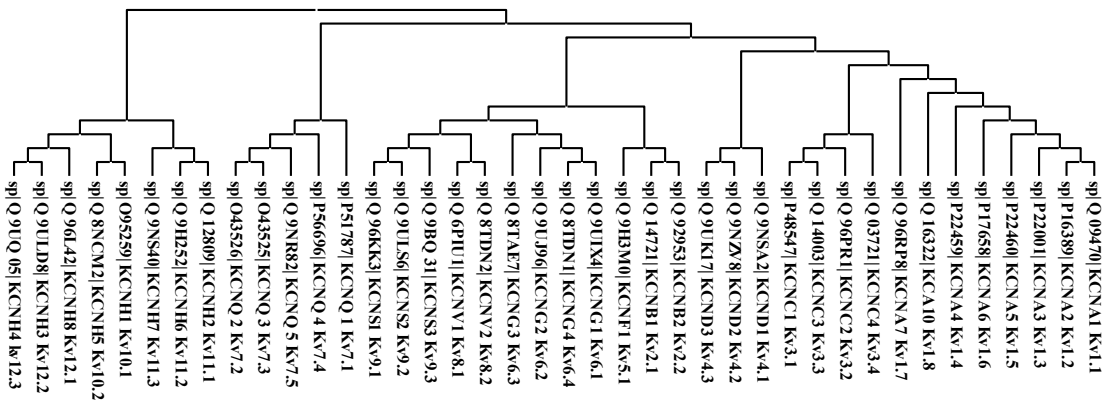


Figure 1.5. Maximum likehood tree of human K_v channels. The amino acid sequence of each were obtained using their accession number displayed below in www.uniprot.org database.

- **Heteromultimerization:** Those channels have the ability not only to combine with themselves; they can also build up functional channels when they blend with other members of the same family forming heterotetramers.

- **Alternative mRNA splicing:** For K_v3, 4, 6, 7, 9, 10 and 11, their coding genes have some exons and introns that can generate an alternative splicing of their mRNA.

- **Post-translational modifications** can lead to the phosphorylation, glycosylation, palmitoylation and ubiquitination of the channels modulating their properties.

- **Modifier subunits:** K_v5, 6, 8 and 9 channels alone are not functional but can be found fused with K_v2, which is a mechanism to improve their synthesis, functionality and diversity.

- **Accessory proteins:** The coupling with accessory proteins such us Calmodulin, AKAP proteins and other β regulatory subunits modulate VGPCs' functionalities

1.3.1 TRANSMEMBRANE DOMAINS OF VGPCS

The pore-forming domain is one of the main characteristic features, shared by all potassium channels. In order to form a functional pore, four pore-forming domains have to assemble together. Since this is a common feature this pore could be described perfectly by the two TM K^+ channels named KcsA. In fact, almost all the information we have about the pore structure relays on the work done by the Nobel Awardee MacKinnon on KcsA channels (Doyle et al., 1998)(PDB: 1K4C). As it is now clear, this structural information has paved the way for the scientific community to understands the selectivity for K^+ ions in a potassium channel amid the rest of ions (10000:1 K^+ to Na^+ ions), how is it able to transport such a large quantity of ions (10^7 ions $channel^{-1}$ s^{-1}) (Sansom et al., 2002) or how channel gating may occur.

1.3.1.1 Selectivity filter and conductivity

The mechanism of the **selectivity filter** is really elaborate (Figure 1.6). When K^+ is near the pore, it is electrostatically bound water molecules. Whenever for this ion to cross the membrane, it needs to break the intermolecular bond with water, which is energetically disadvantageous (around 75 kcal/mol)(Thompson et al., 2009). In order to overcome this problem the channel has disposed some amino acids along the pore region with their carboxyl group exposed to the outside, which are able to displace the water brought, ultimately producing the ion desolvation. Once the ion crosses the membrane it rehydrates again.

The amino acid sequence which completes the selectivity filter for KcsA is located in the P-loop and obeys to the consensus sequence TVGYG. It is worth mentioning that the GYG triplet is highly conserved among all potassium channels. Their carbonyl groups have their oxygens located facing the inner part of the pore (Figure 1.6.B). They constitute four sites to interact with the hydrated K^+ ion, two of them prepared to interact with two water molecules while the other two interacts with two ions in an alternate way. When a third ion reaches the outside part of the pore, it generates new electrostatic interactions that destabilize previous ions favouring their cross through the membrane.

It is almost impossible for anions to pass through the channel because the negative charges of the carbonyl groups would repeal them. Moreover, other than K^+ cations can get inside the pore but, because of their size differences, they do not match the carboxyl groups as well as the K^+ ions do, making the ion passing energetically unfavourable.

How is it possible, though, that the ion **conductivity** process happens so rapidly? MacKinnon and colleges claim that only two K^+ ions can go by simultaneously through the filter. If we hypothetically numbered the ions from the top to the bottom from 1 to 7, it is believed that only pair or odd configurations can be achieved. Shifting from the first to the second configuration is energetically allowed through a mechanism where top dehydrated ions' hydration favours

dehydration of bottom ions. Moreover, K^+ ions stabilize the open conformation of the selective filter which in consequence favours K^+ conduction.

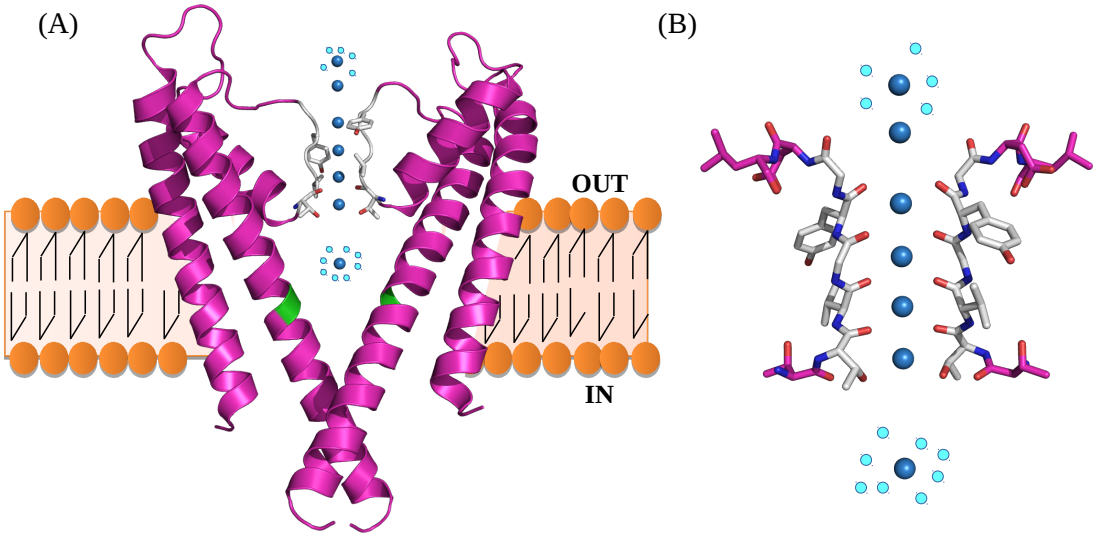


Figure 1.6. Structure of KcsA. (A) Representation of three of the four monomers forming the channel. In green has been highlighted the glycine 99 which take the function of the gating-hinge. In white TVGYG selective filter is highlighted. Dark-blue spheres are K^+ ions and cyan spheres are H_2O molecules and have been placed there to empathise the hydration of ions. (B) Pore region of the channel. The TVGYG motif of the P-loop appears in white. Lateral chains of those amino acids are able to create polar contacts with water and K^+ ions. PDB: 1K4C.

1.3.1.2 Voltage-sensor domain

Action potentials generated by nerve cells depends on several types of voltage-gated channels. Voltage-sensor domain is the molecular component of these type of channels responsible for the channel's response to voltage changes. This domain includes the first four helices from the all transmembrane domains (S1-S4). Throughout the S4 helix, there are located many positively charged Arg or Lys, one out of three amino acids to be more precise.

With the new high resolution structures of K_v channels (PDB: 2A79, 2R9R and 3LUT) several questions about voltage-signal translation to the pore-forming domain found an explanation (Figure 1.7). Structural data is consistent with amino acids in the S4 suffering a rotation and a displacement, changing subsequently of conformation helix S6 and ultimately resulting in the opening of the channel (Bezanilla et al., 2000). Moreover, it seems plausible for S4 to change its structure from a very horizontal position to a vertical one when it detects a depolarization, displacing its charges from the inner part to the bilayer membrane (Figure 1.7D). Furthermore, negatively charged residues located in S2 and S3 could take part on the stabilization of the new conformation. Accordingly, when S4 detects the voltage, segmental motion is produced, from the S4-S5 linker to S5 and then S6, which opens the pore.

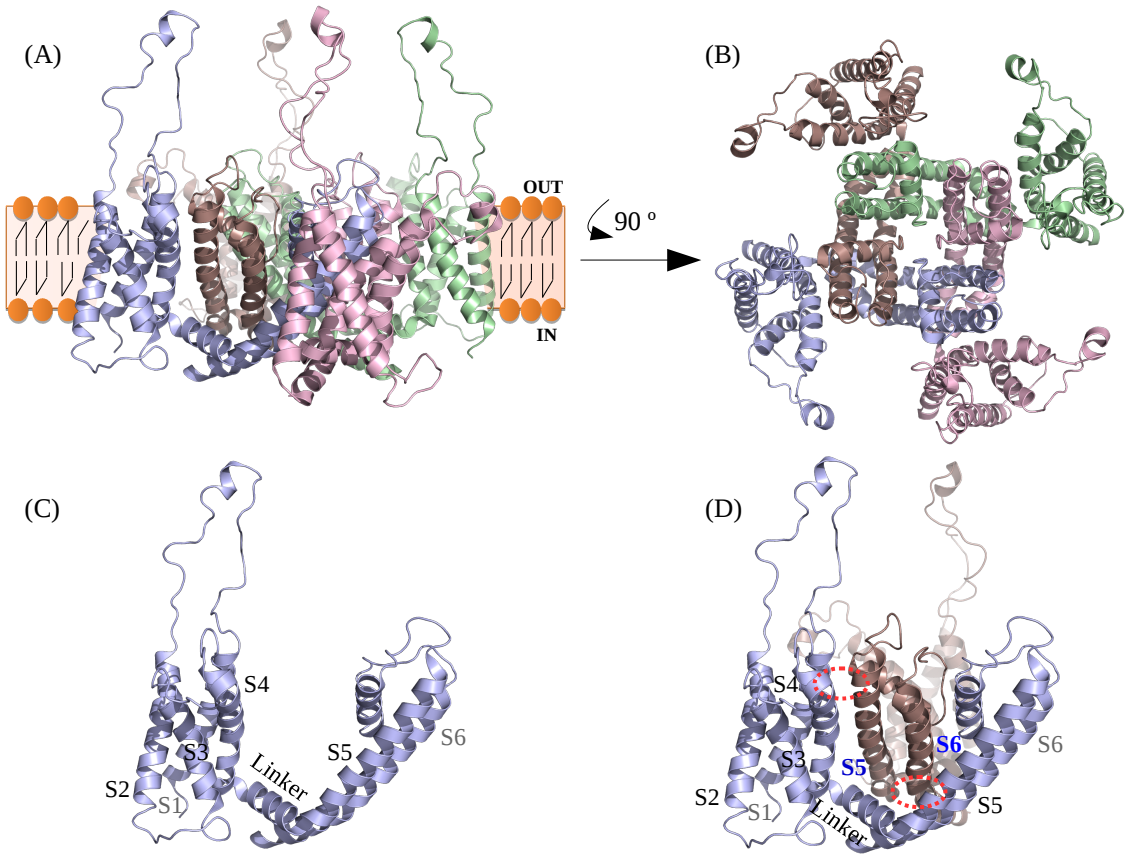


Figure 1.7. Structure of K_v1.2. (A) Lateral view of the tetramer. Every subunit has been shown with a different colour. Only the transmembrane domains are shown. (B) Upper view of the channel where a central pore is visible. All selection filters are looking towards the inner part of the pore. (C) Structure of a single subunit. This includes six transmembrane. Linker between S4-S5 serves as a bridge from the voltage sensing region (S1-S4) and the pore (S5 and S6). (D) Intermolecular contact regions surrounded in red circles. By those contacts voltage-dependent gating occurs in the channel. This regulation is among all the subunits. PDB: 3LUT.

1.3.1.3 Gating

VGPC's gating two proposed main mechanisms:

- Mechanism 1: In the case of KcsA, three crucial hydrophobic residues have been detected in the final part of the second transmembrane domain: V, T and A. Their hydrophobic moieties are located towards the entry of the pore, blocking it. When the pore has to open, these residues move thanks to a glycine which serves as a hinge (Figure 1.6.A). Nonetheless, this glycine is not conserved in all the channels and instead, some evolutionary conserved amino acid sequences can also act as a hinge: PxP, where P is proline and X is any animo acid. In K_v channels, this mechanism also allows the direct interaction between the sixth helix of the pore-forming domain and the voltage sensor domain (Figure 1.7.D). By doing this, the voltage sensor domain allosterically transfers the signal up to the pore domain (Long et al., 2005b).

- Mechanism 2: The inactivating mechanisms are independent blockage systems. Type N inactivation, also known as “ball-and-chain” mechanism, has a specific region in the N-termini of the channel that is able to insert in the inner part of the pore when this is opened, causing a transitory current. Other regulatory subunits such as $K_v\beta$ can mimic the “ball” region when this is not present in the channel. C-type inactivation, however, is slower and it seems to occur due to the movement of the residues located in the selectivity filter (Batulan et al., 2010). Apparently, this method is sensitive to the extracellular K^+ concentrations, which means that this could regulate an excess of K^+ ions.

It is also of interest to understand how different compounds or toxins are able to modulate the gating properties of the channels. For a long time it has been known that some animals such as spiders, snakes, scorpion among others have a huge variety of toxins that affects the neural system of other animals in which humans are included (Strong et al., 1990). Originally, it was thought that some of these toxins were specially affecting ion-channels. This idea was supported by physiological studies, but more recent studies where KcsA was analysed in presence of kaliotoxin (a scorpion toxin) by solid-state NMR revealed some structural rearrangements that can explain the blockage of channels (Lange et al., 2006). In 2013, the group of MacKinnon published a structure of a K_v1.2-2.1 chimera in complex with the scorpion toxin Charybdotoxin (PDB: 4JTA)(Banerjee et al., 2013). The toxin is plugging the ion conduction pathway of the pore rather than changing the channel structure *per se*. Despite the discrepancy with the solid state NMR data, there is no evidence for a unique way in which toxins act on channels and therefore both mechanisms are plausible.

1.3.2 INTRACELLULAR DOMAINS

Despite the high resolution structures of the transmembrane domains from voltage-gated potassium channels, there are no structures of the entire cytoplasmatic domains yet available. The scenario here is more challenging because amino acid sequences for these regions are less conserved than the transmembrane domains, even between closely related family members. Considering all available information, clustering K_v7 channels in two main groups has been suggested (Barros et al., 2012): 1) K_v7 channels with the tetramerization domain in the N-terminal (K_v1-K_v4) and 2) K_v7 channels with the tetramerization domain in the C-terminal domain (K_v7 and K_v10-K_v12).

Several parts of the N-terminus have been resolved and are referred here: **T1 domain** of Shaker (PDB: 1A68)(Kreusch et al., 1998), K_v1.2 (PDB: 1QDW)(Minor et al., 2000), K_v4.3 (PDB: 1S1G)(Scannevin et al., 2004), K_v4.3 with KchIp1 (PDB: 2I2R)(Pioletti et al., 2006), K_v1.3 (PDB: 4BGC)(Kremer et al., 2013). **Distal N-terminus activation structures** of K_v3.4 (PDB: 1ZTO)(Antz et al., 1997), K_v1.4 (PDB: 1KN7)(Wissmann et al., 2003). **PAS domain** of K_v11.1

(PDB: 1BYW)(Morais Cabral et al., 1998), (PDB: 2LOW)(Ng et al., 2011), (PDB: 4HQA & 4HP9)(Adaixo et al., 2013).

Some of the C-terminus have been resolved too. **Coiled-coil segment** of K_v7.4 (PDB: 2OVC) (Howard et al., 2007), K_v7.1 (PDB: 3BJ4)(Wiener et al., 2008). K_v7.1 channel's **helix B in complex with holo-CaM** (PDB: 4GOW)(Q. Xu et al., 2013) and finally the K_v7.1 channel's **helices A and B in complex with CaM** (PDB: 4UMO & 4V0C)(Sachyani et al., 2014).

Even when considering the large divergence among the intracellular domains, a general structural model of K_v7 channels can be devised and it is shown in Figure 1.8. The T1 tetramerization domain of K_v1-K_v4 channels, which is also called NAB, is located in the **N-terminal.** Moreover K_v1.5 channels have a tandem of two SH3 binding domains whereas a secondary inactivation domain is always present in K_v1.4 channels. The N-type inactivation-prevention domain (NIP) of K_v1.6 channels is also represented as it protects potassium channels against rapid inactivation. Finally, in the very beginning of the N-terminus the ball-like structure is located, responsible of the N-type inactivation and the PAS domain of eag-like channels. In the **C-terminal** domain a C-linker/cNBD regions of the K_v related HCN and CNG channels can be found. The localization and C-terminal activation (CTA) domains corresponds to K_v2.1. Finally, a post-synaptic density protein binding domain (PSD-95) is located in the most distal part of the C-terminus which comes from some K_v1 channels. However, a more detailed view of K_v7 channels and the implications of the model in M-current density regulation will be discussed later on.

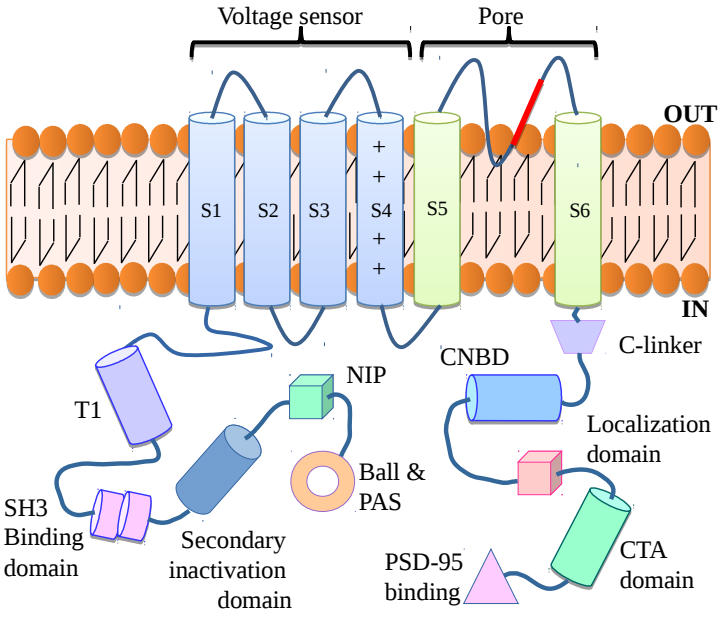


Figure 1.8. Schematic representation of the intracellular domains of K_v channels. Different regions and domains have been highlighted. In the N-terminal, *T1*, tetramerization domain of K_v1-K_v4 channels, *SH3* binding domain from K_v1.5; *Secondary inactivation domain*, from K_v1.4; *NIP*, N-type inactivation-prevention domain from K_v1.6; *Ball-like* structure and *PAS* domain from eag-like channels. In the C-terminal, *C-linker/cNBD* regions from K_v related HCN and CNG channels; *Localization* and *CTA (*C-terminal activation) domains from Kv2.1; and *PSD-95*, post-synaptic density protein binding domain from some K_v1 channels.

1.3.3 CHANNELOPATHIES

This term refers to all the diseases caused by a malfunction of ion channels. The discovery of missense mutations in genes encoding for K_v channels has underscored the vast role diversity channels. Mutations may affect protein expression, trafficking, protein folding or regulation but not very often the ion channel function *per se*. As VGPCs are responsible of pumping positive charges out of the cell, a malfunction of it often causes cellular hyperexcitability. In Table 1.1 some examples of channelopaties are reported.

Protein	Gene	Disease
K_v1.1	KCNA1	Episodic ataxia myokymia
K_v7.1	KCNQ1	Autosomal-dominant long-QT syndrome with deafness Autosomal-dominant long-QT syndrome
K_v7.2	KCNQ2	Benign familial neonatal convulsions (BFNC), also with myokymia
K_v7.3	KCNQ3	Benign familial neonatal convulsions (BFNC)
K_v7.4	KCNQ4	Autosomal-dominant deafness
KCNH2	KCNH2	Long-QT syndrome
Kir1.1/ROMK	KCNJ1	Bartter syndrome
Kir2.1	KCNJ2	Long-QT syndrome with dysmorphic features
Kir6.2	KCNJ11	Persistent hyperinsulinaemic hypoglycaemia of infancy diabetes mellitus
SUR1	SUR1	Persistent hyperinsulinaemic hypoglycaemia of infancy
SUR2	SUR2	Dilated cardiomyopathy
KCNE1	KCNE1	Autosomal-dominant long-QT syndrome with deafness Autosomal-dominant long-QT syndrome
KCNE2	KCNE2	Long-QT syndrome
KCNE3	KCNE3	Hypokalaemic periodic paralysis

Table 1.1. Potassium channels and human diseases. Table of the protein causing the disease, the gene which encodes the protein, and finally the caused disease.

It is important to emphasize that, among all VGPC families, the most frequent one is K_v7 family. In fact, only one out of five members (K_v7.5) has never been related to a specific disease. So, what is so interesting about these channels that are so implicated in the development of different diseases?

1.4 K_v7 CHANNELS AND M-CURRENT

In 1980 a voltage-dependent potassium current was first discovered, which was described to be regulated by second messengers (Brown & Adams, 1980). The name *"M-current"* was designated because at that time it was only known that muscarine was able to suppress it (Figure 1.9).

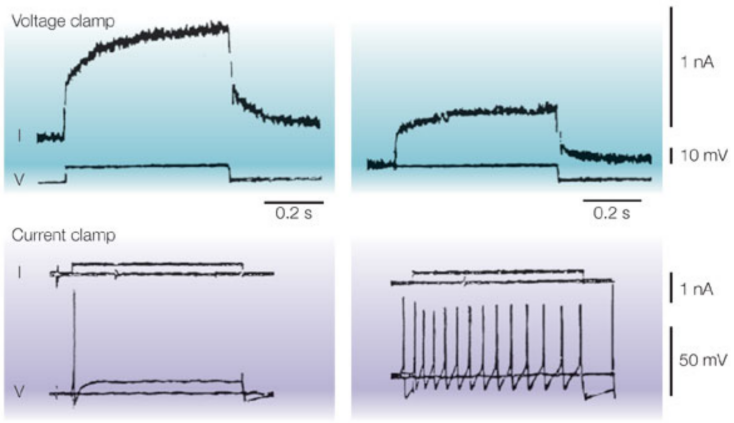


Figure 1.9. Neuronal response to a stimulus in absence and in presence of muscarine. Left: After depolarizing the neuron from the resting potential to -40 mV produces the M-current where only a single spike appears. Right: When $K_v7.2$ and $K_v7.3$ are closed due to the muscarinic effect, M-current is suppressed and therefore a depolarizing stimulus can generate a train of action potentials. Adapted from (Delmas & Brown, 2005)

Immediately, K_v7 channels raised much interest within the scientific community, and only few years after the discovery it was already known that I_M is responsible for the control of the neuronal excitability of the peripheral and central nervous system. In fact, the M-current represents the most important potassium current in the sub-threshold area of the membrane potential, restricting the excitability and the trigger of repetitive APs. Moreover, this current never inactivates, so it takes part on the adjusting of the firing rate.

It wasn't until 1998 that sufficient evidence could support that the $K_v7.2$ and $K_v7.3$ heterotetramers are the molecular components responsible for the current (Wang et al., 1998), and that mutations on these subunits are resonsible for diseases such as BFNC (Biervert et al., 1998; Cooper et al., 2000; Delmas & Brown, 2005). Actually, both channels are expressed in the central and peripheral nervous system overlapping each other. Nevertheless, $K_v7.2$ and $K_v7.3$ often change their proportion within the channel forming tetramer, subsequently altering the M-current properties as a function of the cerebral region in which they have been expressed.

Within the neuron, $K_v7.2$ and $K_v7.3$ channels (encoded by KCNQ2 and KCNQ3 genes respectively) gather in the soma, dendrites, in the initial segment of the axon and finally in the *Ranvier nodes* where the AP regenerates in myelinated nerves. The high availability of both channels in these regions highlights the importance of the I_M in the regulation of neuronal excitability. Nonetheless, it has also been described that even though M-currents are mostly generated by $K_v7.2$ and $K_v7.3$, homomers of $K_v7.2$ and $K_v7.5$ could produce M-currents *in vivo* (Schwarz et al., 2006), whereas in cellular models such as HEK293T and CHO any K_v7 could produce it (Shapiro et al., 2000).

1.4.1 K_v7 FAMILY MEMBERS

1.4.1.1 Characteristics

All K_v7 family members share the typical structure of six transmembrane domains of voltage dependent potassium selective channels (TM/S1-6). S1-S4 segments conform the voltage sensing region, specially the arginine rich, and so positively charged, S4 segment. S5, S6 and the loop connecting them form the pore, which contains the typical signature GYG which confers the selectivity to potassium. Amino and carboxy terminal regions are positioned intracellularly. Four subunits have to assemble to form the functional pore. However, while the majority of K_v channels tetramerize at the N-terminus via a tetramerization domain (T1), K_v7 channels do not have this domain. Instead, the subunit interaction occurs through a localized domain at the C- terminus (Maljevic et al., 2003). It is worth highlighting that the C-terminus of K_v7 channels is exceptionally long and contains many regulatory domains.

1.4.1.2 $K_v7.1$

The gene encoding for $K_v7.1$ protein, KCNQ1, was the first family member to be cloned and it was identified on chromosome 11p15.5 by using a positional cloning approach in families with long QT syndrome type 1 (Wang et al., 1996). As with the rest of K_v7 members, it does assemble into tetramers to generate a functional pore, but $K_v7.1$ is the only member of the family that cannot form heterotetramers with other K_v7 members. Nonetheless it can co-assemble with the auxiliary subunit KCNE1, also known as mink or IsK, to create the cardiac IKs current, which plays an important role in cardiac cell repolarization (Barhanin et al., 1996; Sanguinetti et al., 1996). $K_v7.1$/KCNE1 are also expressed in the inner ear, thyroid gland, lung, gastrointestinal tract, the small intestine, pancreas, forebrain neuronal networks and brain stem nuclei, ovaries and the proximal and distal tubule of nephron, suggesting the importance of these channels in a lot of cellular functions (Jespersen et al., 2005).

Long QT syndrome (LQTS) is a cardiac disorder that affects cell repolarization, often causing strong suppression of WT currents and predisposes the affected individuals to arrhythmias and cardiac sudden death (Maljevic et al., 2010). Two different syndromes have been associated to KCNQ1 mutations: autosomal dominant Romano-Ward and the recessive Jervell and Lange-Nielsen syndrome. In the later one, patients suffer from congenital deafness as well as the already mentioned cardiac deficiency.

1.4.1.3 $K_v7.2$ and $K_v7.3$

Genes encoding for $K_v7.2$ and $K_v7.3$ channels, KCNQ2 and KCNQ3, were identified by using two different approaches. At first one, based on a KCNQ1-derived sequence, screened a human brain cDNA library (Yang et al., 1998) while, alternatively, positional cloning was explored in families with BFNC (Biervert et al., 1998; Schroeder et al., 1998). $K_v7.2$ and $K_v7.3$ are

expressed in different brain regions and can form both homo and heterotetramers. They underlie the neuronal M-current and mutations in these proteins have been related to a series of diseases such as neonatal seizures, Rolandic epilepsy (Neubauer et al., 2008), or epileptic encephalopathy (Weckhuysen et al., 2012). Most of the mutations are located in the KCNQ2 gene, but some polymorphisms in KCNQ3 have also been associated with autism (Gilling et al., 2013).

1.4.1.4 K_v7.4

The KCNQ4 gene was identified from a human retina cDNA library using a KCNQ3 partial cDNA. An inherited autosomal dominant form of nonsyndromic progressive hearing loss was also identified to co-segregate with this gene (Jentsch et al., 2000). Low expression is found in the brain, mainly restricted to nuclei contributing to the central auditory pathway (Kharkovets et al., 2000), but it is predominantly expressed in outer hair cells of the inner ear (Kubisch et al., 1999). DFNA2 mutations show a loss of function either by a haploinsufficiency mechanism or by a dominant negative effect (Maljevic et al., 2010). Finally, K_v7.4 can coassemble with K_v7.3 (Kubisch et al., 1999).

1.4.1.5 K_v7.5

KCNQ5 was the last member to be cloned and is the less characterized, although it shows a large homology to KCNQ3 in a brain cDNA library (Schroeder et al., 2000). It is mainly expressed in the brain and in the skeletal muscle, being regulated during myoblast proliferation (Lerche et al., 2000). K_v7.5 was also found in colon, lungs and uterus. As other K_v7 channels, it can be inhibited by muscarine receptors so, this member of the K_v7 family is also integrated in the M-channel family. They can also form heteromultimers with K_v7.2 and K_v7.3 channels, which are expressed in the nervous central system and peripheral ganglions. The assembly with K_v7.3 produces an increase on the current and provokes small changes on the activation kinetics. Up to date, no diseases have been associated to mutations in K_v7.5. It has also been demonstrated that K_v7.5 can interact with KCNE peptides (KCNE1, KCNE3), increasing the diversity of K^+ currents (Roura-Ferrer et al., 2009).

1.4.2 Regulation Of The M-Current in $K_v7.2$ And $K_v7.3$

The carboxyl-terminal is extensive (320-560 amino acids), varying its length in the following order: $K_v5 > K_v2 > K_v3 > K_v4 > K_v1$. This contrasts to the short intracellular amino end of those channels (~100 amino acids). Although no crystal structure of the whole C-terminus has been yet obtained, the secondary structure predicting programs have revealed that different α-helix, coiled-coil and basic amino acids clusters are present. C-terminal becomes a place where channel trafficking and assembling is controlled inasmuch as other accessory molecules interact and modulate it, turning the region as a "multi-modular" domain (Sachyani et al., 2014).

First analysis revealed that the probability for all K_v7 channels to form four α-helices was large, and therefore helices were generically named A to D (Yus-Najera et al., 2002). Helices are separated by several chains of amino acids with different lengths (Figure 1.10).

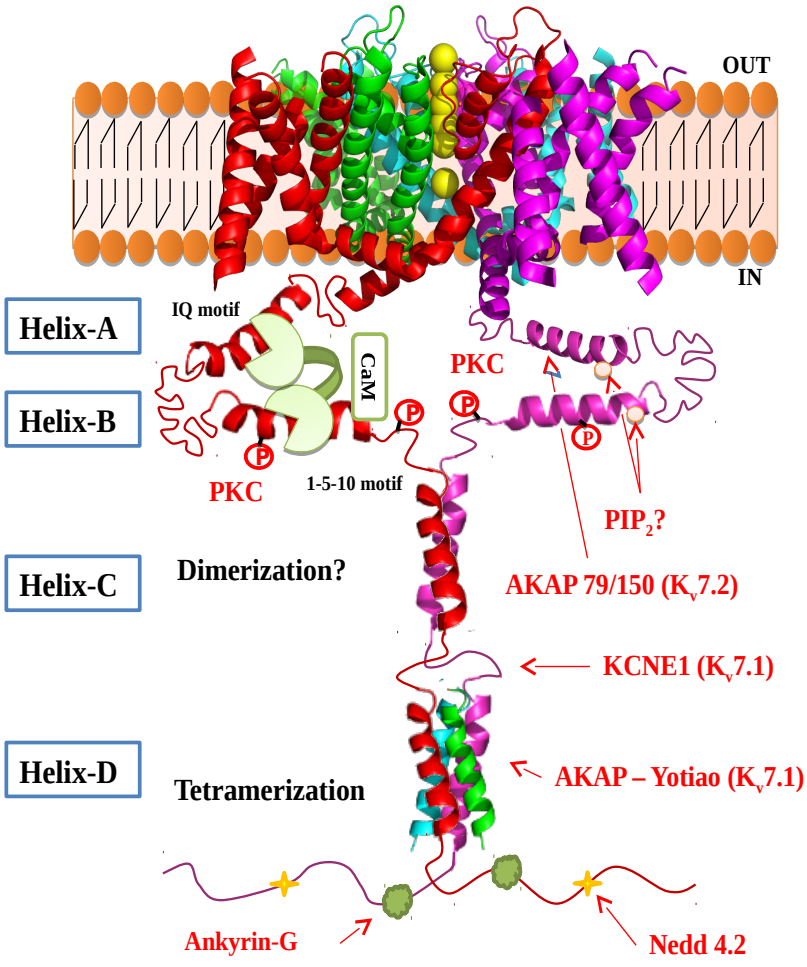


Figure 1.10. Scheme of the C-terminal of K_v7 channels. Two of the four subunits are represented. A and B helix have conserved interaction domain for CaM. PIP_2 interaction sites may exist in those helix too. AKAP-yotiao in $K_V7.1$, AKAP79/150 in $K_V7.2$, Ankyrin-G in $K_V7.2/3$ and Nedd4.2 in $K_V7.1$-3 also interact with K_v channels. Several phosphorylation sites have been reported too. Helix C complex may undergo dimerization whereas helix D tetramers have been seen in the crystal structure.

1.4.2.1 Helix A and B:

Helix A is located just after the sixth transmembrane segment, which corresponds to V320 to Y372 in the case of K_v7.2. Forthcoming this segment, a linker of about a hundred amino acids separates the second helix B which encompasses the amino acids from T501 to E529. Even if there is not a structure of this particular region of K_v7.2 channels, in late 2014 a crystal structure of CaM bound to helices A and B of K_v7.1 was released (Sachyani et al., 2014)(PDB: 4UMO/4V0C). Due to its relevance to our work, more details about the crystal structure are discussed in chapter 3 and 4.

- Calmodulin (CaM)

In 2002 it was demonstrated that CaM interacts with all the members of K_v7. Two different biding epitopes have been described: 1) In helix A, the IQ motif (**IQ**xxx**R**xxxx**R**) and 2) in helix B, two overlapping 1-5-10 motifs (**L**xxx**I**xxxx**V** and **M**xxx**V**xxxx**F**) (Table 1.2).

	Helix A	Helix B
K_v7.1	AASL**IQ**TAW**R**CYAAENPDS	EH**H**RAT**I**KVIRR**M**QYF**V**AKKK**F**QQARK
K_v7.2	AAGL**IQ**SAW**R**FYATNLSRT	PG**L**KVS**I**RAVC**VM**RFL**V**SKRK**F**KESLR
K_v7.3	AAEL**IQ**AAW**R**YYATNPNRI	PT**L**KAA**I**RAVR**IL**QFR**L**YKKK**F**KETLR
K_v7.4	AANL**IQ**SAW**R**LYSTDMSRA	PT**L**KAA**I**RSIR**IL**KFL**V**AKRK**F**KETLR
K_v7.5	AANL**IQ**CVW**R**SYAADEKSV	PP**L**KTV**I**RAIR**IM**KFH**V**AKRK**F**KETLR
	xxxx**IQ**xxx**R**xxxx**R**xxxx	xx**L**xxx**I**xxxx**V**xxx xxx**M**xxx**V**xxxx**F**xxx
	IQ motif	**1-5-10**

Table 1.2. Sequence alignment of K_v7 channels' helix A and helix B. In bold there are highlighted conserved amino acids compiling the CaM binding motifs.

CaM is a Ca^{2+} modulated protein which serves as a sensor of calcium for the organism. In fact, it takes part a plethora of calcium regulated processes, controlling the downstream of many other proteins' duties as muscle contraction, nerve signalling, fertilization and cell division. Calmodulin is a small protein of 148 amino acids (16.7 kDa), soluble, thermostable and acidic (pI ~4). Its structure is dumbbell-shaped composed of two globular domains connected by a flexible linker. Each domain has two high-affinity Ca^{2+} binding motifs known as EF hands, which are composed of a characteristic loop flanked by two α helices. When Ca^{2+} is not available, CaM has a “closed” conformation hiding all the hydrophobic residues (Babu et al., 1988)(PDB: 3CLN), but when Ca^{2+} is present in the cell, every single EF hand will “sequester” one ion.

Typically, CaM wraps around its target proteins holding them tightly with both globular domains when it recognize specific motifs. However, the high functional of this molecule regarding its ability to recognize and regulate proteins couldn't be explained if it wasn't for its

huge structural plasticity. Actually, CaM can adopt different conformations depending on the [Ca^{2+}] because each EF hand has a different affinity for this ion. Moreover, the linker in between both globules is very flexible expanding even more the conformational landscape of the molecule. Therefore, under specific circumstances and within the cellular context it has been reported that CaM is capable of recognizing as far as 300 proteins (Maylie et al., 2004).

Nevertheless, it is still not well established how CaM interact with K_v7 channels. It is commonly thought that calmodulin acts as a constitutively bound auxiliary subunit and that it is essential for their function. Apparently, the presence of calcium is not necessary for CaM to bind the C-terminal of the channels. In fact, originally it was thought that the CaM coupling to the channel was constitutive (Wen & Levitan, 2002). In this context, the presence of calcium could increase (Hernandez et al., 2008) or decrease (Yus-Najera et al., 2002) CaM binding to the channel. When considering the activity regulation of these channels, Ca^{2+}-CaM can inhibit currents of $K_v7.2$, $K_v7.4$ and $K_v7.5$, but not $K_v7.1$ and $K_v7.3$ (Delmas et al., 2005). Nevertheless, CaM still plays an important role for $K_v7.1$ and $K_v7.3$ channels because it has been shown that this protein is necessary for channels folding, assembling and intracellular trafficking towards the membrane (Etxeberria et al., 2008; Spratt et al., 2006). More information about CaM is provided in Section 1.1.5.

- <u>Phosphatidylinositol 4.5 bisphosphate (PIP_2) binding and G protein coupled regulation</u>

When the membrane phosphatidylinositol is phosphorilated at positions 3, 4 and/or 5 of the inositol groups seven different isomers can be generated. They are considered to be a minority lipids in plasma membrane although they are not equally distributed throughout the cell. All of them can be rapidly interconverted by kinases and phosphatases (Hille et al., 2015).

One special feature of K_v7 channels is that they are inhibited by neurotransmitters that act on G_q protein coupled receptors (GPCR). PIP_2, the most abundant phosphoinositide present in the plasma membrane, is involved in the regulation of these channels, since PIP_2 synthesis inhibition prevents the recovery of the M currents (Suh & Hille, 2002). Apparently, PIP_2 stabilizes the opened conformation of the pore without affecting its conductance. In any case, every K_v7 shows a characteristic affinity for the lipid: $K_v7.3 = K_v7.1 > K_v7.2/7.3 > K_v7.2 > K_v7.4 > K_v7.5$.

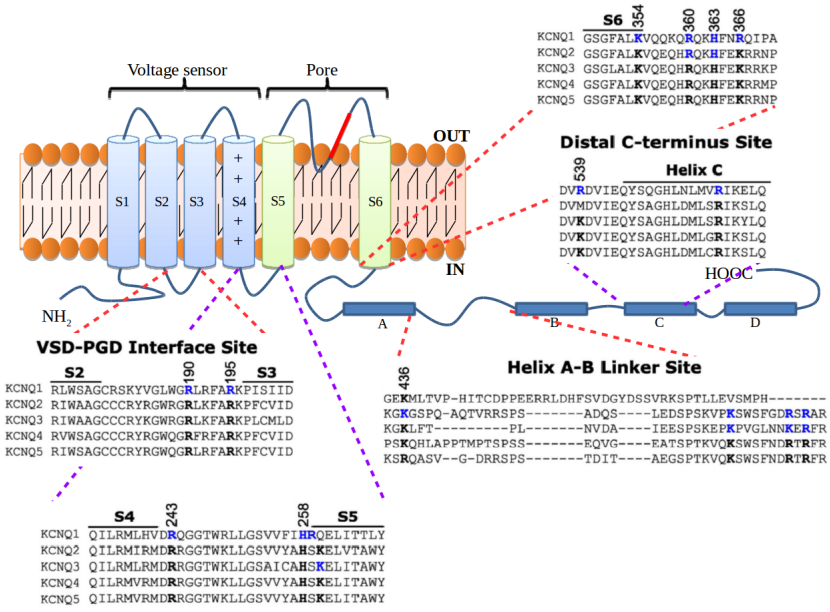


Figure 1.11. Putative PIP_2 interaction sites on K_v7 channels. Blue amino acids represents hot spots where mutations have been reported to affect PIP_2 dependent channel activation. Black amino acids are conserved among other members of the family. Numbers indicates residue number. Adapted from (Zaydman et al., 2014)

Despite the effort, the interaction site the interaction site within the C-terminal domain remains unknown. Based on other lipid-protein interactions, the interaction between the phosphoinositides and protein is mainly electrostatic: the negative charges of the lipid's polar head interact with positively charged amino acids in the protein. When analysing the K_v7 carboxyl-termini, several clusters of basic and hydrophobic amino acids can be found (Figure 1.11). Based on that, two recent investigations claim that interacting epitopes may locate between helices A and B (Hernandez et al., 2008). However, this is in conflict with some functional data showing that deletion of the linker still yields functional channels with unaltered PIP_2 sensitivity, and it is likely that there may be more than one positively charged clusters in which the interaction might occur (Aivar et al., 2012; Zaydman et al., 2014).

It is worth to reconsidering whether the potential PIP_2 interacting domain and the CaM domain may physically overlap, and therefore both element would compete for the same locus. In the case of $K_v7.1$, there are some evidences supporting this hypothesis in which PIP_2 and PIP_3 attach to helix A and B and they shift CaM from its interacting domain (Kwon et al., 2007).

- <u>M-current regulation in which either Calmodulin and PIP are involved</u>
    - Muscarinic and angiotensin II AT_1 regulation

When acetylcholine binds to M1 muscarinic receptors that are coupled to Gq/11, PLC is activated and, subsequently, PIP_2 is cleaved generating diacylglycerol (DAG) and inositol-3-phosphate (IP_3) (Figure 1.12). In the same way, DAG also activates PKC. Often IP_3 reaches IP_3-receptors in the endoplasmatic reticulum where they are activated releasing Ca^{2+}, but without

raising the Ca^{2+} levels. Depletion of PIP_2 is ultimately responsible of the inhibition of M current (Hernandez et al., 2008), and the time course of K_v7 current recovery and the resynthesis of the PIP_2 do perfectly match reinforcing this idea (Falkenburger et al., 2013). On the other hand, it has also been shown that the muscarinic stimulation of neuroblastoma cells depletes up to 75 % of PIP_2 within 30 to 60 seconds (Xu et al., 2003), similarly to Angiotensin II AT1 activation (Zaika et al., 2006).

As mentioned before, channels show different affinity for PIP_2, in the order $K_v7.3 > K_v7.2 > K_v7.4$. The EC_{50} values range from three to hundreds of μM, with intermediate values for $K_v7.2/3$ channel. This values help understands why at resting conditions, the open probability of $K_v7.3$ (which displays a high-affinity for PIP_2) is almost one; however this membrane PIP_2 levels are too low for $K_v7.2$ (which shows small- affinity for PIP_2). Considering that current channels are formed by different subunits (homomeric or heteromeric) at different neuronal loci, differences in PIP_2 affinity may result in distinct inhibition levels in response to different neurotransmitters (Delmas et al., 2005).

- Calcium and Bardykinin / Purinergic inhibition

Under resting conditions Ca^{2+} levels in cell are around ~0.1 μM. Upon PIP_2 hydrolysis IP_3 is translocated to the cytoplasm and releases Ca^{2+} form intracellular stores as described previously, which has a direct effect in K_v7 currents (Figure 1.12). Other studies have demonstrated that in sympathetic neurons the Ca^{2+} IC_{50} is ~100 nM, only fractionally above of the resting Ca^{2+} concentration of ~70-80 nM (Selyanko & Brown, 1996).

Bradykinin receptors and purinergic P2Y are also coupled to Gq/11 proteins, leading to the activation of PLC generating DAG and IP_3 upon hydrolysis of PIP_2. However, unlike the muscarinic regulation, bradykinin inhibition can be prevented by buffering intracellular Ca^{2+} (Cruzblanca et al, 1998).

So, how is possible that either muscarinic and bradykinin inhibition activate the same PLC while the second messenger is different? The rapid Ca^{2+} rise upon bradykinin addition activates PIP_2 re-synthesis, through the activation of neuronal Ca^{2+} sensor protein (NCS1) which finally activates PI4K adding new PIP_2 to the membrane (Santagata et al., 2001). In the case of the muscarinic inhibition the transient increase of Ca^{2+} in the submembrane does not reach the requiered levels to generate a response of NCS1 and therefore the membrane PIP_2 levels are not rapidly restored.

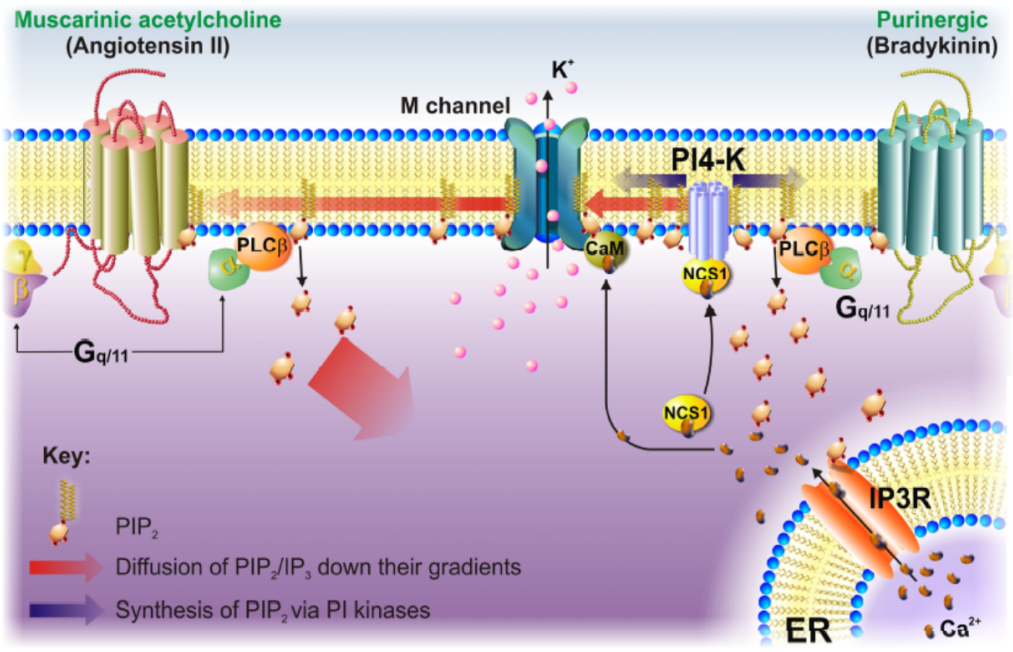


Figure 1.12. M-current inhibition in sympathetic neurons. On the left muscarinic aceltylcholine and AT_1 angiotensin II receptor, where even if PLC activation creates IP_3 there is not activation of Ca^{2+} signals, propably because receptor are situated far from the ER. Nevertheless, PIP_2 depletion causes M channels inhibition. On the right bradykinin B2 and purinergic P2Y6 recerpertors, where they colocalise spatially with ER creating Ca^{2+} signals. The increase of Ca^{2+} binds to NCS1 and CaM. Taken from Hernandez et al., 2008

- AKAP79/150 and PKC

A-Kinase Anchoring Protein (AKAP) family handles the coordination of molecular complexes between phosphorylate channels and receptors. At the helix A and in the intermediate helix it has been defined an interacting domain for AKAP79/150, whereas helices C and D has been suggested as the $K_v7.2$-AKAP-PKC forming point. The main effect of these phosphorylations is that M-channels' currents decrease in presence of muscarinic agonists (Kosenko et al., 2012).

Nevertheless, the situation becomes more complicated since AKAP can interact with PIP_2, CaM, Calcineurin and Protein Kinase A (PKA) as well. It has been hypothesized that these molecules can be organize a macrocomplex, which would eventually be involved in the regulation of K_v7.

1.4.2.2 Helices C and D

Although helices A and B encode the CaM binding domain and motifs for other auxiliary proteins, they are not able to form multimerize by their own. *Subunit interaction domain* (SID) was identified in helices C and D and it is involved in the exclusive assembling of different K_v7 subunits. Nonetheless, the molecular mechanisms that determine the K_v7 subunit assembly selectivity remain imperfectly understood. For example, it is well known that $K_v7.3$ is able to for heterotetramers $K_v7.2$ $K_v7.4$ and $K_v7.5$ but never with $K_v7.1$.

Within this domain, the helix C is the most conserved among all the channels. This region is delimited from M537 to V561 in $K_v7.2$, and it has been described to be absolutely necessary for

the assembling of homomers and heteromers. Helix D (R594-R619), contains a bigger and more variable region that can explain the restricted association between all the K_v7 subunits. The crystal structure of helix D (available for $K_v7.1$ and $K_v7.4$ channels) show a four-stranded parallel coiled coil as well as a self-assembling capacity (Howard et al., 2007). Moreover, helix D is not only important in the assembly between $K_v7.2$ and $K_v7.3$ but it also determines the expression increment on the cell surface (Wiener et al., 2008).

1.4.2.3 Distal extreme:

In this region, it has been found that $K_v7.2/3$ channels possess a binding motif for *ankyrin-G* (motif C3, ~10 amino acids). The interaction of the channel with this large adaptor is the mechanism for the selective retention of K_v7 at the axon initial segments and Ranvier Nodes, given by ankyrin-G bound to cytoskeletal actin. Therefore, it is believed that ankyrin-G is essential for the regulation of M channels' concentration, in these neuronal localizations.

For $K_v7.1$-3, it is also known that they interact with the ubiquitine ligase *Nedd4-2*. This interaction decreases ion fluxes in the complexes KCNQ1/KCNE1, KCNQ2/KCNQ3 and KCNQ3/KCNQ5 (Miranda et al., 2013)

1.4.2.4 Direct regulation of M current

- K_v7 inhibitors: Linopirdine and XE-991

Linopirdine is a reference blocker for M channels. This drug was first synthesised in the 80s and since then it has been used as a elementary tool. It increases cognitive performance of laboratory animals, by blocking the M-current and a subsequent increase in neurotransmitter release (Tam & Zaczek, 1995) and it was a potential drug for treatment of neurodegenerative disorders. Unfortunately, side effects observed during clinical trials precluded its pharmacological use. Instead, a second generation analogues of Linopirdine (XE-991) are currently widely used in *in vitro* and *in vivo* studies (Miceli et al., 2008).

- K_v7 activators:Flupirtine and Retigabine

M-current enhancement is an attractive mechanism for treating hyperexcitability diseases. In 1984 it was discover that Flupirtine actives M-current and since then it has been successfully used in Europe. Optimization of flupirtine yielded retigabine. In K_v7 channels (with the exception of $K_v7.1$), it causes a shift towards hyperpolarizing direction of the channel activation process (Miceli et al., 2008).

1.4.3 CHANNALOPATHIES RELATED TO $K_v7.2$ / $K_v7.3$

Numerous mutations in the $K_v7.2$ and $K_v7.3$ genes have been associated to the clinical phenotype of BFNC (Maljevic et al., 2008). With time, other different phenotypes have been also reported, such as peripheral nerve hyperexcitability, myokymia or Rolandic epilepsy. In addition, KCNQ2 mutations have recently been linked to children suffering refractory epilepsy and mental retardation introducing epileptic encephalopathy as a new clinical phenotype.

Usually BFNC starts before the fifth day of life of the newborn (Rett & Teubel, 1964) and it consists on generalized focal seizures, which have transient expression. After several weeks or even months seizures remit spontaneously. It has an autosomal dominant inheritance, with around 85 % of penetrance. About 15 % of patients may have recurring seizures later in life (Maljevic et al., 2010).

The mode in which those mutations have been studied has been in heterologous systems. Often, a reduction in the K^+ current has been reported, although sometimes there is a complete loss of functionality (Jentsch et al., 2000; Maljevic et al., 2008). Co-expression of WT $K_v7.2$ or $K_v7.3$ with a mutant subunit, in a 2:1:1 ratio, largely mimics the patients situation with one of the four alleles mutated, revealed that a mere 20-25 % reduction in current is sufficient to cause seizures (Schroeder et al., 1998). Families with more severe phenotype also display a mild reduction in the $K_v7.2/3$ current, suggesting that other factors may also be implicated in the final outcome of the disease (Steinlein et al., 2007).

To date, more than 130 described deleterious mutations in $K_v7.2$ are present at different parts of the protein including the cytoplasmic C-terminus, the pore region (S5-S6), voltage sensor (S4) and S1-S2 region. In contrast, for $K_v7.3$ only six mutations have been reported, five of which are located in the pore, while the last one is in the S6 (Soldovieri et al., 2014)(Figure 1.13).

As previously described, C-terminal domain posseses different regions for binding regulatory proteins, as well as the tetramerization domains. Mutations in this part of the protein possibly alter the channel functionality by impairing tetramerization and/or the surface expression (Etxeberria et al., 2008; Schwake et al., 2000). On the other hand, mutations in the pore region probably affect the conductance by a haploinsufficiency mechanism and the ones in S1-S2 and S4 have been linked to changes in channel gating (Soldovieri et al., 2007; Wuttke et al., 2008).

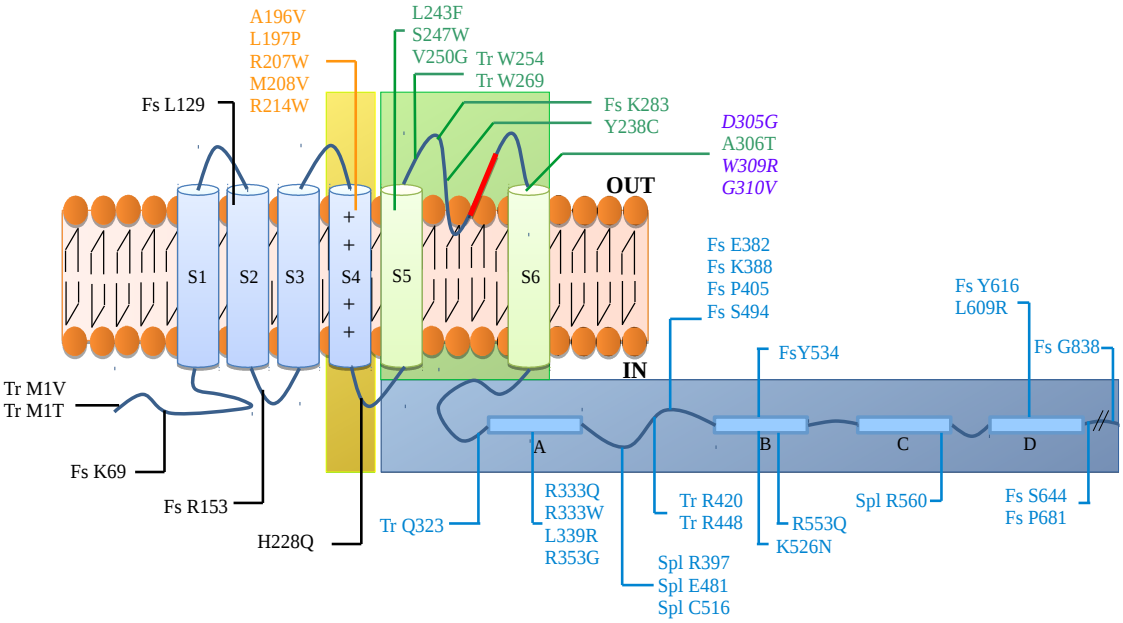


Figure 1.13. Localization of BFNC-linked mutations. Fs, frame shift mutation; Spl, alternative splicing variants; Tr, truncation. In purple and italic $K_v7.3$ related mutations are shown.

Nevertheless it is difficult to find a molecular explanation for the fact that the seizures disappear by themselves after a short period. Based on the up-regulation found in rodent brain within the first postnatal weeks, some scientist claim that the expression pattern $K_v7.2/3$ changes during development (Geiger et al., 2006; Maljevic et al., 2008; Weber et al., 2006). In that case, the expression of a mutant subunit of $K_v7.2/3$ channel would lead to a reduction in current that is insufficient for the normal functioning of the neurons at the first stages of the newborn. However, in adulthood, as K_v7 channels are more abundantly expressed and other K^+ channels are also upregulated, the overall increase in potassium conductance can compensate the small loss of function giving arise a normal phenotype. The proposed excitatory action of GABA in the immature brain could aggravate this effect until the proper inhibitory GABA system is completely developed (Okada et al., 2003).

1.5 CALCIUM SIGNALLING

Calcium ion is by far one of the most important intracellular messenger in cells. Either in prokaryotic or eukaryotic cells use this ion in order to regulate different activities such as gene expression, cell growth, development, migration, survival and death. The evolutionary choice of Ca^{2+} as a versatile messenger is not by a coincidence. For the first living cells on earth in theory four different cations (Na^{+}, K^{+}, Ca^{2+} and Mg^{2+}) would have been available when the need of internal cellular regulation came (Carafoli, 2002). Moreover, this regulation and signalling would have become even more accentuated when the multicellular organisms where developed. Taking into the account that cellular signalling should be a reversible process, the messenger ion should interact with another component of the cell in specific, rapid and reversible manner. Obviously, the acceptors for this messenger resulted to be proteins as they have a structural complexity to form interaction sites for the messenger. But, what makes Ca^{2+} different from the rest of cations? The unique combination of charge and size of Ca^{2+} permits the compliant reversible complexation with the proteins' interacting cavities where different amino acids, side chains and angles are allowed for Ca^{2+} binding.

Another important fact about Ca^{2+} is that the cell is not able to create or destroy it, it can only control the cytoplasmic concentrations using as Ca^{2+} sources the extracellular media and intracellular organelles such as endoplasmic reticulum (ER), sarcoplasmic reticulum (SR) or mitochondria. In fact, even though Ca^{2+} binding proteins could buffer Ca^{2+} levels inside the cell, small Ca^{2+} intracellular free concentrations achieved essentially thanks to ion transporters such as Na^{+}/Ca^{2+} exchangers that are constantly pumping cytoplasmic Ca^{2+} to the extracellular or intracellular organelles. Thus, fast and high-affinity regulation of Ca^{2+} levels is achieved. Additionally, neurons have developed Ca^{2+} signalling to an exquisite control as it orchestrates transmission of information through the nervous system by managing neurotransmitters release. Calcium is also involved in axon outgrowth, changes in synaptic strength.

In a resting neuron, extracellular Ca^{2+} concentration is around 1 mM while the cytosolic concentration is around 10 to 100 nM. In a response to a chemical or electrical signal intracellular Ca^{2+} levels can be rose to 1 to 2 μM rapidly. The mechanism by which this phenomena occurs is by voltage-dependent, receptor-dependent, calcium-dependent and TRP channels that open their pores allowing Ca^{2+} influx in favour of the electrochemical gradient (Figure 1.14). Besides, ER and SR also liberate calcium when they are stimulated by other type of second messengers such us IP_3 or ryanodine as we have already described in PIP_2 regulating section.

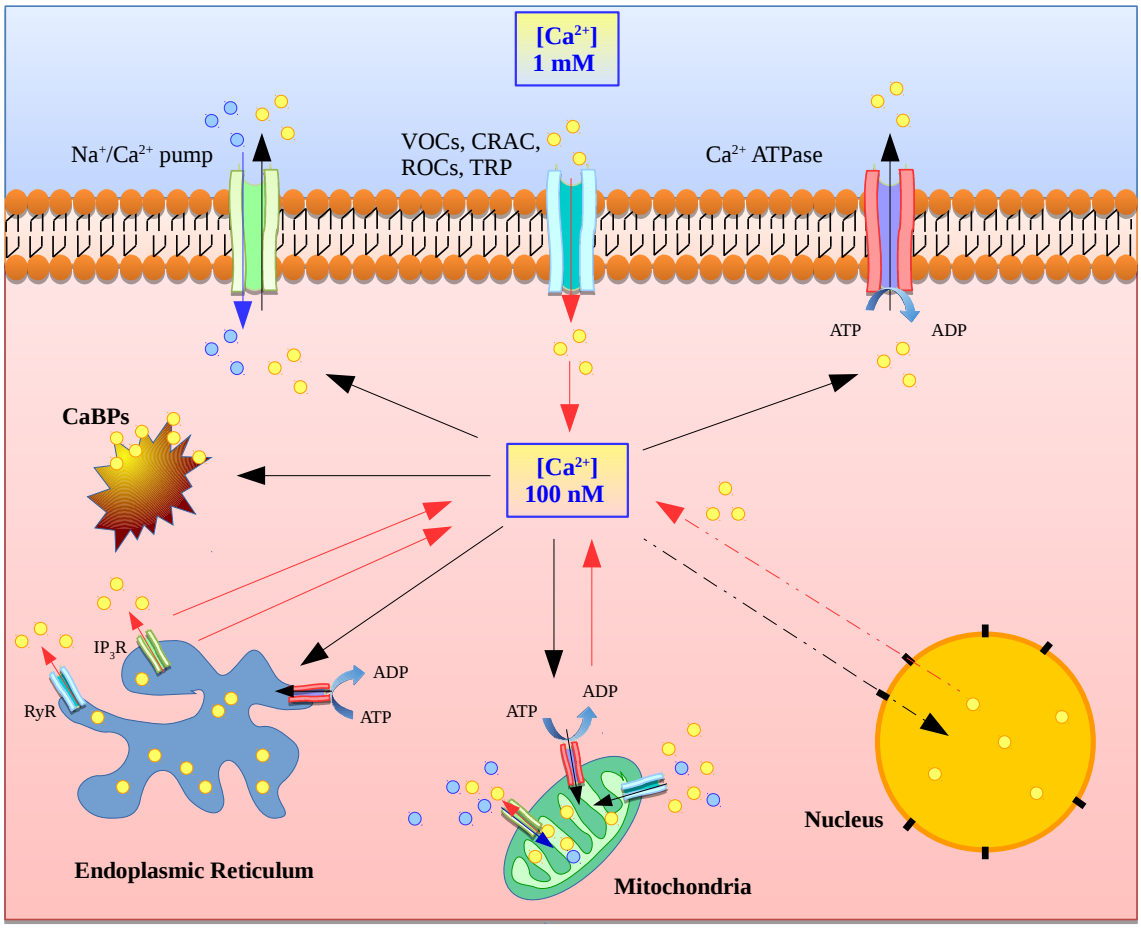


Figure 1.14. Calcium homeostasis in cells. After a stimuli, Ca^{2+} concentration rapidly rises when Ca^{2+} channels such as Voltage-Operated Channels (VOCs), Receptor-Operated Channels (ROCs), Calcium Release-Activated Channels (CRACs) or Transient Receptor Potential Channels (TRP) open. Moreover, more Ca^{2+} could be released to the cytoplasm through the activation of inositol 1,4,5-triphosphate receptors (IP_3P) or Ryanodine receptors (RyR). At resting conditions, intracellular Ca^{2+} levels are maintained at submicromolar levels thanks to Ca^{2+}-ATPases and to Na^+/Ca^{2+} exchangers located in plasma membrane, endo/sarcoplasmatic reticulum and mitochondria. Nuclear pores are opened every time and can contribute to buffering Ca^{2+} levels.

Nevertheless, how is possible that the Ca^{2+} signalling can generate a such variety of responses when the triggering factor is just a intracellular Ca^{2+} influx? The physiological outcome depends on the location, duration, amplitude and Ca^{2+} binding protein. In specialized cells, Ca^{2+} entrances are polarized which means that the Ca^{2+} transporters are localized in certain microdomains. The duration and signal amplitude activates or inactivates different signalling pathways. On the other hand, specific calcium binding proteins are likely to be activated by specific Ca^{2+} signals. Thereby, by the combination of all of them give rise to the communication complexity in which Ca^{2+} triggered signalling and transduction are correlated.

1.5.1 EF-HAND PROTEIN FAMILY

In general, those proteins able to reversibly and specifically bind Ca^{2+} and contributes to Ca^{2+} signalling are clustered in this protein family. All those calcium binding proteins (CaBPs) are located in the cytoplasm or inside the endoplasmic reticulum. In some cases Ca^{2+} is able to directly regulate the function of a CaBPs such as PKCs, PI_3K, PLCs, calpain or calcineurin. Nonetheless, CaBPs have often an intermediary role in which they are able to sense Ca^{2+} levels and release the signal to their targets.

The mechanism by which CaBPs senses Ca^{2+} is by a highly extended helix-turn-helix structural element called EF hand (Grabarek, 2006; Kawasaki & Kretsinger, 1994). A single EF-hand is composed of a N-terminal helix (helix E) followed by a small coil and a C-terminal helix (helix F). The structure has a peculiar similarity to a hand in which the pointer finger would be helix F, the thumb helix E and the remaining fingers would be the loop Ca^{2+} binding (Figure 1.17.A).

Even if all those proteins are able to bind calcium, the mechanisms by which they react may be different. A group of CaBPs (parvalbumin, careticulin, calbindin,...) does not suffer a conformational change after Ca^{2+} binding and they are more related to a buffering function rather than signalling function. Second group would be the real calcium sensors which decodes calcium signalling two structural changes upon Ca^{2+} binding. First, Ca^{2+} binds to CaBP opening its structural conformation exposing an interaction site for a target protein, and second, CaBP binds target protein changing its structure again. However, as we will see not all the CaBP proteins need Ca^{2+} to interact which their targets.

Taking into the account all the calcium fluctuations that a single neuron may suffer, EF-Hand protein family has evolved so as to cover a big range of calcium levels detection, which encloses different affinities for Ca^{2+} ($K_d = 10^{-9} – 10^{-5}$). Moreover, the number of EF-Hands in CaBPs may be different (four in CaM, five in calpains and six in calbindin) and the length of the helices E and F may vary but the turn in which Ca^{2+} is binding is pretty similar.

Besides CaM, among all those CaBPs two main groups are important for neuronal function.

- Synaptotagmins: are transmembrane protein associated with synaptic and secretory vesicles. They are composed of two C2 domains which are able to bind up to 3 Ca^{2+} cations with a relatively low affinity ($K_d > 10$ μM). This binding process requires coordination by the protein and the membrane lipids.

- Neuronal Calcium Sensors (NCS): Those are related to CaM. They have four EF-hands but not all of them are functional. Unlike CaM, they have a compact conformation when they are binding Ca^{2+} and open conformation when they are not. They also have some motifs that allow membrane association which is regulated by Ca^{2+} binding.

1.5.2 CALMODULIN

Without a doubt, CaM is the most studied member among all the EF-hand protein family. Calmodulin is a very soluble and thermostable small acidic protein (pI ~ 4) of 148 residues (first methionine is not taken into account) which is ubiquitously expressed in all the cells. It has been highly conserved during the evolution as it can be found in all eukaryotes and for vertebrates de 100 % of the amino acid sequence is identical. Cells often have a intracellular CaM concentration around 10 µM but if we have a look on neurons this concentration rises up to 100 µM. However, in most cells the number of targets exceeds that of CaM, and the majority is already engaged in a complex (Villarroel et al., 2014). Although the amount of free holo-CaM (Ca^{2+} loaded) is likely limiting (~ 50 nM), the concentration of apo-CaM (Ca^{2+} free) available is highly debated, with estimates ranging from 60 to 8800 nM. Recent functional studies suggest that free apo-CaM is accessible at sub-micromolar concentrations.

Calmodulin is involved in a big number of cellular processes in which it senses the calcium levels within the cell and relays the calcium signals on its targets. Among all the processes it has been shown to be implicated in inflammatory processes, immune response, smooth muscle contraction, cell division and proliferation, gene expression, hormones secretion, neurotransmitters secretion, short-term and long-term memory and apoptosis. Moreover, CaM directly interacts and regulates (activating or inhibiting) too many proteins such as enzymes, cytoskeleton, transporters, receptors and ion channels (Figure 1.15)(Table 1.3).

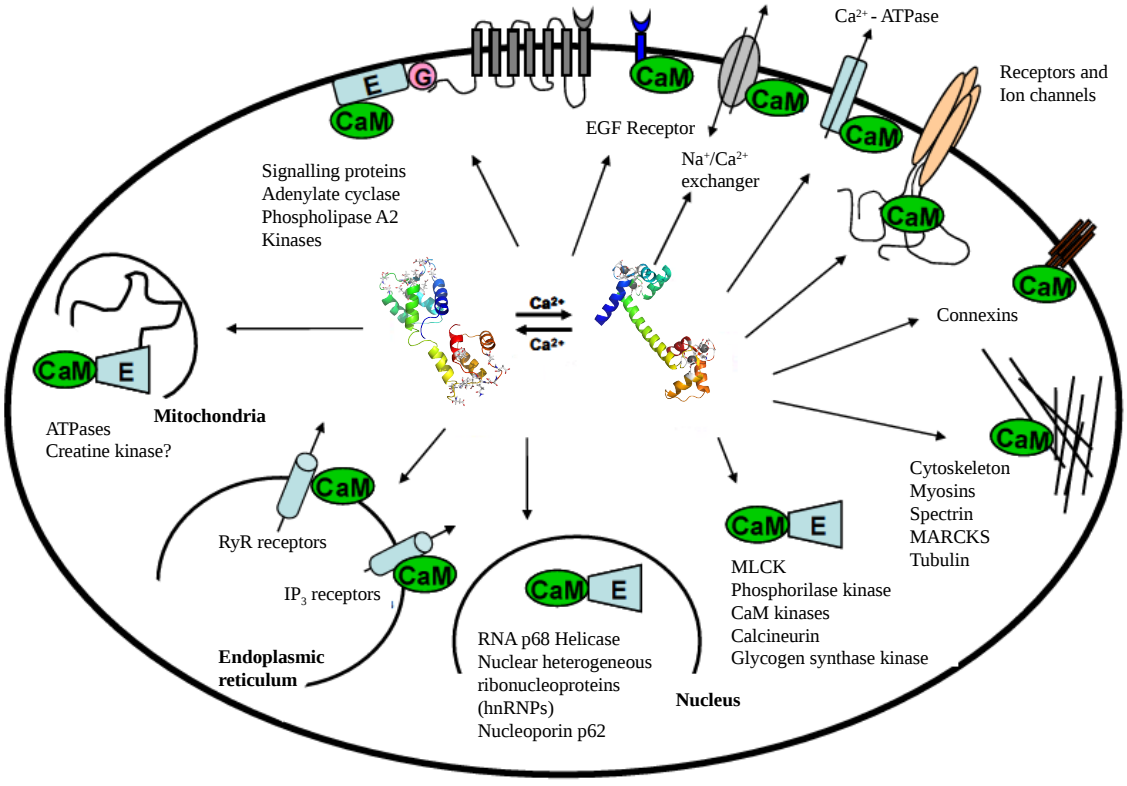


Figure 1.15. Net of proteins regulated by CaM.

Type	Proteins
Kinases	Myosin light-chain kinases (MLCKs)
	Ca^{2+}-CaM dependent protein kinases (CaMKs)
	Phophorylase kinases
	Phosphofructokinases
	G-protein coupled receptor kinases s(GRKs)
	NAD+ kinases
	Glycogen synthase kinases
Other enzymes	Calcineurins
	Adenylate cyclases
	Glutamate decarboxylases
	Nitric oxide synthases (NOS)
	Phosphodiesterase
Transporters	Ca^{2+}-ATPases
	Na^{+}/Ca^{2+} exchangers
Reticular receptors	Inositol 1,4,5-triphosphate receptors (IP_3P)
	Ryanodine receptors (RyR)
Ion channels and receptors	Small conductance calcium-activated potassium channels (SK, IK)
	Cyclic nucleotide gated channels (CNGs)
	Transient receptor potential channels (TRPs)
	NMDA receptors
	Ca_v channels (L, P/Q, R)
	Na_v channels
	K_v channels
	Epidermal growth factor receptor (EGFR)
	Connexins (GAP junctions)
CaM reservoir proteins	Neuromodulin
	Neurogranin
	Regulator of calmodulin signalling (RCS)
Genes expression systems	RNA helicase p68
	Heterogeneous nuclear ribonucleoproteins (hnRNPs)
	Nucleoporin p62
Cytoskeleton and other proteins	Caldesmon, Spectrin, Syntrophin, Dystrophin, PEP-19
	MAP2, Adducin, Marck, Tubulin, Myosin, Actinin, etc

Table 1.3. Examples of proteins regulated by CaM.

As previously shown, CaM is able to interact which a large quantity of different proteins but, how is possible for a single protein to recognize and to interact with all those different types of proteins? As it will be outpointed later, the big secret remains in the unique versatility of its structure.

1.5.2.1 Architecture

Calmodulin is formed by two mostly globular domains that are presumably originated through a sequential double-gene duplication event. Each domain, often referred to as lobes, contains two Ca^{2+} binding EF-hand motifs and are connected by a very flexible linker which confers the structural plasticity either at individual side chain or at domains orientation level. This ability, as we will see, is crucial for target proteins recognition and regulation. Moreover, this linker gives almost full autonomy for both lobes to fold, to bind Ca^{2+} or to interact with target proteins.

In the absence of Ca^{2+}, apo-CaM shows a compact or closed conformation where de hydrophobic residues of the helices are in the inner part of the structure (Figure 1.16.A). If we have a closer look we will find the structures between both lobes are slightly different: The N-lobe has a very closest conformation whereas the C-lobe has a semi-open conformation in which some hydrophobic residues are exposed to the solvent. This could enable the protein target binding of CaM in the absence of calcium as it has been observed in SK2, neurogranin, Ca_v and Na_v channels.

In the presence of Ca^{2+}, holo-CaM suffers a big structural rearrangement changing its conformation from closed to opened. As a consequence, the hydrophobic residues that were hidden before now they are totally exposed to the solvent. Moreover, previously disordered linker now forms a big α-helix providing continuity to EF2 and EF3 (Figure 1.16.B).

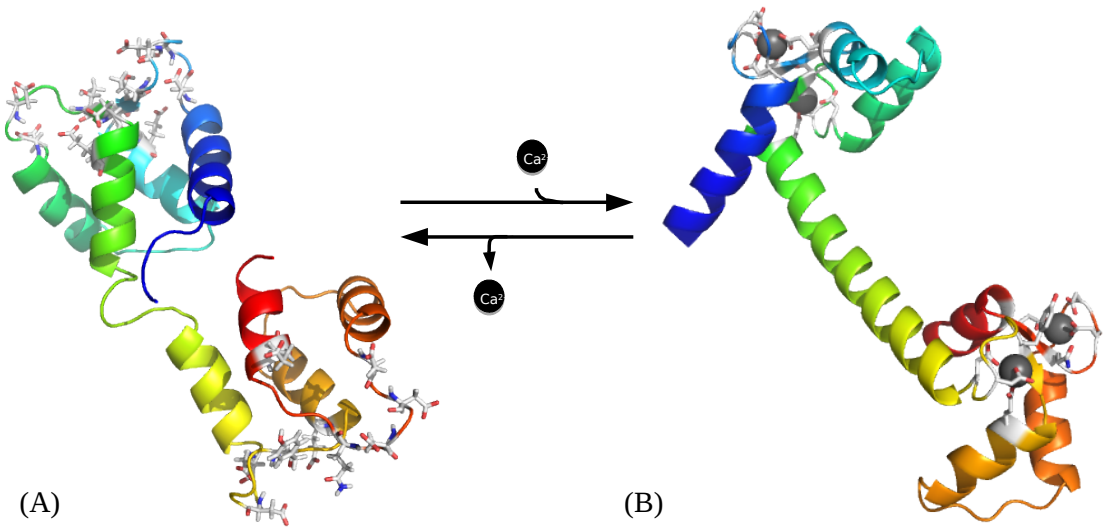


Figure 1.16. CaM structure. (A) Apo-CaM. (B) Holo-CaM. The N-lobe and the C-lobe of CaM are connected by a flexible linker which is not structured in the absence of Ca2+ and is structured in the presence of it. In grey side chain atoms have been highlighted from the residues forming the EF-hands.

1.5.2.2 Calcium binding and affinity

As mentioned before CaM has four fully operative EF-hands which are composed by the amino acid segments 20-31 (EF-1), 56-67(EF-2) in the N-lobe and 93-104 (EF-3) and 129-140(EF-4) in the C-lobe (Figure 1.17). This canonical sequence is composed of 12 amino acids which starts at a Asp and ends in at Glu. The oxygen atoms of half of the canonical amino acids take part in Ca^{2+} binding (residues 1-3-5-7-9 and 12). Even if in theory all EF-hands should have same affinity for Ca^{2+} different affinities have been reported for both lobes. In fact, due to a minor amino acidic differences between CaM's lobules, C-lobe has a higher affinity for Ca^{2+} than the N-lobe (K_d ~ 0.2 µM and ~ 2 µM respectively). This phenomena empathises the idea that both lobules are able to work either independently or cooperatively.

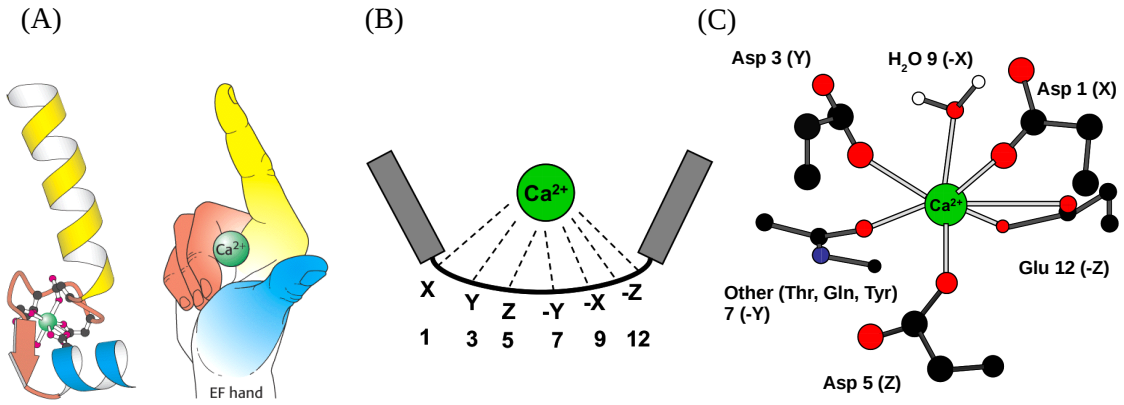


Figure 1.17. EF-hands structure. (A) Representation of the origin of the terminology "EF-Hand". (B)(C) Overall Ca2+ binding motif corresponding to a EF-Hand. The characters (X, Y, Z,...) and numbers correspond to relative positions in the EF loop. (Pictures provided by Dr. Alaimo).

The ample array of ionic conditions and CaM concentration employed results in different estimates for Ca^{2+} binding affinity, with reported K_d values ranging between 300 and 5000 nM. Under near-physiological intracellular ionic conditions, the K_d is about 1 µM, a value that allow CaM to respond to intracellular Ca^{2+} oscillations. Nonetheless, this affinity may vary when CaM is complexed with its target.

In general, CaM affinity for Ca^{2+} is higher when complexed with a target although the opposite effect has also been described. Some values for the Ca^{2+} affinity of target-bound CaM are lower than 50 nM. This would mean that in resting cell conditions in which Ca^{2+} concentration is around 100 nM the CaM/target complex may be Ca^{2+} saturated. Reciprocally, the affinity of many targets for CaM increases when it is loaded with Ca^{2+}.

1.5.2.3 Calmodulin target recognition

CaM is able to recognize and regulate a huge variety of proteins which are implicated in many different cellular processes. Generally, CaM targets formed by 16-30 amino acids and are enriched in basic and hydrophobic residues (Figure 1.18). Most of them have an amphipathic helical structures but others are considered to be disordered in solution. However, last groups of targets often present a tendency to adopt an alpha-helical configuration upon interaction with CaM. On the other hand CaM recognized its targets by some residues or “hot spots” which often creates hydrophobic pockets in which Van der Waals and other kind of interactions occur (Figure 1.19).

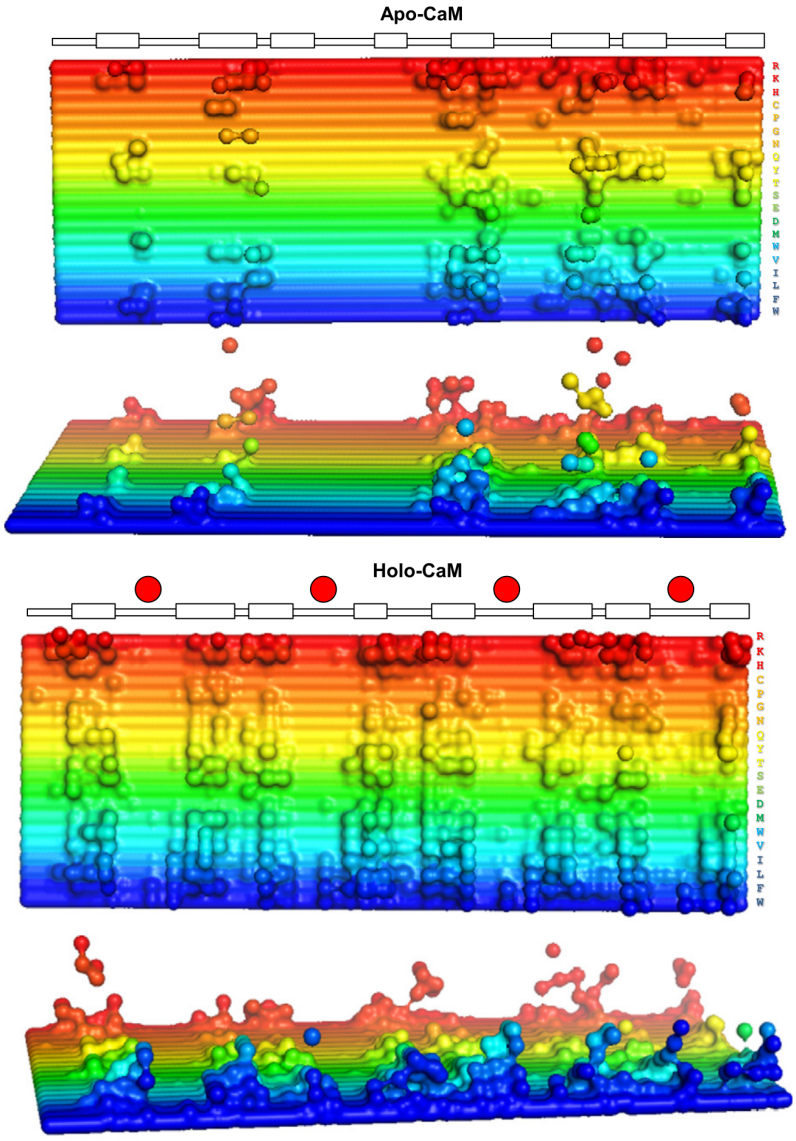


Figure 1.18. Interaction preference of each CaM residue with amino acids from target peptides. Aerial and side view of contact surface area (z-axis) 3D plots of different target residues (y axis) in contact with each of the vertebrate apo-CaM or holo-CaM residues (x-axis). The Ca^{2+} binding loopos are indicated by the red spheres, and the CaM α-helices are boxed on top of the plot. Taken from (Villarroel et al., 2014).

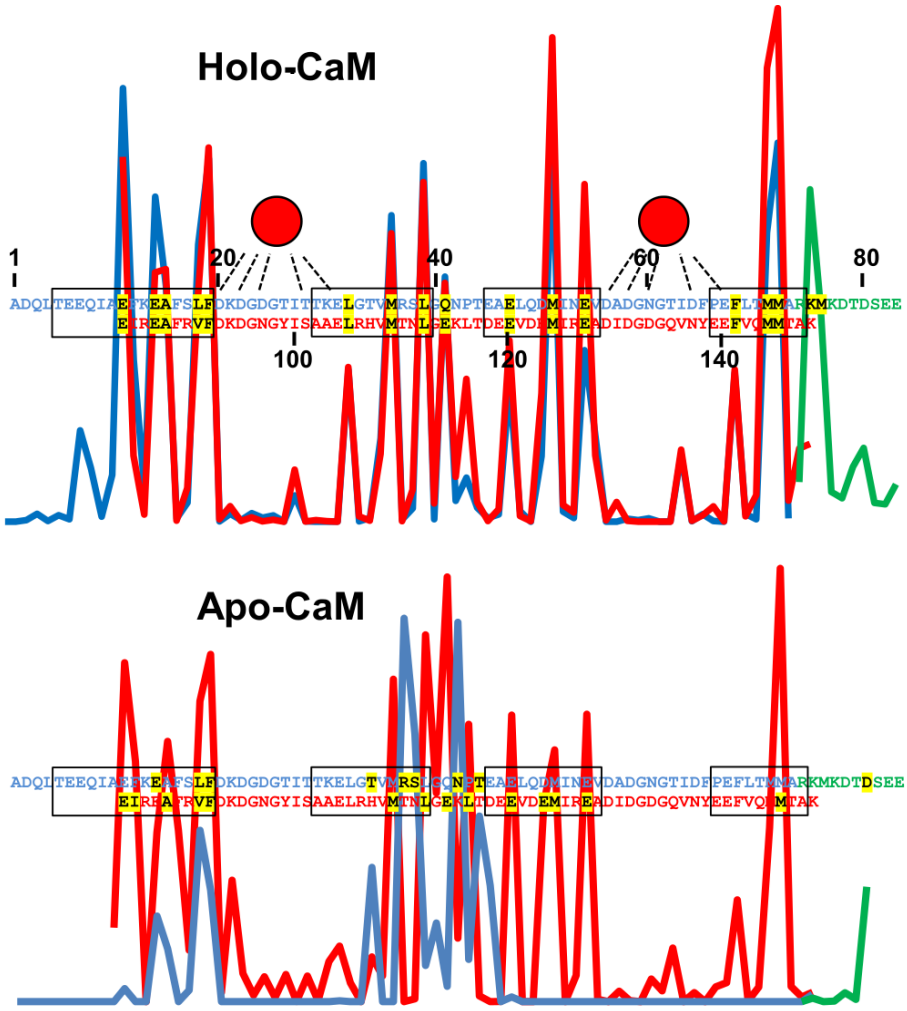


Figure 1.19. CaM's averaged contact surface with targets. It defines eight "hot spots" that correspond to the eight alpha helices of CaM represented as boxes. Plot of the weighted averaged CSU of the residues of holo-CaM and apo-CaM with targets. Red: C-lobe residues, blue: N-lobe residues. Ca^{2+} is represented by the red spheres in (a), indicating the binding residues by broken lines. The main residues interacting with the target are shaded grey. The boxes indicate the CaM alpha helices. Two residues of the flexible inter-lobe linker (Lys75 and Met76) providing significant contributions to target binding are shown by the green line. The two inter-hand linkers also provide significant contributions through Gln41 (N-lobe) and Glu114 (C-lobe).

Due to the high amount of binding motifs for CaM and their amino acid sequence diversity it has not been possible to classify them in their totality. Nevertheless, it is clear that those motifs could be divided in two mayor groups: Those interacting in the presence and those interacting in the absence of Ca^{2+}:

1.5.2.3.1 Ca^{2+} dependent binding

In this regard CaM's targets are often small helices of around 20 residues in which basic and hydrophobic amino acids are predominant. Often, two hydrophobic residues called "anchors" are essential CaM binding and recognition as those two anchoring residues interact with the hydrophobic pockets of one of the two lobes of CaM. The way in which those motifs have been classify is by numbering the amino acids between those hydrophobic residues. Some examples are illustrated in Table 1.4.

Even though there are some exceptions, for the majority of cases Ca^{2+} signalling is divided in two conformational changes (Figure 1.20). First, apo-CaM binds calcium in the absence of the target, which generates a conformational change from the apo-CaM to the holo-CaM. Second, when the hydrophobic patches of CaM are exposed it is now able to recognize target's binding motives. Afterwards, CaM makes contact with the target and wraps it. Moreover, depending on the target type this wrapping effect may result in the target folding to an α-helix.

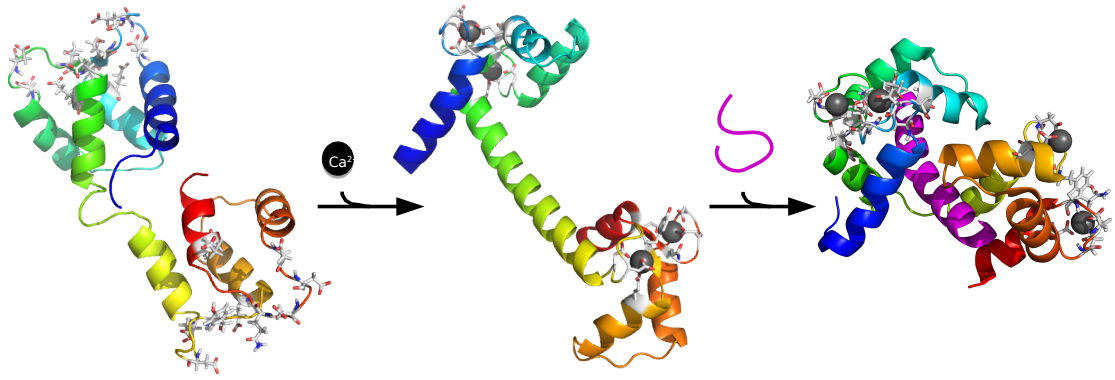


Figure 1.20. CaM complexing with its target. Apo-CaM (PDB: 1CFC) binds Ca^{2+} ions which provokes a conformational change (PDB: 3CLN). Hydrophobic residues are exposed and target recognition occurs (PDB: 2JZI).

1.5.2.3.2 Ca^{2+} independent binding

Many of the targets that bind to CaM in the absence of Ca^{2+} have the well-established IQ motif. On the contrary to the previous group, IQ motif is more defined with a binding sequence IQxxxBGxxxBxxX where B means basic amino acid (Lys or Arg) and X is a hydrophobic residue (Phe, Ile, Leu, Val, Trp or Tyr)(Table 1.4). Target proteins with this motif usually interact with the apo form of CaM, even if there are some exceptions.

There is much less structural information about complexes with apo-CaM. Therefore, the mechanisms by which Ca^{2+} signalling happens in less intuitive as the previous one which give rise to several hypothesis of how this process occur. As previously mentioned the C-lobe of apo-CaM has a semi-open conformation which could allow the interaction with its target (Figure 1.21). Once the interaction occurs the affinity of the C-lobe for CaM may increase but this binding of Ca^{2+} could only stabilize the C-lobe binding to the target. The N-lobe, however, is still free to bind Ca^{2+}. In this case is more probable to suffer a conformational change upon Ca^{2+} binding with may expose hydrophobic residues to attract other targets.

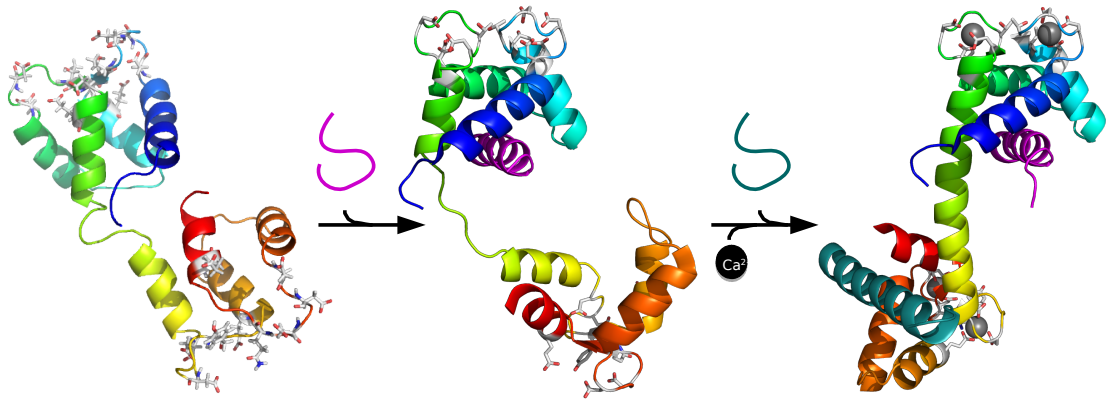


Figure 1.21. CaM complexing with its target. Apo-CaM (PDB: 1CFC) binds a target protein (SK2) which provokes a minor conformational change (PDB: 1G4Y). However, after Ca^{2+} binding a drastic conformational change appends, by which second target is able to binds (PDB: 3SJQ). In this last structure a dimer is formed which has not been represented.

Protein	Type	Sequence
	1-14	XX**X**XXXXXXXXXXXX**X**XX
CaMK Iα		SK**W**KQAFNATAVVRH**M**RK
	1-5-8-14	XX**X**XXX**X**XX**X**XXXXX**X**XX
MLCK (muscle)		RR**W**KKN**F**IA**V**SAANR**F**KK
	1-8-14	XX**X**XXXXXX**X**XXXXX**X**XX
NOS (neuronal)		IG**F**KKLAEA**V**KFSAK**L**MG
MLCK (smooth muscle)		RK**W**QKTGNA**V**RAIGR**L**SS
	1-10	XX**X**XXXXXXXX**X**XXXXXX
IP_3 kinase		KS**W**RKIKNMVH**W**SP
	1-5-10	XX**X**XXXX**X**XXXX**X**XXXXX
CaMK IIα		RRK**L**KGA**I**LTTM**L**ATR
	1-16	XX**X**XXXXXXXXXXXXX**X**XX
CaMKKα		PS**W**TTVILVKSMLRKRS**F**GN
	2x 1-5-10	XX**X**XXX**X**XXXX**XX**XXX**X**XXXX**X**XX
$K_v7.2$ (Helix B)		PG**L**KVS**I**RAVC**VM**RFL**V**SKRK**F**KE
	1-12	XX**X**XXXXXXXXXX**X**XX
NMDA receptor (NR1aC1)		AT**F**RAITSTLASS**F**KR
MARCKS	Basics	KKKKKRFSFKKSFKLSGFSFKKNKK
EGFR		RRRHIVRKRTLRRLLQ
NMDA receptor (NR1aC0)	Others	RHKDARRKQMQLAFAAVNVWRKNLQ
	IQ	XX**IQ**XXX**B**XXXX**B**XX**X**XXX
Neurogranin		AK**IQ**ASF**R**GHMA**R**KK**I**KSG
Neuromodulin		TK**IQ**ASF**R**GHIT**R**KK**L**KGE
$Ca_V1.2$		FL**IQ**EYF**R**KFKK**R**KEQGLV
$K_v7.2$ (Helix A)		GL**IQ**SAW**R**FYATNLSRTDL
$Na_v1.2$ (Helix A)		II**IQ**RAY**R**RYLL**K**QK**V**KKV

Table 1.4. Overall motif for CaM recognition. Adapted from (Ishida & Vogel, 2006).

1.5.2.4 CaM and Disease

In humans as in other vertebrates three CaM genes are available in the genome (CALM1-3) which are translated into an identical CaM protein. Considering that during the evolution the amino acid sequence has remained identical for vertebrates, we could imagine that any mutation in CaM would derive in a cellular malfunction and more likely to death. Nevertheless, it is noteworthy that in the last years up to ten CaM mutations have been related to arrhythmogenic human diseases (Villarroel et al., 2014)(Figure 1.22). In the heart, two missense mutations (N53I and N97S) were identified in CALM1 gene in patient with catecholaminergic polymorphic ventricular tachycardia (CPVT). Moreover, A102V mutation in CALM1 too has been also related to CPVT.

Following such discovery, exome sequencing on two unrelated infants with Long QT synthrome revealed the occurrence of two *de novo* missense mutations in CALM1 and in CALM2 (D129G and D95V, respectively) . In this study, the same CALM1 D129G mutation was also found in a 3-year-old boy who had suffered multiple cardiac arrests beginning at 1 month of age. Another CALM1 missense mutation (F141L) was revealed in a 14-year-old boy with recurrent episodes of CPVT. Same study identified CALM2 mutations which caused LQTs and CPVTs synthoms (D131E, D133H and Q135P) (Marshall et al., 2015). Besides, another mutation (F89L) was associated with the idiopathic ventricular fibrillation.

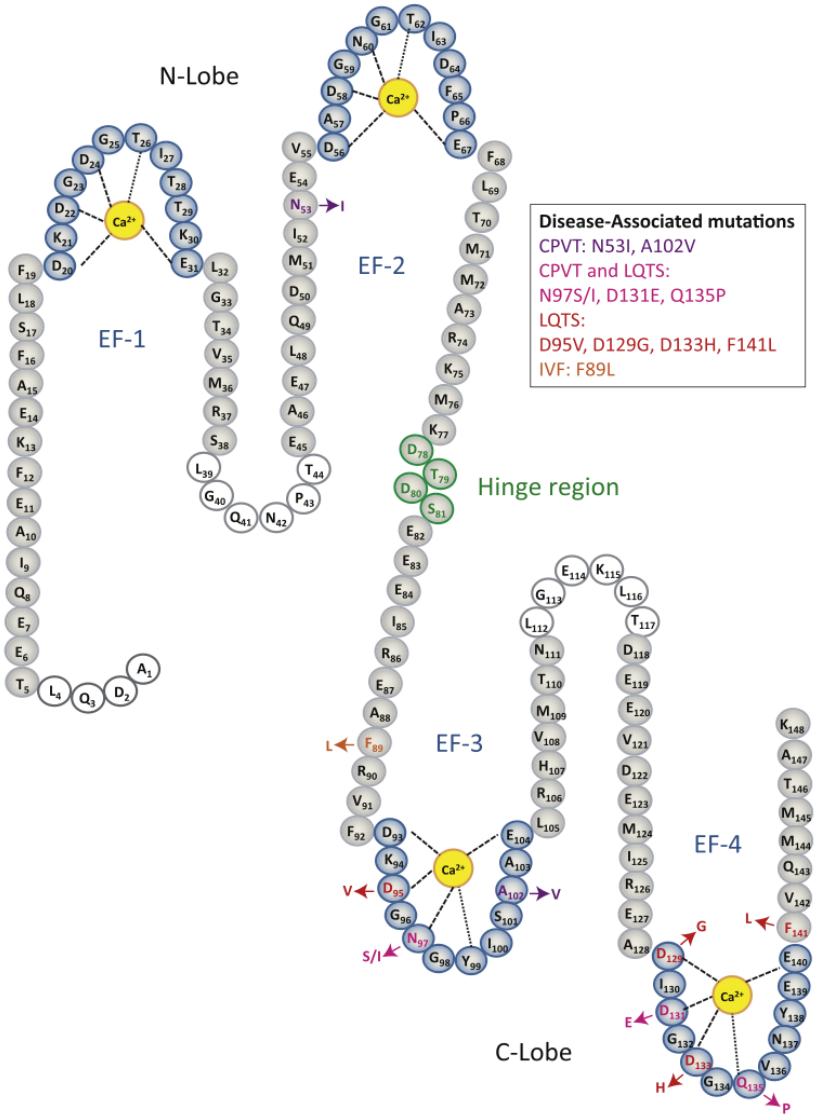


Figure 1.22. CaM sequence and disease-associated mutations. The hinge region between N-lobe and C-lobe is colored in green. Four EF-Hands are marked in blue where those residues in direct contact with Ca^{2+} have a dashed line towards Ca^{2+} ion. Most of the disease-associated mutations (pointed with arrows) are located in the EF-3 and EF-4 hands, in the C-lobe. Image taken from (Marshall et al., 2015).

OBJECTIVES

The aims of the present doctoral thesis are:

1- To improve the recombinant protein expression of intracellular C-terminus of K_v7 channels in *Escherichia coli*.

2- To structurally characterize the protein complex formed by CaM and the CaM Binding Domains of $K_v7.2$ channels by NMR.

3- To determine the calcium effect on the protein complex.

2 IMPROVING RECOMBINANT PROTEIN EXPRESSION OF K_V7 INTRACELLULAR SEGMENT IN *ESCHERICHIA COLI*

2.1 INTRODUCTION

2.1.1 LACK OF INFORMATION ABOUT K_v7 CHANNELS: THE CHALLENGE OF WORKING WITH MEMBRANE PROTEINS

Outcomes of various genome projects have revealed that up to 30 % of proteins encoded by eukaryotic cells are membrane proteins. However, despite the fact that more than 112.000 structures have been solved according to the RSCB Protein Data Bank (www.pdb.org) in October 2015, only around 3000 structures (less than 3 %) correspond to membrane proteins. Moreover, around 700 structures correspond to channels and from these 113 correspond to K_v channels. But the low numbers are not precisely for lack of interest. In fact, the 60 % of pharmaceutical drugs' effects relays on these proteins (Fagerberg et al., 2010). Therefore, a better understanding of the relationship between membrane proteins' structure and function would entail unimaginable contributions to structural biology, pharmacology and medicine.

Working with K_v7 channels is difficult too as only 7 structures has been released. Therefore, whereas the understanding of their regulation and their electrophysiological properties has increased dramatically in the last years, still little is known about K_v7 potassium channels structure, mainly because obtaining the large amounts of purified protein required for crystallography or NMR has proven to be very challenging.

The largest limiting factor in most functional and structural characterization is the production and purification of high amounts of soluble, functional and highly pure protein. Unfortunately natural protein sources do not usually satisfy these conditions. In these cases researchers often think about recombinant technology to overcome this problem. Even if there are several protein expression systems described in the literature (Rai & Padh, 2001), *Escherichia coli* is still the preferred choice among the scientific community for several reasons: it has short-life cycle (30 minutes), low complexity medium, high expression levels, easy isotope labelling protocols for NMR and selenomethionine incorporation for X-ray, well-known genetics, easy genetic manipulation, and the availability of a wide variety of commercial tools and protocols. However, there are many advantages using *E. coli*: it is not able to produce some eukaryotic post-translational modification such as glycosylation or phosphorylation, it could be a problem when expressing eukaryotic proteins due to codon-bias, membrane and/or big protein over-expression may result in formation of insoluble and misfolded aggregated proteins known as inclusion bodies and finally even if we obtain soluble protein it may not be functional due to an impropriety folding.

Since there is no universal solution to face these problems here we describe a scheme that we have followed in order to successfully over-express the C-terminus of K_v7 channels in *E. coli*. This workflow could be employed in any other protein and could help any researcher who has to start a new project in which a protein purification is needed or any other researcher who is stacked in this task (Figure 2.1).

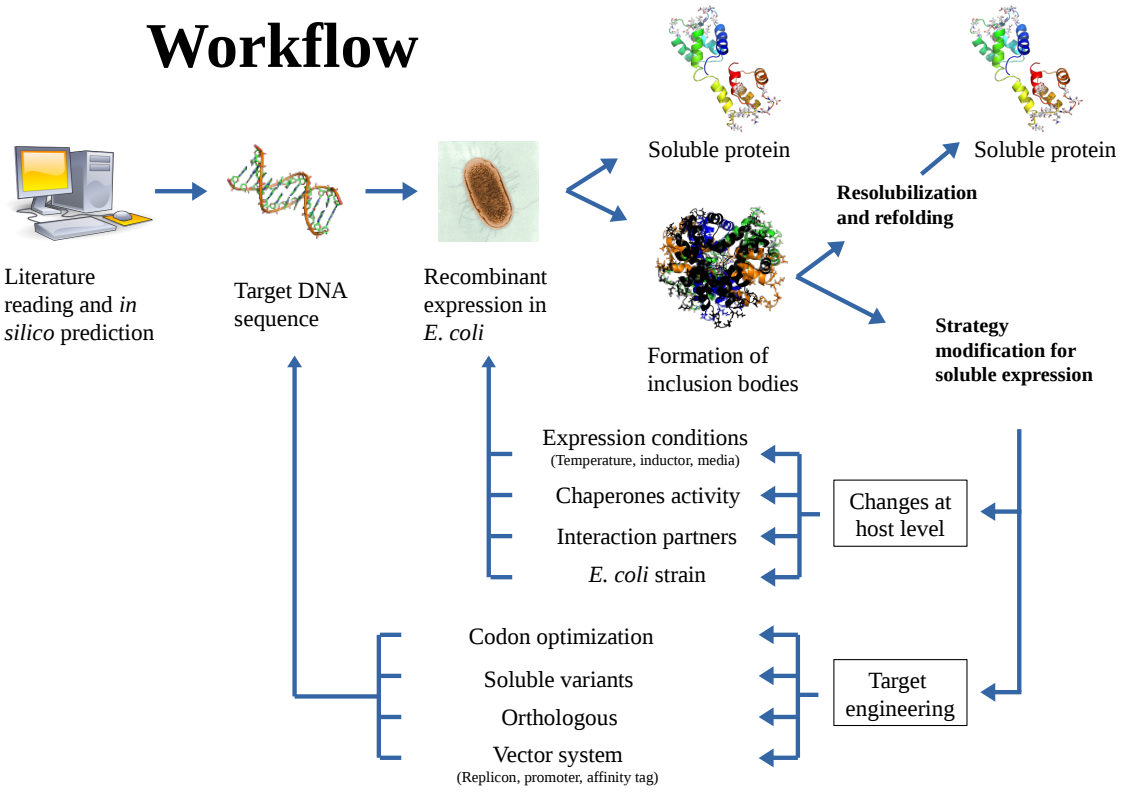


Figure 2.1. Workflow proposed to improve recombinant protein's solubility. After acquiring more knowledge about our target protein first attempts of expression should be done. In case of inclusion bodies formation we need to reconsider to make changes either at host level or at protein DNA sequence level. There is not an specific path to follow and parallelization is recommendable so as to save time and effort. Every single point will be described later.

2.1.2 LITERATURE READING AND *IN SILICO* PREDICTION BASED ON SEQUENCE

Usually when we are in front of a new project and we do not really know how to start, the easy way to begin with is to read and learn more about our protein of interest. At the beginning checking the literature may result in a time inversion but soon rewards the scientist. Sometimes we think about a new and original idea and we find out that someone thought about it before loosing all the appealing to continue with it. In other occasions, we can find that some others have characterize our protein of interest which can save time and effort. And finally we can check whether other scientists have face the same problems as you are dealing with and how the have solved them or how they have not.

Even if we are in the worst scenario where there is no information about a certain protein in the literature, there are some softwares that could help. Due to the increasing expansion of protein databases, several programs try to predict proteins' features such as solubility, stability, secondary structure, intrinsically disordered regions, domains and even function from the amino-acid sequence.

In order to predict protein solubility and stability programs such as ESPRESSO (Hirose & Noguchi, 2013), PROSO (Smialowski et al., 2012), ccSOL (Price et al., 2009) and CamSol (Sormanni et al., 2015) are highly recommendable. They often use several features obtained from the primary amino-acid sequence such as molecular weight, net charge, isoelectric point, signal peptides, GRAVY values, single residue decomposition, dipeptide decomposition and many others as described (Habibi et al., 2014). Secondary structure can be predicted using protein structural databases as reference to detect regions or domains that are most likely to form stable secondary structural elements. Programs using this principle are for example PsiPred (Jones, 1999), GOR5 (Sen et al., 2005), and PREDATOR (Frishman & Argos, 1997). Intrinsically disordered region are characterized of been charged and hydrophobic (Dunker et al., 2001). PONDR (Romero et al., 1997) defines these regions as these that do not appear in PDB structures whereas DisEMBL (Linding et al., 2003) and IUPred (Dosztányi et al., 2005) look for loops with high B-factors. FFAS (Xu et al., 2014) provides functional and structural information based on profile-profile sequence alignments and fold recognition to identify remote structural homologies. Finally there are other platforms that provides different *in silico* services which try to gather all of them such as SCRATCH predictors (Cheng et al., 2005)

2.1.3 STRATEGIES FOR OVER-EXPRESSION AT HOST LEVEL

2.1.3.1 Changing expression conditions

In normal conditions (37 ºC) *E. coli* transcription and translation are highly coupled, synthesizing as far as ~ 60.000 polypeptides chains per minute (Lorimer, 1996). At this speed the recently synthesized recombinant polypeptide is exposed to a new cellular environment which differs from its original source. Hydrophobic regions could be exposed in misfolded proteins which can interact among them causing a protein aggregation and inclusion bodies formation. Besides, when over-expressing a foreign gene in *E. coli*, all the bacterial machinery is recruited and causes a "metabolic burden" (Sørensen & Mortensen, 2005) by which bacteria loses its capacity to control and express house-keeping genes provoking which ultimately provokes a metabolic destabilization.

Growth **media** is essential for bacterial life. All nutrients, oxygen and pH should be supplied for correct protein production. Furthermore, if our target protein requires a specific cofactors should be added.

The reduction of cellular protein production or concentrations favours proper folding (Rosano & Ceccarelli, 2014). Either decreasing the **temperature**, decreasing **inductor concentrations** or both leads to higher amounts of soluble protein. The main explanation for this effect is that lowering down the inductor levels provides more time for recombinant protein to fold. Besides, lowering down the temperature, temperature-dependent hydrophobic interactions are reduced. However, expression times are longer and the produced biomass is lower.

2.1.3.2 Increase chaperones activity

Chaperones are specialised proteins that aid unfolded polypeptides to reach their final conformation. In bacteria GroEL (Hsp60 family), Dnak (Hsp70 family), HtpG (Hsp90 family) and ClpB (Hsp100 family) are the most common chaperones present in the cytoplasm. these chaperones operate according to their host requirements and they are more active when their host is in a stress condition. In fact, when the temperature of protein expression is changed, researcher has to be aware that bacteria have to adapt to the new temperature and produce these heat-shock proteins before we induce our protein.

As mentioned before, when inducing the production of the target protein bacterial machinery suffers. At this moment the chaperones availability is lower than the newly synthesized protein and that is why sometimes the system collapses leading to the formation of inclusion bodies. However, chaperones co-expression with target proteins has been proven to be effective (Nishihara et al., 2000).

There is another strategy to improve protein solubility from inclusion bodies without the necessity of over-expressing chaperones. Protein expression can stop removing all the media by centrifugation and replacing it with fresh media with chloramphenicol, which is an inhibitor of protein synthesis (Carrió & Villaverde, 2001). Thus, molecular chaperones can freely continue with the refolding of the proteins from inclusion bodies without the input of new unfolded material.

2.1.3.3 Interaction partners

When we are expressing individually a protein which requires other proteins or partners for stability and folding it may result either in not expression or formation of inclusion bodies. As described before cytoplasmic protein solubility depends on the distribution of hydrophobic residues on protein's surface. When, for example, two proteins adopt a complex conformation in nature usually they protect each other from exposing these hydrophobic patches due to the interaction between them. Precisely, these hydrophobic patches constitute the selective interaction, recognition or binding regions of one protein to the other. Hence, if we fail trying to obtain soluble protein and we know that our target protein requires its counterpart in order to be stable and perhaps soluble, could be advisable to identify and co-express both proteins in two compatible vectors or in any duet or bicistronic vectors (Gopal & Kumar, 2013).

Another possible scenario where interaction partners may favour protein solubility is when we express our protein as a fusion protein and we try to proteolytically removed the tag. Often we may have soluble protein attached to a fusion partner such as thioredoxin, maltose binding protein, etc. and when we cleave it our target protein precipitates. For example, it has been shown in pheromaxein A and C proteins that when they were produced separately with their fusion tags were soluble. However after cleaving their fusion proteins separately both proteins precipitated. However, when both protein were together and the cleavage was performed, both proteins, which are interacting partners, remained soluble (Austin, 2003).

2.1.3.4 *E. coli* strain

The choice of the host strains depends more on the nature of the heterologous protein (Table 2.1). Generally, BL21 and their derivatives strains are by far the preferred choice for expression. BL21 has Lon and OmpT proteases deleted which causes an inhibition in foreign protein degradation. DE3 version has a lysogen that encodes T7 RNA polymerase in order to express genes under the control of T7 promoters such as in pET systems. In order to control the leaky expression of the target gene pLysS strains include a plasmid expressing T7 lysozyme which is a natural inhibitor of T7 polymerase.

In case of willing to over-express a protein which forms a disulfide bonds **Origami** strains have a more oxidizing cytoplasmic environment as they are double mutant in thioredoxin reductase

(trxB) and glutathione reductase (gor). If dealing with a membrane-bound proteins **C41** and **C43** strains are a beneficial choice. They constitutively express F_1F_0 ATP synthase subunit b which is accompanied with a big membrane proliferation that could aid the expression of membrane proteins. Moreover, taken into account that the protein folding is much better at low temperature there is a strain designed for this purpose. **ArticExpress** (Agilent) has Cpn10 and Cpn60 chaperones from *Oleispira antarctica* which enable cell growth and protein expression at low temperatures helping protein folding.

Strain	Features	Usage	Consideration
BL21	Deficient in Lon and OmpT proteases	General expression	
BL21(DE3)	T7 RNA polymerase based expression	General expression	
C41(DE3) C43(DE3)	Big membrane proliferation	Membrane or Toxic proteins	
BL21(DE3)pLysS	Co-expression of T7 lysozime	Toxic proteins. Controls basal expression	Extra plasmid. Usually Cam resistance
CodonPlus(DE3) Rosetta(DE3)	Codon supplements	Overcome codon bias	Extra plasmid. Cam resistance for CodonPlus. Kan resistance for Rosetta
Origami(DE3)	Thioredoxin reductase (trxB) and glutathione reductase (gor) mutants	Disulfide bond formation	Extra plasmids. Kan resistance for trxB plasmid and Tet for gor plasmid
Rosetta-Gami(DE3)	Codon supplements and Thioredoxin reductase (trxB) and glutathione reductase (gor) mutants	Overcome codon bias and Disulfide bond formation	Extra plasmids. Kan, TeT, Str, and Cam
ArcticExpress	Cold adapted chaperones Cpn10 and Cpn60 from *O. antarctica*	Misfolding and Insolubility. Expression at low temperatures	Extra plasmid. Gen resistance

Table 2.1. Table of most used *E. coli* strands. Each strain shows its main characteristic, usage and considerations when using them. Cam: Chloramphenicol; Kan: Kanamycin; Tet: Tetracycline; Str: Streptomycin; Gen: Gentamicin

Many times protein expression is not successful due to the codon bias. Usually the target protein which has to be over-expressed has its origin in different organisms sources such as bacteria, eukarya or also archeabacteria. Despite the fact that the genetic code is considered to be universal, during the evolution these distantly related organisms have acquired different codon frequency preferences. During the translation of a foreign DNA in bacteria, the ribosome might truncate due to the poor availability of these rare codon in the host, leading to translational errors, frame shifting events or even premature termination.

Two alternative strategies are available so as to face codon bias: First approach could be to change the host to *codon-supplemented strains*. these bacteria are engineered to contain extra copies of genes that encode the tRNAs that often are not available for the cell. **CodonPlus** (Stratagene) and **Rosetta** (Novagene) commercial strains have an additional plasmid which contains these tRNAs. Second approach could be the *codon optimization of the target gene* where “rare” codons are replaced by these that are common to the tRNA frequency of the host as we will see later.

2.1.4 Strategies For Over-Expression Engineering Target DNA

2.1.4.1 Codon optimization

Another way of facing codon bias is optimizing our target cDNA. Nowadays we can modify the rare codons of our protein encoding gene so as to reflect or adapt to the codon usage of the host. This strategy is done either introducing silent point mutations by site-directed mutagenesis or *de novo* resynthesising the whole gene. Site-directed mutagenesis is convenient when few codons have to be changed. However, when many recombinant proteins are been studied it is more useful to synthesize the gene. There is a big list of available softwares for this purpose (Puigbò., 2007). Usually they do not only optimize the codon usage to the selected host but they also take into account the mRNA stability and secondary structure, the G/C content of it and codon harmonization. Another advantage of synthesizing a gene is that the researcher can introduce or remove restriction sites for specific restriction enzymes. This could be very interesting when we decide to perform a parallel cloning of a target DNA into different expression vectors.

2.1.4.2 Screening soluble variants

After the collection of enough information about the target or an homologous protein either by sequence-homology alignments, secondary structural predictors, intrinsically disordered regions predictors, structural information in PDB or by experimental or functional approaches, **rational site directed mutagenesis** can be employed to design different sub-constructs. The aim of this idea is to find an alternative version of the target protein which shows improved folding, solubility and stability profile. Mutation methods are based on these found in the natural evolution which consist on point mutations, deletions and truncations.

A priori it may be difficult to predict which one of these is more appropriate but there are some tips that could help (Hart & Tarendeau, 2006):

1. Point mutagenesis is preferable to be done in single-domain proteins.

2. Often when we are working with multidomain proteins is desirable to introduce truncations.

3. Unfolded or disordered regions that lead to expression problems could be overcame by truncation or deletion.

4. If we are willing to expresses separately a protein which forms a complex with other proteins may be favourable to truncate or remove this interacting region.

5. Removal of signal peptide coding sequences which don't have any other role in the new host cell may increase the expression and solubility.

When structural information is available in the Protein Data Bank it is much easier to improve the folding, stability and solubility of a target protein by rational site directed mutagenesis. In fact, apart from the intrinsic solubility prediction CamSol method identifies protein mutants with enhanced solubility *in silico* (Sormanni et al., 2015). This program identifies which amino acid substitution or insertion has the highest impact in protein solubility allowing a fast solubility screening of protein libraries.

A more unspecific method to generate soluble variant randomly is by **directed evolution**. This well established method generates protein libraries by random point mutations, deletions, truncations and gene fragment shuffling (Xiao et al, 2015). Indeed, this technique has been widely applied in the improvement of protein solubility as well as industrial enzymes where kinetics, thermostability and substrate specificity have been enhanced (Lutz, 2010). Surprisingly one noteworthy outcome of several directed evolution projects is that the result wouldn't be predicted from the beginning or that the results are different from these done by rational design (Tobin et al., 2000). Therefore this leaves room for the doubt of which strategy of both is more advantageous to find suitable soluble proteins.

Nevertheless, these soluble proteins obtained by any of these methods can not necessarily be biologically active or properly folded. It should be checked that the new soluble variant has the same activity and behaviour of the wild type protein. Otherwise in case the new protein differs from the first one, scientists should evaluate whether the result they are going to report is representative of the WT rather than of an artefact protein.

2.1.4.3 Screening orthologous

A developing strategy to improve protein expression is to analyse different sequence homologue of the same protein from various species. The base of this approach is that among these

orthologous we could find an amino-acid sequence variation that do not affect protein function but have important effects on the stability and solubility.

A functional protein is the result of evolution, and the complexity of it is often related to the organism complexity. Therefore there is a belief that orthologous proteins from more primitive species might display more soluble and stable properties (Brelidze et al., 2012). Furthermore, there is a tendency of studying protein orthologous from extremophile species such as thermophilic organisms. Their proteins usually have more salt bridges and less disordered regions that provide them the stability to face these severe environments. these adaptations could be useful so as to obtain soluble proteins for structural analysis making them better candidate for purification and crystallization (Savchenko et al., 2003).

2.1.4.4 Changing expression vectors

Plasmids are small circular double-stranded DNA which are not essential for bacteria but often provide their host of a specific feature by which they can adapt to the environment easily. these plasmids are capable of automatically replicate in a host cell and they are widely used to introduce and produce our target protein in *E. coli*. Nowadays all these expression vectors available are outcome of the combination of different features common in all engineered vectors such as a selectable marker, a origin of replication (replicon), a promoter, a multiple cloning site (MCS), a tag and fusion protein and a protease cleavage site.

In order to determine whether or not our bacteria have incorporated our target plasmid a selectable media is used depending on the plasmid's selectable marker. In *E. coli* antibiotic resistance genes are commonly used for this purpose. Growing medium is then supplemented with the appropriate amount of antibiotic to kill plasmid-free cells (Table 2.2). Using a bigger amount of antibiotic may cause a slow growing of the culture. Special care has to be taken when using ampicillin because protein which provides the resistance to the antibiotic, b-lactamase, is secreted by the bacteria to the media degrading all the antibiotic and in the absence of selective pressure plasmids tend to be lost from the host. Therefore in case where a long period of expression is needed it is recommendable to change ampicillin for carbenicillin which is more stable.

Antibiotic		Concentration	Antibiotic		Concentration
Amp	Ampicillin	100 µg/ml	Kan	Kanamycin	30 µg/ml
Cab	Carbenicillin	100 µg/ml	Rif	Rifampicin	200 µg/ml
Cam	Chloramphenicol	34 µg/ml	Str	Streptomycin	50 µg/ml
Gen	Gentamycin	20 µg/ml	Tet	Tetracycling	12.5 µg/ml

Table 2.2. Antibiotics. Table of the most used antibiotics and the appropriate concentration to kill plasmid free bacteria according to the antibiotic and plasmid providers.

A replicon is a particular sequence in the plasmid where the replication is initiated. Every plasmid has a unique replication site and the number of copies of each plasmid will be determined by the replicon type. The number of plasmid copies available for the host could determine as well the expression levels of our protein: the more copies there are the more mRNA could be synthesize at the same time. When more than one plasmid is present in the host it is very important to take into account the compatibility of both origin of replication and the selection marker (Held et al., 2003). This situation could be reached when we are willing to co-express different proteins cloned in different vectors or when we are using some *E. coli* strains that have an additional plasmids containing tRNAs or other specific proteins as mentioned before (Table 2.3).

Replicon	Plasmid	Copy number
ColE1 (pBR322)	PET, pETDuet, pGEX, pMAL, pProEX	~ 40
ColE1 (pUC)	PETBlue, pRSET	> 500
P15A (pACYC184)	pOKD4, pLysS, pRARE, pACYC, pACYCDuet	10-12
CloDF13	pCDF, pCDFDuet	20-40
RSF1030	pRSF, pRSFDuet	> 100
COLA (ColA)	pCOLADuet	20-40

Vectors combination						Host Strain Group
ColE1 (pBR322)	ColE1 (pUC)	P15A (pACYC184)	CloDF13	RSF103	COLA (ColA)	A
ColE1 (pBR322)	ColE1 (pUC)	CloDF13	RSF1030	COLA (ColA)		B
ColE1 (pBR322)	ColE1 (pUC)	CloDF13	COLA (ColA)			C

Table 2.3. Replicon types and combination. In the first table is shown which replicon is compatible with others and the copy numbers stimated in bacteria. In the second table 3 host groups are described this the possible vectors combination. Group A: BL21(DE3), C41(DE3), C43(DE3); Group B: BL21(DE3), C41(DE3), C43(DE3), BL21(DE3)pLysS, BL21(DE3)CodonPlus, BL21(DE3)pLysS CodonPlus, Rosetta(DE3), Rosetta(DE3)pLysS; Group C: BL21(DE3), C41(DE3), C43(DE3), BL21(DE3)pLysS, BL21(DE3)CodonPlus, BL21(DE3)pLysS CodonPlus, Rosetta(DE3), Rosetta(DE3)pLysS Origami(DE3), Origami(DE3)pLysS, Rosetta-gami(DE3), Rosetta-gami(DE3)pLysS.

The responsible of initiating the transcription if a particular gene is the promoter which is located 10 to 100 nucleotides upstream of the ribosome binding site. Our protein encoding gene should ideally be under the control of a promoter which should be easily inducible and cost-effective, with a strong RNA polymerase sequestrating ability to over-express the target protein (10-50% of the total cellular protein) and finally with a low basal expression levels specially for these toxic proteins for bacteria. Depending on our interest we can choose among different available promoters (Balzer et al., 2013) (Table 2.4). T7 promoter is perhaps the most used in expression plasmids because its expression levels could be really high at which the target protein could represent the 50% present of the total cell protein amount. In this system phage T7

RNA polymerase recognizes the promoter and starts the transcription, therefore it is a requirement for bacteria to posses this polymerase. In order to do this a prophage (λDE3) encoding for the polymerase is placed in the bacterial genome under the control of the transcriptional of lacUV5 promoter. Basal expression could be controlled by lacIQ repressor or by co-expression of T7 lysozyme provided in a compatible plasmid pLysS or pLysE. Another phage promoter would be pL (leftward promoter). This promoter is highly repressed by λcI repressor protein which is integrated in the bacterial chromosome or in an additional plasmid. Transcription induction is achieved by adding IPTG. A mutant of this protein (λcI857) is temperature-sensitive and at higher temperatures than 37ºC (40-42ºC) it becomes unstable allowing a moderate to high gene transcription. CspA promoter is a example of low-temperature and IPTG inducible promoter which is based in the same principle as the previous one. This promoter is interesting so as to produce target proteins at low temperatures. However, it has a leak at 37ºC which could jeopardize bacterial life in case the protein is toxic. AraBAD promoter could be recognized by bacterial polymerase. It has a variable expression level and it is governed by AraC protein, which has a dual repressor/activator role. When the inducer l-arabinose is absence AraC protein binds to two sites in the promoter creating a loop. Nevertheless, when the inducer is present AraC changes and promotes the transcription. Finally, the synthetic trc (hybrid) promoter has derived from trp and lacUV5 promoters. It was design to be stronger and to regulate the transcription tightly. As T7 promoter, trc can be inhibited by lacI or lacIQ repressor and it can be induced by IPTG. The expression levels achieved are moderate to high.

Promoter	Description	Inductor	Repressor	Advantages	Disadvantages
T7	T7 RNA polymerase dependent promoter (λDE3).	IPTG	*lacIQ*	High level exp.: up to 50 % of total protein. Well characterized. Widely used.	Leaky exp.: pLysS strain has to be used in case of toxic protein exp.
pL	Bacterial polymerase dependent promoter. Thermal sensitivity due to λcI857	Temperature shift to 40-42ºC	λcI857	Efficient exp. at high temperatures.	Leaky exp.
CspA	From the major cold-shock protein in *E. coli*. Bacterial polymerase dependent promoter.	Temperature shift to 10-25ºC		Efficient exp. at low temperatures.	Leaky exp.
araBAD	Bacterial polymerase dependent promoter. Relies on positive control.	l-arabinose	TetAraC	Tight regulation. Low basal exp.: toxic protein exp.	Repressed exp. state is not always zero. Gene dependent
trc (hybrid)	Combination of *trp* and *lac*UV5 promoters.	IPTG	*lacI or lacIQ*	High level exp.: 15 – 30 % of total protein.	Very leaky exp.

Table 2.4. Table of most used promoters and their characteristics. A brief description is provided. Exp = Expression.

Usually all these plasmids have an specific region where different and unique DNA restriction sites are located so as to make molecular cloning easier. Unfortunately, not all plasmids have the same **MCS** and when we are willing to do a parallel cloning into different vectors we have to take into account the possible enzyme combinations among them as well as the insertion of our DNA fragment in the correct reading frame in each plasmid.

Although originally they were developed to facilitate the recombinant proteins detection and purification, nowadays tags and fusion proteins have become positive elements for the folding of their fusion partners, solubility and consequently yield. Therefore, affinity tags are for structural and functional proteomics indispensable tools (Table 2.5).

Short peptide tags are convenient for protein detection and purification but do not often improve protein solubility. They don't usually interfere with fused partner but they may affect in the tertiary structure of it depending on the amino acid composition or position. For this reason, in case we have structural information of the protein is recommendable to analyse in which terminus is more accessible to the solvent. Poly-His, Poly-Arg, S-tag, FLAG and Calmdulin Binding Peptide (CBP) are commonly used nowadays as they have commercial tools such as antibodies for detection or specific affinity matrix for the purification.

Even if the mechanism why they improve the solubility of fused partners remains unclear, fusion proteins are widely used for this purpose. Some of them have to be expressed with another short peptide tags as they have not any kind of affinity for a specific resin. Most popular fusion proteins are Glutathione S-Transferase (GST), Maltose-Binding Protein (MBP), N-Utilization Substance (NUS), Thioredoxin (Trx), Ubiquitin, Green-Fluorescence Protein (GFP) and Small Ubiquitin-like Modifier (SUMO). Nonetheless there are some drawbacks about fusion proteins: Due to the large size of fusion tags they might jeopardise the final structure and activity of target protein. In case of MBP it has been reported that this fusion protein may form soluble inclusion bodies leading to a erroneous detection of soluble protein. For several applications such crystallization, NMR measurements and antibody production tags have to be removed because they may interfere in data acquisition and final result. Sorrowfully, when tag removal, it is unpredictable to know whether the target protein will remain soluble or will aggregate.

Tag	Size (AA)	kDa	Location	Expression	Solubility enhancer	Affinity tag	Affinity matrix	Metabolic burden	Secretion	Detection	Comments
Poly-His	6-10 (Usually 6)	0.84	N, C, internal	No	No	Yes	Immobilized metal resin (Ni^{2+} or Co^{2+})	Low	No	Yes	Inexpensive IMAC is not highly specific
Poly-Arg	5-15 (usually 5)	0.8	C	(refolding)	No	Yes	Cation-exchange resin	Low	No	No	Inexpensive. Low specificity
S-tag	15	1.75	N, C, internal	No	No	Yes	S-fragment of RNaseA	Low	No	Yes	Expensive resin. High affinity.
FLAG	8	1.01	N, C	No	No	Yes	Anti-FLAG monoclonal antibody	Low	No	No	Expensive resin. High affinity.
CBP	26	2.96	N, C	No	No	Yes	Calmodulin	Low	No	Yes	Not expensive. High affinity
GST	214	26	N	Yes	Yes	Yes	Glutathione agarose	High	No	Yes	Inexpensive. Homodimeric.
MBP	396	40	N, C	Yes	Yes	Yes	Amylose resin	High	Yes	Yes	Inexpensive. Efficient translation initiation
NUS	495	54	N	Yes	Yes	No	-	High	No	No	Not affinity tag, needs one.
Trx	109	11.7	N, C	Yes	Yes	Yes	Phenylarsine oxide agarose	Medium	No	No	Inexpensive. Medium affinity
Ubiquitin	76	8.4	N	Yes	Yes	No	-	Low	No	No	Not affinity tag, needs one.
GFP	220	24.2	N, C	No	No	No	-	High	No	Yes	Expressed for visualization. It may improve solubility. Not affinity tag, needs one.
Sumo	100	11	N	Yes	Yes	No	-	Medium	No	No	Not affinity tag, needs one.

Table 2.5. Table of most used tags and their characteristics. A brief description is provided. Exp = Expression.

Tag removal can be done by chemical methods or by enzymatic digestion (Hwang et al., 2014). As chemical methods may be too aggressive for the survival of target protein enzymatic digestion is the one largely used. In this case expression vectors are provided by a specific protease cleavage site usually between the fused tag or protein and our target protein. Most frequently used proteases are enterokinase, thrombin, factor X_a and tobacco etch virus (TEV). Enterokinase and thrombin were the first proteases used for this purpose. Cleavage takes place at temperature ranges from 20 to 37ºC. However, depending on the final conformation of the fused protein and the protease site's exposure to the solvent non-specific cuts are common while using these. Factor X_a shows more selectivity than previous two and cleavage can be carried out from 4 to 25ºC. Nonetheless, cleavage of the tag by this protease has been proven to be ineffective or non-specific. TEV protease is perhaps the best choice for this purpose due to its high specificity and the temperature range at which cleavage is done (from 4 to 37ºC degrees where 34ºC is the optimum).

2.1.5 Screening And Selection Of Recombinant Protein

The combination of all these strategies described before to improve protein solubility may lead to the generation of a huge collection of variants of our target protein. Moreover it is very difficult to foresee which of these combinations will result in a soluble protein. Therefore, the more clones are tested the more a the probabilities of obtaining estimated our goal. Depending on the amount of recombinant proteins to be checked and their physio-chemical properties we could choose the more suitable method of assaying for protein solubility.

If we have to deal with a small amount of clones we could analyse protein solubility by hand. Perhaps the most used method for this purpose is the **SDS-PAGE** visualization of the physical fractionation of the supernatant and pellet of a bacterial lysate. Another useful method is to do small expression test in a **96-well plates**. After expression all these plates are filtered so as to separate the soluble from the insoluble fraction and after could be analyse by SDS-PAGE. However, as this method generates to many samples the easiest way to confirm protein solubility it to perform a dot-blot from the soluble fraction which could be robotised making the screening faster. Nevertheless, dot-blot analysis does not provide information about protein size, fragmentation or quality of it, therefore positive samples are further studied by SDS-PAGE or other techniques. Finally, probably the fastest way to analyse a big amount of clones is the **colony filtration blot** (Cornvik et al., 2005) where target protein's solubility is directly studied from bacteria grown in a Petri dish. Whilst only protein presence in the supernatant is reported.

When we are dealing with huge libraries and we need to increase our throughput the most convenient tactic is to attach a fusion protein as a solubility reporter. The addition of **GFP** has been described to be very beneficial for this purpose (Waldo et al., 1999). The main advantage about this method is that protein solubility can be tested directly by visualization of GFP

fluorescence in the bacteria. When the fluorophore is attached to an insoluble protein GFP fluorescence decreases due to the protein maturation inhibition, resulting in a lack or low fluorescence levels in bacteria. When GFP is linked to a soluble protein green cells are detectable. Native fluorescent proteins forms dimers in solution which could lead to misleading results. However, nowadays there are monomeric mutants which prevent this situations. Similarly, fusion of **Chloramphenicol Acyl Transferase** (Maxwell et al., 1999) (CAT) reports protein solubility. Based on the resistance to the presence of increasing amounts of chloramphenicol in the growing media protein solubility may be accompanied with the bacterial survival. Nonetheless, CAT forms trimers which could generate unpredictable effects but this method is preferable to analyse even more clone numbers than GFP due to its high selectivity. Unfortunately both techniques may score false positives due to the solubility perturbation produced by these fusion proteins on the target. Therefore fusion protein removal is essential for further analysis.

2.2 MATERIALS AND METHODS

2.2.1 MOLECULAR BIOLOGY TECHNIQUES

The human **K_v7.2** (Q2) (UniProtKB/ Swiss-Prot: 043526) and **K_v7.3** (Q3) (UniProtKB/ Swiss-Prot: 043525) were kindly provided by Prof. Thomas J. Jentsch (Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany) (Biervert et al., 1998). The rat **CaM** cDNA was obtained from Prof. Adelman (Vollum Institute, Portland, OR, USA).

From the **constructs of K_v7.2 channels' C-terminus** were done:

- Helices ABCD-L → residues G310-K844 → WT version of the C-terminus from
- Helices ABCD → residues G310-G632 → Truncated after helix D
- Helices ABc → residues G310-D549 → Truncated in middle of helix C

Following **deletions between helices A and B** were combined with previous constructs:

- Del2 (Δ2) → ΔT359_T501
- Del6 (Δ6) → ΔR374_K493
- Del6 (Δ6L) → ΔR374_T501

From the **constructs of K_v7.3 channels' C-terminus** were done:

- Helices ABCD → residues G349-K661 → Truncated after helix D
- Helices ABc → residues G349-D556 → Truncated in middle of helix C

Codon optimization:

Q2ABCD Δ2, Q2ABCD Δ6L and Q3ABCD sequences were optimized for *E. coli*'s codon-usage. these genes synthesis was entrusted to the company *ShineGene*. Later on, optimized constructs were shorted to Q2ABc Δ2, Q2ABc Δ6L and Q3ABc in order to obtain more variability.

Synthesis of orthologous K_v7 channels:

DNA sequences were optimized and artificially produced by *ShineGene* corporation according to the amino acid sequences available in UniProt database: for *Caenorhabditis elegans*, Q3MQC5 and G5EE44; for *Ciona intestinalis*, C0JSV5 and C0JSV8; for *Danio rerio*, B8JIR6 and F1QD08; for *Drosophila melanogaster*, B7YZR3; and finally, for *Petromyzon marinus*, C0JSV7 and C0JSV6.

List of **plasmids** used in this chapter for cloning:

Name	Provider	Size (bp)	Resistance	Tags	Cleavage
Easy-T	Promega (A1360)	3015	Amp		
pProEX-HTc	Invitrogen (10711018)	4780	Amp	His	TEV
pProEX-CaM (Homemade)	Based on pProEX-HTc (69864-3)	5211	Amp	His-CaM	Double TEV
pGEX-6P-1	GE Healthcare (28-9546-48)	4984	Amp	His-GST	PreScission
pET-44a(+)	Novagen (71122-3)	7311	Amp	His-Nus	Thrombin Enterokinase
pET28-HMT	(Van Petegem et al., 2004)	6473	Kan	His-MBP	TEV
pThioHis-A	Thermo-Fischer (K360-01)	4400	Amp	His-Trx	Enterokinase

***E. coli* strains** used:

Strain	Provider	Humanized?	Extra plasmid resistance?
BSJ	SUNY Stony Brook	No	No
BL21(DE3)	Novagen (69450-3)	No	No
C41(DE3)	(Arechaga et al., 2003)	No	No
C43(DE3)	(Arechaga et al., 2003)	No	No
CodonPlus(DE3)(RIL)	Agilent (#230240)	Yes	Cam
Rosetta(DE3)pLysS	Novagen (70956-3)	Yes	Kan
Rosetta-Gami(DE3)pLysS	Novagen (71057-3)	Yes	Kan, Tet, Str and Cam

Different cloning strategies have been followed to obtain the constructs studied in the present work:

2.2.1.1 PCR technique

The polymerase chain reaction (PCR) is a biochemical technique to amplify DNA several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. Moreover, by this technique different mutations can be introduced in the target DNA fragment. The amplification process is based on a set of cycles of heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA (T100tm Thermal Cycler, Bio-Rad). Different kinds of PCR reactions were assayed:

1. Simple PCR: to amplify sequences in cassettes not longer than 700 bp which allows the addition of new restriction on the edges. The polymerase used was the Expand High Fidelity polymerase (Ref.: 03300242001, Roche) because introduces deoxiadenosines in the 3' termini, which facilitates the insertion in the EasyT cloning vector.

2. Overlapping nested PCRs: useful to mutate residues located in the inner part of a fragment that is flanked by restriction sites. First, two simple PCR reactions using the Pfu Turbo polymerase (Ref.: 600250, Agilent) are done. Each PCR corresponds to half of the fragment where we want to introduce the mutations. The primers from each PCR are specifically design to hybridize between them at least in one extreme. Next step is to make a third PCR with the Expand polymerase to amplify the whole fragment with the deoxiadenosines at the extremes.

Primers were designed and synthesized by IDT (Integrated DNA Technologies). The PCR protocol consist on (Protocol P.1):

Protocol P. 1: PCR

DNA	H_2O	Primers	Deoxynucleotides	Polymerase Buffer	Polymerase
[0.1 µM] 0.2 µl	2.9 µl	[0.1 µM] 0.2 µl x 2	[2.5 mM] 2 µl	[10X] 2 µl	0.3 µl

Temperature cycles:

- 30 sec. at 98 ºC

- 10 sec. at 98 ºC
- 30 sec. at 57 ºC
- 30 sec. at 72 ºC

x 30

- 10 min. at 72 ºC
- Cooling: 4 ºC, ∞

With the PCR product, and depending on the case two different work-pathways were followed:

1. These PCR can be first sub-cloned in the pGEM-EasyT vector (Ref.: A1360, Promega). This is a linear vector that has a 3' terminal thymidine in both ends. These single 3’-T overhangs at the insertion site greatly improves ligation efficiency of a PCR product into the plasmids by preventing re-circularization of the vector and providing compatible overhang for PCR products generated by the Expand polymerase. It also contains a region within the alpha-peptide coding region of the enzyme beta-galactosidase.

2. PCR product can be directly digested by enzymes for further manipulation.

2.2.1.2 Fragment digestion, purification and ligation

cDNA-s are digested for at least 2 h at 37 °C with the adequate restriction enzymes and buffer (Protocol P.2). After neutralization of the digest reaction by adding 10 X loading buffer (Glycerol 50 %, 0.2 M EDTA pH 8, bromophenol blue 0.05 %), the different fragments are separated by electrophoresis in a 1 % agarose (Ref.: 8008, Laboratorios Conda) gel (120 ml TBE 1 X [1 L: 54 g Tris Base, 27.5 Boric Acid, 20 ml of 0,5 EDTA] + 1.2 g agarose). A Lambda

DNA marker (Ref: SD0011 Fermentas) digested with BstE II / Eco91 (ref: ER0391, Fermentas) is used as a reference of DNA weight. The agarose gels are stained with ethidium bromide which acts as an intercalating agent in between the nucleic acids, and which suffers an almost 20-fold increase in fluorescence upon DNA binding. Bands corresponding to the vector and the fragment respectively are removed from the gel and purified following manufactures instructions using the GeneJet Gel Extraction Kit (ref: K0692, Fermentas) . Once the vector and fragment are purified, the ligation is performed by using the T4 DNA ligase (Ref.: EL0014, Fermentas) overnight (O/N) at room temperature (RT) (Protocol P.3).

Protocol P. 2: Fragment digestion 2 h 37 ºC

	DNA	H_2O	Buffer [10X]	Restriction enzyme I	Restriction enzyme II
Fragment	[0.1 µg/ul] 2 µl	13.5 µl	2 µl	0.5 µl	0.5 µl
Vector	[0.1 µg/ul] 4 µl	11.5 µl	2 µl	0.5 µl	0.5 µl
PCR product	[0.05 µg/ul] 8 µl	7.5 µl	2 µl	0.5 µl	0.5 µl

Protocol P. 3: Fragment ligation O/N, RT

DNA Fragment	DNA Vector	Ligase Buffer [10X]	Ligase
[1 µg/ul] 7 µl	[1 µg/ul] 1.5 µl	1 µl	0.5 µl

2.2.1.3 Bacterial transformation

Already ligated vector and fragment are transformed in *E. coli* BSJ strain, which has been previously prepared to be competent. The purpose of this technique is to introduce foreign DNA into the bacteria and use it as a host to amplify the plasmid in order to obtain large quantity of plasmid copies. Three main mechanisms are used to transform bacteria, such as electroporation, chemical transformation or heat shock, but for cloning and DNA expression we use a heat shock. When we heat bacteria for 45 sec. in the presence of the ligated plasmid and we place them back in ice, we create small pores in the bacterial wall making the DNA uptake easier. Afterwards, to help bacterial cells recover from the shock, the cells are briefly incubated with non-selective growth media for around 40 min, spin down, get rid of the excess of supernatant, resuspended and finally, the cells are seeded into LB agar plates with the specific antibiotic (Protocol P.4).

Protocol P. 4: Heat shock transformation

- Thaw one vial (or more) of cells.
- In a Eppendorf of 1.5 ml add 50 µl of competent cells
- Add 5 µl of Ligation or 1 µl of a miniprep (1 µg/µl)
- Keep on ice for 20 mins.
- Place them in a thermoblock at 42 ºC for 45 secs.
- Place them back on ice for 5 mins.
- Add 500 µl of LB without antibiotics.
- Take it to the shaker at 37 ºC for 40-60 mins.
- Spin down the cells for 5 mins at 2500 g.
- Pour out 450 µl of the supernatant.
- Resuspend the pellet and seed a Petry dish with the corresponding antibiotic.
- Incubate the plates inverted at 37 ºC until colonies have grown (6 – 12 h)
- Plates can be stored at 4 ºC for weeks.

2.2.1.4 DNA extraction and quantification

To obtain small amount of DNA (minipreps), each colony is grown in 3 ml of LB medium with the appropriate selective antibiotic at 37 °C and 220 rpm O/N. After that 2 ml of the culture are transferred to a 2 ml eppendorf tube and spin at maximum speed for one minute to pellet cells. This pellet is then resuspended in TE 1 X (TE 10 X: 100 mM Tris-HCl pH 8, 10 mM EDTA), and then lysed with SDS/NaOH (0.2 M NaOH/ 1 % SDS), SDS denatures proteins while the NaOH denatures DNA. In less than a minute lysis time, cold neutralization buffer (3 M potassium acetate, 5 M acetic acid) is added and mixed immediately. This step neutralizes the mixture, allowing plasmid DNA to re-anneal. Next, to separate lipids from proteins, a solution made of phenol/chloroform (25:24) is added. The aqueous phase is removed and transferred to a new microcentrifuge tube. By adding isopropyl alcohol and centrifuging at maximum speed the DNA is precipitated, afterwards cleaned with 70 % ethanol and resuspended in TE 1 X with RNAse to degrade the RNA. With this procedure we can obtain between 1 and 2 µg of DNA per microliter.

Protocol P. 5: DNA extraction protocol

- Pick a colony and let it grow O/N in 3 ml of LB with antibiotics.
- Next day transfer 2 ml of culture to 2 ml Eppendorf and spin it to pellet the cells.
- Discard supernatant and respuspend the pellet in 150 µl of TE 1X.
- Add 150 µl NaOH/SDS lysis buffer. Mix gently inverting the tube 5-6 times.
- Add 150 µl of neutralizing buffer and mix.
- Add 80 µl of phenol/chloroform.
- Spin at maximum speed for 5 mins.
- Extract around 400 µl of the upper phase and place it in other eppendorf.
- Add 200 µl of isopropyl alcohol to the new eppendorf.
- Spin at maximum speed for 5 mins.
- Get rid of the isopropyl alcohol and wash the pellet with 70 % EtOH.
- Get rid of EtOH and dry the pellet.
- Dissolve the plasmid in 30 µl of TE+RNAse buffer.

2.2.2 *IN SILICO* ANALYSIS

The amino-acid sequences used in all the softwares were the following:

>K_v7.2 C-terminal

GILGSGFALKVQEQHRQKHFEKRRNPAAGLIQSAWRFYATNLSRTDLHSTWQYYERTVTVPMYR
LIPPLNQLELLRNLKSKSGLAFRKDPPPEPSPSQKVSLKDRVFSSPRGVAAKGKGSPQAQTVRRSP
SADQSLEDSPSKVPKSWSFGDRSRARQAFRIKGAASRQNSEEASLPGEDIVDDKSCPCEFVTEDLT
PGLKVSIRAVCVMRFLVSKRKFKESLRPYDVMDVIEQYSAGHLDMLSRIKSLQSRVDQIVGRGPAI
TDKDRTKGPAEAELPEDPSMMGRLGKVEKQVLSMEKKLDFLVNIYMQRMGIPPTETEAYFGAKE
PEPAPPYHSPEDSREHVDRHGCIVKIVRSSSSTGQKNFSAPPAAPPVQCPPSTSWQPQSHPRQGHG
TSPVGDHGSLVRIPPPPAHERSLSAYGGGNRASMEFLRQEDTPGCRPPEGTLRDSDTSISIPSVDHE
ELERSFSGFSISQSKENLDALNSCYAAVAPCAKVRPYIAEGESDTDSDLCTPCGPPPRSATGEGPFG
DVGWAGPRK

>K_v7.3 C-terminal

GILGSGLALKVQEQHRQKHFEKRRKPAAELIQAAWRYYATNPNRIDLVATWRFYESVVSFPFFRK
EQLEAASSQKLGLLDRVRLSNPRGSNTKGKLFTPLNVDAIEESPSKEPKPVGLNNKERFRTAFRM
KAYAFWQSSEDAGTGDPMAEDRGYGNDFPIEDMIPTLKAAIRAVRILQFRLYKKKFKETLRPYDV
KDVIEQYSAGHLDMLSRIKYLQTRIDMIFTPGPPSTPKHKKSQKGSAFTFPSQQSPRNEPYVARPS
TSEIEDQSMMGKFVKVERQVQDMGKKLDFLVDMHMQHMERLQVQVTEYYPTKN

2.2.2.1 Solubility predictors

The web-interface of the following list of free available softwares were used.

Software name	Creator/Described in	Web Page
CamSol	(Sormanni et al., 2015)	http://www-vendruscolo.ch.cam.ac.uk/camsolmethod.html
ccSol	(Price et al., 2009)	http://s.tartaglialab.com/grant_submission/ccsol
ESPRESSO	(Hirose & Noguchi, 2013)	http://mbs.cbrc.jp/ESPRESSO/TopPage.html
PROSO II	(Smialowski et al., 2012)	http://mips.helmholtz-muenchen.de/prosoII/prosoII.seam

2.2.2.2 Secondary structure predictors

The web-interface of the following list of free available softwares were used.

Software name	Creator/Described in	Web Page
PsiPred	(Jones, 1999)	http://bioinf.cs.ucl.ac.uk/psipred/
GOR5	(Sen et al., 2005)	http://gor.bb.iastate.edu/
PREDATOR	(Frishman & Argos, 1997)	http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::predator
Scratch	(Cheng et al., 2005)	http://scratch.proteomics.ics.uci.edu/

2.2.2.3 Intrinsically disordered region predictors

The web-interface of the following list of free available softwares were used.

Software name	Creator/Described in	Web Page
DisEMBL	(Linding et al., 2003)	http://dis.embl.de/
IUPred	(Dosztányi et al., 2005)	http://iupred.enzim.hu/
PONDR	(Romero et al., 1997)	http://www.pondr.com/

Output results were later analysed using R2014a MatLab scripts.

2.2.3 PROTEIN EXPRESSION AND SOLUBILITY ANALYSIS (PILOT EXPERIMENTS)

Generally, from a Petry dish separate colonies were picked and transferred to 3 mL of LB with appropriate resistance O/N at 37 ºC. Next day, a 1:50 dilution was performed into a 10 mL of fresh LB with the corresponding antibiotics. Bacteria were grown until OD_{600} reached 0.6-0.8 (2-3 h approx.). Afterwards, IPTG was added into a final concentration of 0.5 mM. Protein expression was done at 37 ºC for 3 hours. Later on, cells were harvested by centrifugation for 5 minutes at 5000 g. Supernatant (SN) was disabled. Bacterial-pellet was resuspended with 500 µL of PBS 1X with protease inhibitors (Roche, Ref.:04693116001). Later, samples were sonicated (20 s ON, 10 OFF, 3 cycles). In order to test the solubility, lysates were centrifuged at 20,000 g for 15 minutes. The SN was stored, and subsequently the pellet was washed two times with the same buffer using the same volume finally resuspending the pellet in 500 µl of buffer. From each sample (SN and Pellet) 40 µl were added to 10 µl of 5X Loading Buffer (Tris-HCl 0.25 M pH 6.8, 50 % glycerol, 10 % SDS and 0.0125 % of 2-betamercaptoethanol). These new samples were placed in a thermal block for 10 minutes at 95 ºC.

Protein separation was done by **PolyAcrylamide Gel Electrophoresis (PAGE)**. By this technique proteins migrate through the polyacrylamide gel depending on their hydrodynamic radius and charge. When SDS-PAGE gels are used (Laemmli, 1970) SDS is intercalated in the protein giving more-less the same charge to all the proteins in proportion to their size. Taking this into account, different polyacrilamide percentages were employed in the preparation of SDS-PAGE gels based on the protein size estimation. 15 µl of boiled protein samples were loaded into the gels and one cell was reserved to load pre-stained protein standards in order to estimate the Molecular Weight (Kaleisdoscope Prestained Standards, BioRad, Ref.: #1610324). Protein staining was done by a dissolution of Coomassie Brilliant Blue R-250 0.25 % (w/v) (BioRad, Ref.: 161-0436), metanol 45 % (v/v) and Acetic Acid 10 % (v/v).

Pictures were taken from all stained gels with Versadoc Imaging System (BioRad, Ref.: 170800) and analysed using Image J software (Schneider et al., 2012). In order to meassure the protein solubility percentage, target protein band intensity was measured either in the supernatant and pellet fractions. Consequently protein solubility percentage was measured as follows:

$$Protein\,solubility\,(\%) = \frac{Soluble\,protein\,\text{in}\,supernatant}{Total\,protein\,\text{in}\,supernatant\,\text{and}\,pellet}\,x\,100$$

2.2.4 PROTEIN PURIFICATION

2.2.4.1 CaM purification

From the Petry dish, 10 mL of LB were inoculated with competent cells from a single colony O/N at 37 ºC in presence of the appropriate antibiotic. Next day, a 1:50 dilution was performed into a 1 L of LB media with antibiotics. Bacteria were grown until OD_{600} reached 1. Afterwards, IPTG was added into a final concentration of 0.4 mM IPTG. Protein expression was done at 37 ºC for 6 h. Cells were collected by centrifuging at 5300 g for 15 min. SN was discarded. Cells were resuspended in 50 mL of *Lysis buffer* (50 mM Tris-HCl pH 7.5, 2 mM EDTA and 0.2 mM PMSF).

Cell lysis was carried out by 3 passes through French press machine (Avestin, Emulsiflex-C5) Lysated sample was submitted to 3 cycles of freezing and defrosting using dry ice with EtOH and a water bath of 37 ºC. Due to the high thermostability of CaM this process does not affect CaM solubility. Sample was centrifuged at 25000 g for 30 min. SN was later heated at 70 ºC for 10 min. Samples were centrifuged again at 25000 g for 30 min. The SN was directly poured into 10 mL of Phenyl-Sepharose resin which was previously pre-equilibrated with 10 CV of *Equilibrium buffer* (50 mM Tris-HCl pH 7.5, 5 mM $CaCl_2$, 0.1 M NaCl). The flow through was again loaded into the column. The second flow was stored and a first wash was done using 20 CVs of *Wash Buffer* (50 Mm Tris-HCl pH 7.5, 0.1 mM $CaCl_2$, 0.1 M NaCl). The column was washed newly with 10 CVs of *High Salt buffer* (50 Mm Tris·HCl pH 7.5, 0.1 mM $CaCl_2$, 0.5 M NaCl). Finally, the protein was eluted using *Elution buffer* (50 mM Tris·HCl pH 7.5, 1 mM EGTA). During the purification samples were saved from each step, and at the end these were loaded and analysed using and SDS-PAGE gel, staining it with Coomassie. Positive fractions were selected, dialysed using dialysis cellulose membrane (Sigma-Aldrich, Ref.: D927-100FT) against ddH_2O. Protein was later concentrated using Amicon Ultra-15 centrifugal filter units with a 3 kDa cut-off (Sigma-Aldrich, Ref.: Z677094-24EA) and protein concentration was measured using NanoDrop 3300 (ThermoScientific, Ref.: ND-3300).

2.2.4.2 CaM dansylation

CaM was diluted to 1 mg/ml in 10 Tris-HCl (pH 8.5) and 2 mM of $CaCl_2$ buffer. Dansyl chloride (2.17 mg/mL dissolved in acetone)(Sigma-Aldrich, Ref.: 39220) was added to a final concentration of 100 µM. Incubation was done in the dark at room temperature for 2 h vortexing every 20 min. Unincorporated dansyl was eliminated using a 1 ml G-25 sepharose column (Sigma-Aldrich, Ref.: G25150).

2.2.4.3 His fusion proteins

Different C-terminal fragments and mutants of $K_V7.2$ channel sequences were cloned into pProEX-HTc vector. 10 mL of LB were inoculated with competent cells O/N at 37 ºC in

presence of the appropriate antibiotic. Next day, a 1:50 dilution was performed into a 1 L of LB. Bacteria were grown until OD_{600} reached 0.6. Then, bacteria were chilled at 20 ºC for 30 mins and afterwards, IPTG was added into a final concentration of 0.3 mM IPTG at 20 ºC O/N. Cells were collected by centrifuging at 5300 g for 30 min. SN was discarded. Cells were resuspended again using 20 mL of *Buffer A* (120 mM KCl, 20 mM Tris-HCl pH 7.4, 20 mM imidazole, 2 mM DTT, protease inhibitor without EDTA (Roche, 04693132001)). Lysis was carried out by sonication (20 s ON, 10 OFF, 12 cycles). Centrifugation was done at 30,000 g for 30 min. Finally, the supernatant (soluble fraction) was transferred to a fresh tube while the insoluble cell lysate (inclusion bodies) was treated separately.

The inclusion bodies isolated by centrifugation were resuspended three times in 10 ml of the same buffer used for sonication and then centrifuged to remove any remaining soluble material. The precipitate was dissolved for 30 min at 4 °C in 10 ml of 20 mM Tris-HCl (pH 7.4), 120 mM KCl, 2 mM DTT and 6 M urea, mixing occasionally. The solution was centrifuged at 20.000 g for 20 min at 4 °C and the supernatant was diluted to a final concentration of approximately 1 mg/ml. Refolding was done with an “urea concentration gradient dialysis”. The solution of denatured protein was dialysed against 2 L of freshly made 4, 2, 1, 0.5, and 0 M Urea, with 20 mM Tris-HCl (pH 7.4), 120 mM KCl and 2 mM DTT. For each step the protein was dialyzed 10 h at 4 °C.

Refolded and soluble fractions were incubated separately for 2 hours at 4 ºC in agitation with 1 mL of Ni-Sepharose pre-equilibrated resin. Afterwards, the resin was placed into a column where the sample was loaded twice. A first wash was done using 15 mL of Buffer A. The column was washed with 20 mL of Buffer B (120 mM KCl, 20 mM Tris-HCl pH 7.4, 2 mM DTT, 60 mM imidazole). Finally, the protein was eluted with 20 mL of Buffer C (120 mM KCl, 20 mM K-HEPES pH 7.9, 2 mM DTT and 500 mM imidazole). Eluted volumes were collected. During the purification samples were saved from each step, and at the end these were loaded and analysed using and SDS-PAGE gel, staining it with Coomassie. Protein was stored.

2.2.4.4 His-CaM, His-MBP, His-NUS and His-Trx fusion proteins

Different C-terminal fragments and mutants of $K_v7.2$ channel sequences were cloned into pProEX-CaM, pET28-HMT, pET-44a(+) and pThioHis-A vectors respectively. These vectors provides the possibilities of purifying fusion proteins using poly-histidines as an affinity tags. In this case the purification procedure were based on His-tag and the purification protocol was same as *His fusion proteins*. However, inclusion bodies were not refolded. In fact, only soluble fraction was used in all the cases. During the purification samples were saved from each step, and at the end these were loaded and analysed using and SDS-PAGE gel, staining it with Coomassie. Proteins were stored.

2.2.4.5 GST fusion proteins

Different C-terminal fragments and mutants of $K_v7.2$ channel sequences were cloned into pGEX-6P-1 GST expression vector. Cells were grown at 37 °C in 1 l of LB media with antibiotics till OD_{600} 0.6-0.8. The expression of the fusion proteins was induced with 0.3 mM of IPTG for 3 h at 30 °C. Cells were harvested by centrifugation and suspended in 20 ml of chilled *GST buffer* (20 mM Tris-HCl (pH 7.4), 100 mM NaCl, 0.5 % Triton X-100, 2 mM DTT, with Protease inhibitors (Roche, Ref.:04693116001). After lysis by sonication, cell debris was removed by centrifugation at 30000 g for 30 min at 4 °C. Finally, the supernatant (soluble fraction) was transferred to a fresh tube while the insoluble cell lysate (inclusion bodies) was treated separately.

The inclusion bodies isolated by centrifugation were resuspended three times in 10 ml of the same buffer used for sonication and then centrifuged to remove any remaining soluble material. The precipitate was dissolved for 30 min at 4 °C in 10 ml of 20 mM Tris-HCl (pH 7.4), 100 mM NaCl, 5 mM DTT and 6 M urea, mixing occasionally. The solution was centrifuged at 20.000 g for 20 min at 4 °C and the supernatant was diluted to a final concentration of approximately 1 mg/ml. Refolding was done with an “urea concentration gradient dialysis”. The solution of denatured protein was dialysed against 2 L of freshly made 4, 2, 1, 0.5, and 0 M Urea, with 20 mM Tris-HCl (pH 7.4), 100 mM NaCl and 2 mM DTT. For each step the protein was dialyzed 10 h at 4 °C.

The soluble fraction and refolded proteins were incubated separately with glutathione-sepharose 4B (GE Healthcare, Ref.: 17-0756-01) previously equilibrated with *GST buffer*. Bound proteins were washed and then eluted with 50 mM Tris-HCl (pH 8.5) and 15 mM of reduced glutathione (Sigma-Aldrich, Ref.: G4251). Fractions were analysed by 10 to 12 % SDS-PAGE gels, dialysed, concentrated and stored at -20 °C. Protein concentration was measured using NanoDrop 3300 (ThermoScientific, Ref.: ND-3300). The initial experiments (see later) were performed with proteins purified from the soluble fraction to ensure that they were properly folded. In latter experiments, fusion proteins were solubilised from inclusion bodies with 6 M urea and then refolded. The results obtained with these proteins were indistinguishable from the soluble material. This protocol was implemented by Dr. Alessandro Alaimo.

2.2.5 SPECTROSCOPY

2.2.5.1 Dynamic Light Scattering

Light scattering is a physical phenomena in which light waves suffer a splitting in different frequencies when they go though a specific material. The way in which this dispersion happens is highly dependent on the structure of the material and, in solution, allows the determination of the size of the particles. Dynamic Light Scattering (DLS), or photon correlation spectroscopy, enables us to obtain the size distribution profile of different particles within a dissolution and also estimate the percentage of every single species present in sample.

DLS photometer strikes a monocromatic laser beam on our sample with a wavelenght of 650 nm approximately, which induces the polarization of the molecules in the solution becoming them as dispersion sources. A particle in brownianic movement will scatter the light in a series of fluctuations: the bigger the hydrodinamic radious (RH) of the particle is the bigger the fluctuation is. these fluctuations are measured and analysed over time. As a function of the recordered fluctuations the translational diffusion speed (D) is determined, and based on this, the Rh of the theoretical sphere that represent the particle based on Stokes-Einstein equation:

$$D = \frac{k_B T}{6 \pi \eta R_H}$$

where k_B is Boltzmann constant, T is the temperature, R_H is the hydrodynamic radius of the particle and η the viscosity of the media.

In order to know the heterogeneity of the purified proteins, dispersion of the samples was evaluated by Dynamic Light Scattering using a Zetasizer spectrometer (Malvern Instruments) All measurements where done using 80 µl of sample volume. The prismatic plastic cuvette was incised by a 5 mW and 633 nm He-Ne laser beam. Scattered light was detected by a photomultiplier located at 173 degrees. A digital analysis of the data gives rise to the size distribution based on the particle's volume. By the correlation function and the polydispersity index protein size and possible aggregates were determined

2.2.5.2 Fluorescence Spectroscopy

The fluorescence properties of a molecule depends on the electronic and vibrational states in which its energy levels are. Samples are first excited using an specific wavelength in which a photon is absorbed changing the relaxed electronic state to a excited electronic state. The way in which the excited molecules recovers the basal energetic state is by collisions with other molecules, and emitting its energy in form of fluorescence. Lost photons will have different energies and thereby different frequencies depending on the biochemical environment. Thus, analysing the different frequencies of the emitting light and their relative intensities by the

spectrometer we can determine the structural effect of a labelled protein with a ligand: if a structural change takes place the chemical environment of the target fluorophore in the labelled protein may change affecting also the emission spectra.

Dansyl-calmodulin (D-CaM) and GST-K_v7.2, MBP-K_v7.2 and Trx- K_v7.2 fusion proteins were dialysed against 2 L buffer containing 25 mM Tris-HCl (pH 7.4), 120 mM KCl, 5 mM NaCl and 2 mM $MgCl_2$ for 48 h, changing the buffer every 10–12 h. Final concentration of D-CaM used was 12.5 nM. The fluorescence emission spectrum was obtained with an Aminco Bowman Series 2 fluorescence spectrophotometer (SLM Aminco) in a final volume of 100 µl (using quartz cuvette) at 25 °C. The excitation wavelength was 340 nm, and emissions were registered from 400 to 660 nm.

For the titration experiment, increasing concentrations of each fusion protein were added to a cuvette containing D-CaM alone (in the presence of EGTA 10 mM) or in the presence of excess of calcium (4 µM free Ca^{2+}) in D-CaM binding buffer (25 mM Tris-HCl (pH 7.4), 120 mM KCl, 5 mM NaCl, 2 mM $MgCl_2$, 10 mM EGTA).

Fluorescence enhancements were plotted against fusion/D-CaM ratio or [fusion] in order to obtain concentration-response curves. Data were fitted to the Hill equation by curvilinear regression to estimate EC_{50} values for each curve. Data are shown as an average of ± S.E.M. of at least three independent experiments.

2.3 RESULTS

2.3.1 *IN SILICO* ANALYSIS OF THE C-TERMINUS OF K_v7.2 CHANNELS REVEALS PUTATIVE SOLUBLE PROTEIN

Due to the high complexity of this protein we decided to produce and purify in E.coli just the C-terminus of the channel. Even if we knew many things about our protein we examined its sequence with some *in silico* tools mentioned above.

Regarding *secondary structure* the closest program to the literature was SCRATCH, where four well-defined alpha helices can be found (Figure 2.2). In this case helix A appears fragmented in two helices, one from G315 to K332 and the other P335 to T349. Moreover, a small helix appear in between helix A and B, from W360 to E364. The appearance of this small helix caught our attention which will be discussed in next chapter. Nevertheless, helix C is a bit smaller compared to the definition found in the literature. In all the cases helix A appears splitted in two and the best defined helix is helix D.

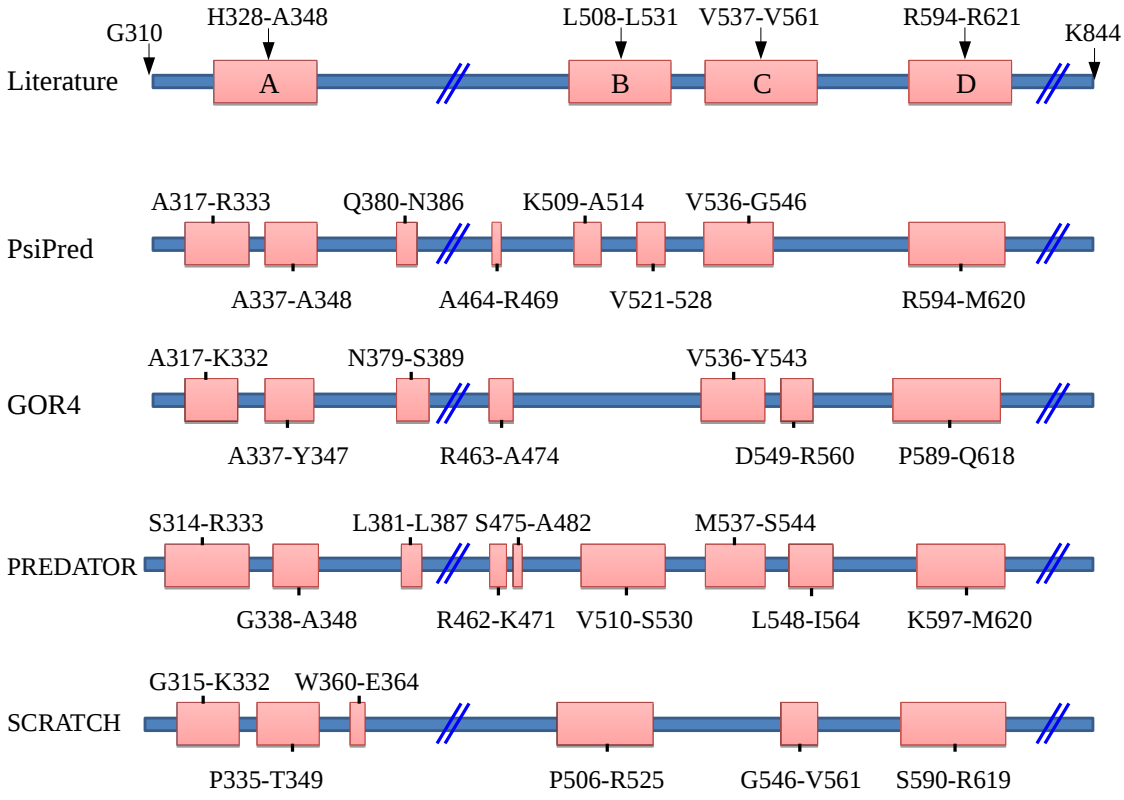


Figure 2.2. Secondary structure prediction *in silico*. Only results for the C-terminus of K_v7.2 are shown. Among PsiPred, GOR4, PREDATOR and SCRATCH softwares, the closer one to the literature is the result obtained from SCRATCH.

When we analysed the sequence in order to find putative *Intrinsically Disordered Regions* (IDR) with PONDR, DisEMBL and IUPred we saw that there is a big consensus among them (Figure 2.3). Accordingly to previous secondary structure predictions, the regions around helix A (H328-A348), helix B (L508-L531), helix C (V537-V561) and helix D (R594-R621) are defined as a ordered region (T349-K388) followed by a disordered region (S389-G507)(values bellow 0.5). We can notice that between helix C and D there is another disordered region. From the helix D to the carboxyl end disordered regions are mostly observable. Nevertheless other two small ordered regions appears (D653-T668 and Q781-R803). As we will discuss later, the presence of several IDRs may be a problem in the future when expressing the protein in a soluble manner as they are known to be formed by low content of hydrophobic amino acids and a high proportion of polar and charged amino acids.

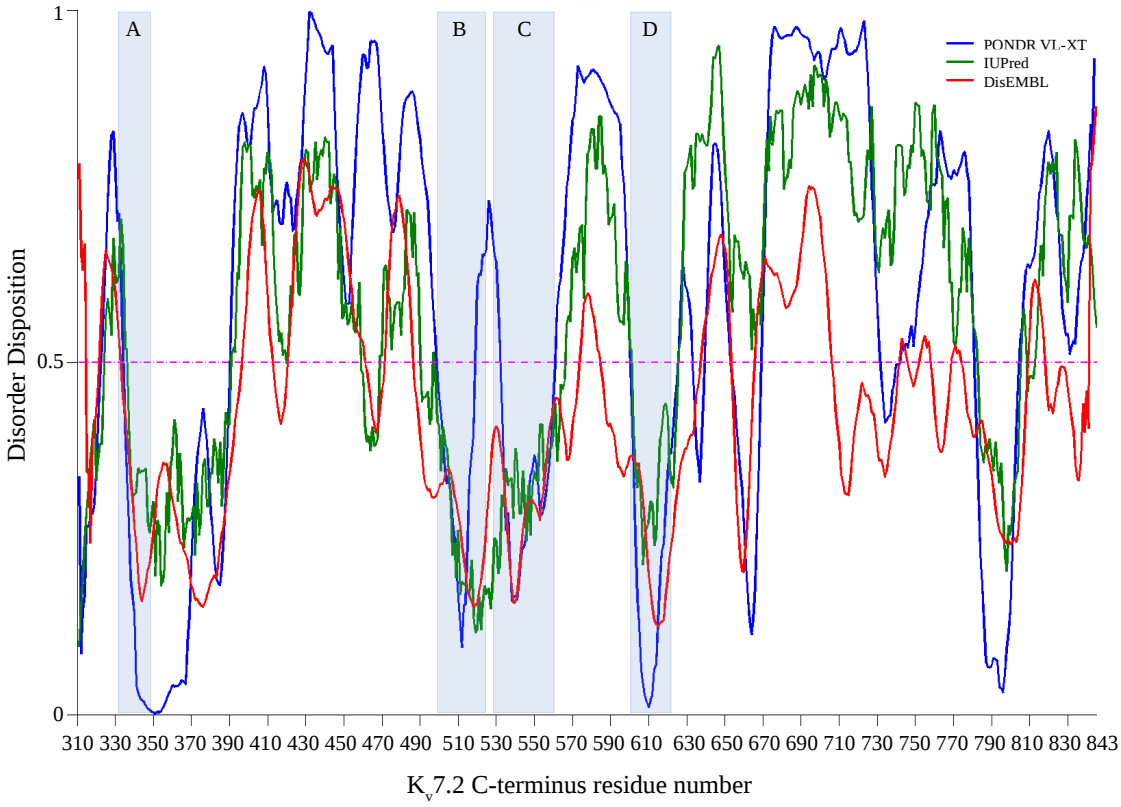


Figure 2.3. Intrinsically Disordered Regions prediction *in silico*. PONDR, IUPred and DisEMBL results are highly comparable. Values bellow 0.5 are considered to be ordered. Values above 0.5 are considered to be disordered. In light blue boxes previously defined helices A, B, C and D regions are highlighted. More than the 50 % of the C-terminus of K_v7.2 channels are predicted to be disordered.

Moreover, when we examine the protein *solubility and stability* using ccSOL, ESPRESSO and PROSOII programs we find out that the three of them claim that the C-terminus should be soluble and stable when expressed in *E. coli*, with a probability of the 97%, 74.2% and 85.7% respectively.

2.3.2 C-Terminal Constructs Are Intrinsically Insoluble When They Are Over-Expressed

Two versions of the carboxyl end of K_v7.2 channels (Q2ABc and Q2ABCD) were amplified by PCR, and later cloned in pProEX-HTc vector. This plasmid adds six histidines in the N-terminal end of the protein which does not influence the intrinsic properties of the protein. Besides this tag is helpful to purify out target protein using Ni-Sepharose resin as we have previously mentioned. these recombinant proteins were transformed into *E. coli* BL21(DE3) strain and induced first at 37ºC with 1 mM IPTG for 3 h. We lysated the cells by sonication. Soluble and insoluble fractions were separated by centrifugation and analysed by SDS-PAGE. In contrast to the solubility predictors, no protein was visible in the supernatant.

Afterwards, we proceeded to express the protein at 18 ºC with 0.05 mM of IPTG O/N so as to help the protein in its folding. So far, we found that the outcome was as poor as the first one. Therefore we decided to resolubilise the inclusion bodies in which our target protein would be misfolded. In order to recover the solubility of the protein urea was used as a chaotropic agent in high concentration (6 M). Later on, urea was removed by dialysis. It has to be mentioned that during this process protein aggregates started to appear inside the dialysis membrane. After eliminating the urea the remaining protein was incubated with Ni-Sepharose resin recovering almost nothing from the inclusion bodies (Figure 2.4).

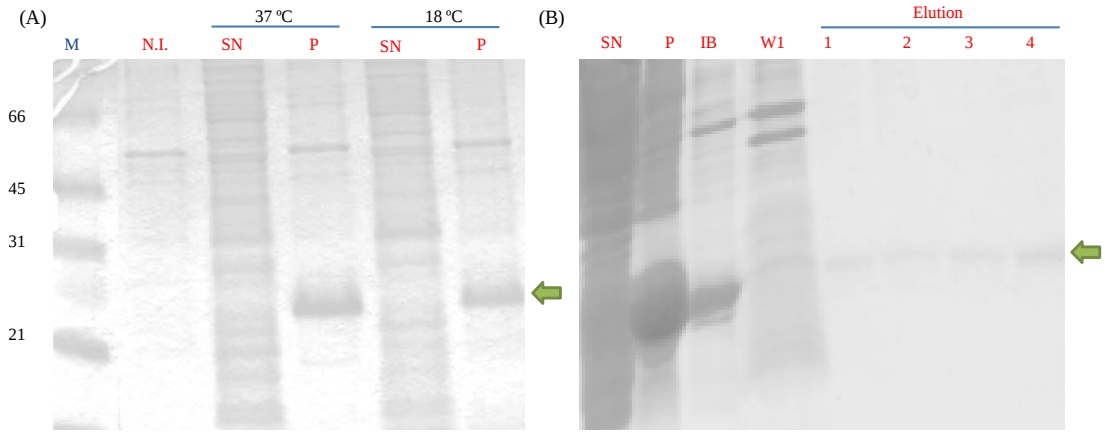


Figure 2.4. Over-expression and purification of His-Q2ABc protein (28 kDa). (A) SDS-PAGE gel at 15 % of polyacrylamide where different temperatures are tested at 37 ºC and 18 ºC. M, Marker; N.I., Not Induced; SN, supernatant and P, Pellet. (B) Resolubilisation of inclusion bodies. SN, supernatant; P, Pellet; IB, Inclusion bodies resolubilisation done with 6 M of urea; and W1, wash with 60 mM imidazole. Elution was done with 400 mM of imidazole.

Due to the fiasco of obtaining soluble protein without any kind of fusion protein, we proceeded to explore new strategies as mentioned in the introduction of this chapter.

2.3.3 GST-Tagged Proteins Were Soluble After The Resolubilisation Of Inclusion Bodies

Different constructs were cloned into pGEX6P1. GST tag is located in the N-terminal of the recombinant protein and is well-known to be able of improved protein solubility and folding. This is why we first tried to get soluble protein with this tag. these recombinant proteins were transformed into *E. coli* BL21 strain as above. We tested several temperatures so as to find out which one could be beneficial. We found out that in cases were the temperature was colder we could obtain a very little portion of soluble protein (at 20ºC 0.05-0.5 mg/ml) (Figure 2.5.A). However, the bigger part of the protein was still targeted to inclusion bodies. We also checked if the IPTG concentration could influence the solubility, but we didn't obtain any good result. Therefore, we decided to resolubilise protein aggregates in order to produce larger amounts.

Proteins resolubilisation was carried out using big amounts of urea as mentioned in materials and methods. Later on, urea was removed by dialysis and protein purification was done using a Glutathione-Sepharose column (Figure 2.5.B). In order to determine whether or not the proteins were properly folded we completed Dynamic Light Scattering (DLS) approaches. We selected the samples that presented values of 53,1 ± 1,9 kDa and 0,2 polydispersity (PDI). By this method we were able to select properly folded proteins discarding aggregates.

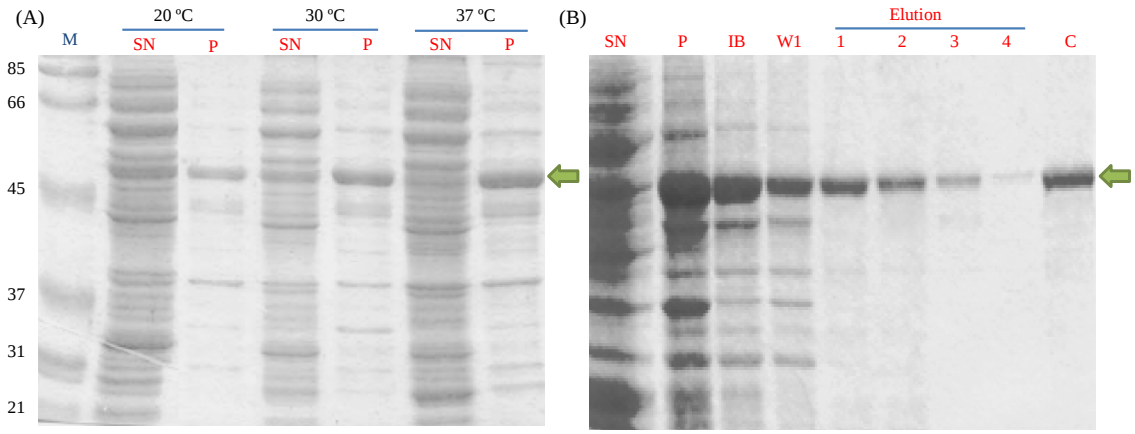


Figure 2.5. Over-expression and purification of GST-Q2ABc protein (28 kDa). (A) SDS-PAGE gel at 12 % of polyacrylamide where different temperatures are tested at 20 ºC, 30 ºC and 37 ºC. M, Marker; SN, supernatant and P, Pellet. (B) Resolubilisation of inclusion bodies. SN, supernatant; P, Pellet; IB, Inclusion bodies resolubilisation done with 6 M of urea; and W1, wash with 5 mM glutathione. Elution was done with 20 mM of glutathione. C, Concentrated sample.

Nevertheless, even though we succeeded in obtaining soluble protein, still, the amounts were small, and the effort to obtain these few mg were not worth it. Moreover, proteins resolubilisation didn't always properly work and the yields were too oscillatory. Thereby, refolding of proteins was not enough to set up a high throughput project for structural determination. This is why we continued our search for better alternatives following Figure 2.1 scheme.

2.3.4 Humanized Strains Are A More Convenient Alternative To The Codon-Usage Optimization Of The Target DNA Sequence

In order to screen which strain was more suitable for the expression K_v7 channels' C-terminus we decided to adopt the strategy of parallel expression of 3 non-optimized genes (His-Q2AB, GST-Q2ABc, GST-Q3ABc) and 3 optimized genes (GST-Q2ABCD Δ2, GST-Q2ABCD Δ6L, GST-Q3ABCD) in *E. coli* BL21(DE3), C41(DE3), C43(DE3), CodonPlus(DE3)(RIL), Rosetta(DE3)pLysS and Rosetta-gami(DE3)pLysS strains. Pilot experiments were done as mentioned in materials and methods. Equivalent amounts of lysated bacteria were loaded into a SDS-PAGE gel. After staining with Coomassie, total protein expression was taken into account when gels were analysed using ImageJ (Table 2.6). So as to compare the protein production among these strains, BL21(DE3) was used as a reference strain due to its simplicity.

Construct	BL21	C41	C43	CodonPlus	Rosetta	Rosetta-gami
GST-Q3A	1 ± 0.085	0.861 ± 0.117	0.948 ± 0.079	2.561 ± 0.348	2.06 ± 0.179	1.655 ± 0.229
GST-Q2ABc	1 ± 0.077	1.192 ± 0.084	1.253 ± 0.131	1.171 ± 0.08	1.21 ± 0.09	0.901 ± 0.072
His-Q2AB	1 ± 0.115	1.22 ± 0.289	1.013 ± 0.087	3.134 ± 0.216	2.302 ± 0.202	2.206 ± 0.208
GST-Q2ABCD Δ2	1 ± 0.087	1.92 ± 0.149	1.911 ± 0.117	1.667 ± 0.186	2.019 ± 0.277	1.215 ± 0.098
GST-Q2ABCD Δ6L	1 ± 0.091	1.426 ± 0.108	2.139 ± 0.441	2.129 ± 0.255	1.832 ± 0.142	1.303 ± 0.279
GST-Q3ABCD	1 ± 0.102	1.167 ± 0.275	1.51 ± 0.18	1.168 ± 0.096	1.712 ± 0.13	0.924 ± 0.09

Table 2.6. Bulk protein expression comparative to BL21(DE3) strain. 3 optimized genes (GST-Q2ABCD Δ2, GST-Q2ABCD Δ6L, GST-Q3ABCD) and 3 non-optimized genes (His-Q2AB, GST-Q2ABc, GST-Q3ABc) are expressed in 3 humanized strains [CodonPlus(DE3)(RIL), Rosetta(DE3)pLysS and Rosetta-gami(DE3)pLysS)] and 3 non-humanized strains [BL21(DE3), C41(DE3), C43(DE3)]. Therefore 4 groups are created: In orange, non-optimized genes expressed in non-humanized strains; in blue, non-optimized genes expressed in humanized strains; in green, optimized genes expressed in non-humanized strains; and in purple, optimized genes expressed in humanized strains. Upper value in each cell represents times over-expressed and bellow the standard error.

Extracted data were plot using a script in MatLab. Subsequently, it was analysed whether or not the codon usage could influence the protein expression in *E. coli*. On one hand, when we compared the behaviour of non-optimized sequences with all these strains we clearly observed a protein expression improvement in humanized strains (Figure 2.6.A). Among them, BL21(DE3)CodonPlus was the best strain to express our target recombinant proteins. On the other hand, when we studied the optimized sequences we saw that not statistical differences were observable among them (Figure 2.6.B).

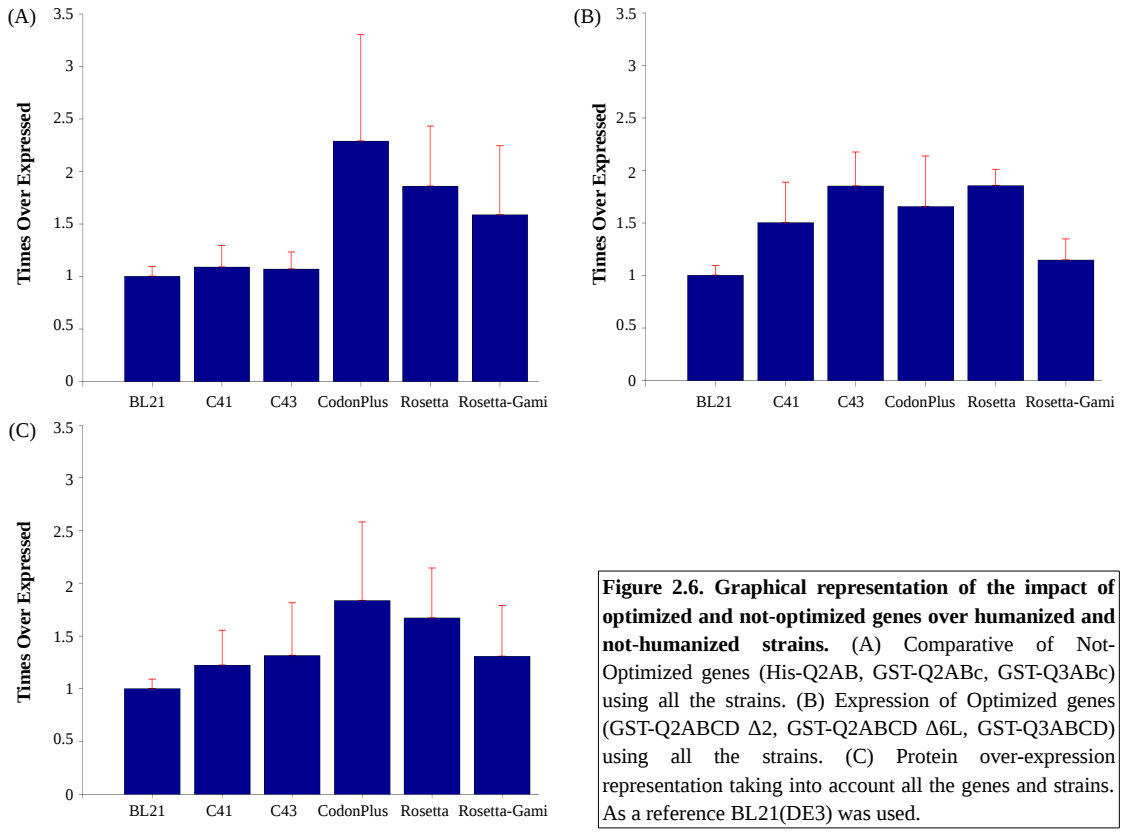


Figure 2.6. Graphical representation of the impact of optimized and not-optimized genes over humanized and not-humanized strains. (A) Comparative of Not-Optimized genes (His-Q2AB, GST-Q2ABc, GST-Q3ABc) using all the strains. (B) Expression of Optimized genes (GST-Q2ABCD Δ2, GST-Q2ABCD Δ6L, GST-Q3ABCD) using all the strains. (C) Protein over-expression representation taking into account all the genes and strains. As a reference BL21(DE3) was used.

However, it is worth to be mentioned that one of the biggest trouble comparing optimized and not-optimized sequences is that we don't have a single construct in both versions (optimized and not-optimized). Therefore, we are not able of doing a direct comparison. Nonetheless, we could assume that some of constructs we used should have similar nature.

Taken into account our handicap and after summarizing the general outcome, our data strongly suggest that in our case BL21(DE3)CodonPlus is the best strain to express our target recombinant proteins (Fig 4.C). Rosetta2(DE3)pLysS and Rosetta-Gami(DE3)pLysS are good strains too. Their protein production is a little bit lower due to their pLysS activity specially in Rosetta-Gami(DE3)pLysS. Nevertheless, the protein background is lower too so they could be interesting strains to work with, always taking into account the constructs and the researcher's purposes. On the contrary, C41 and C43 cells have shown a better behaviour when compared to BL21, but when expressing non-optimized sequences they may have some problems.

As a conclusion, codon-supplemented strains are more suitable alternative to the codon-usage optimization of the target DNA and there are two reasons that support it: first, we have seen that there is not a significant protein expression improvement when we used optimized sequences in not-humanized bacteria, and second, optimizing codon-usage is much expensive than acquiring a humanized strain.

2.3.5 Unlike Other Tags, CaM Fusion Proteins Are Soluble And Properly Folded

Once we knew that BL21(DE3)CodonPlus had a bigger expression, next step was to characterize six different fusion proteins in order to detect which one had the ability to promote the solubility of their fusion partners. Different alternatives of the C-terminus of K_v7.2 and K_v7.3 channels were cloned and expressed in parallel with Poly-His tag, GST, MBP, NUS and Trx. We also decided to use CaM as a fusion partner.

By **pilot experiments**, 38 constructs were tested: 4 for Poly-His, 13 for CaM, 10 for GST, 2 for MBP, 7 for NUS and 2 for Trx. Equivalent supernatant and pellet amounts were loaded into SDS-PAGE gels. After staining with Coomassie, gels were analysed using ImageJ. The effects of different fusion proteins upon solubility are shown (Appendix P1). An example of Q2ABc protein behaviour is also given when it is expressed with these tags (Figure 2.7).

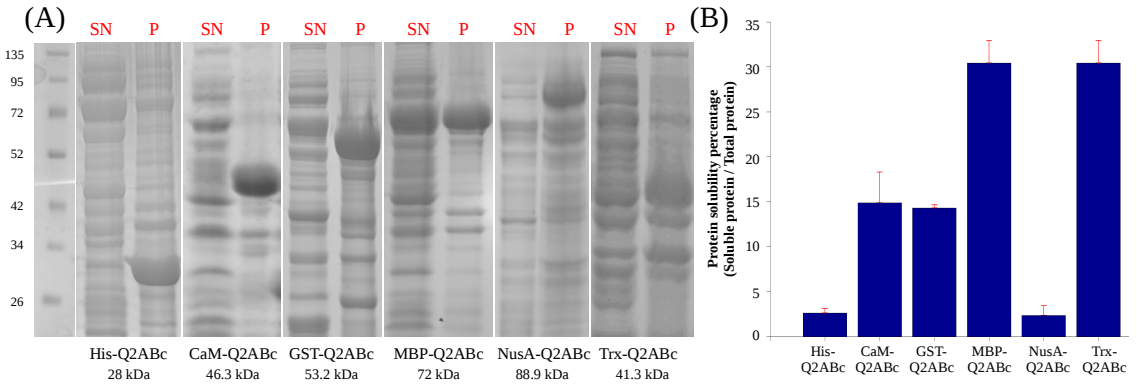


Figure 2.7. Q2ABc protein solubility using different fusion proteins. Proteins were expressed at 37 ºC for 3 h with 0.5 mM of IPTG. (A) SDS-Polyacrilamide gels showing SN, Soluble; and P, insoluble fractions. Marker sizes are given in the first column. (B) Chart with measured protein solubility for each case. MBP and Trx are the best solubility enhancers.

Results reveal that fusion protein technology is beneficial in order to improve solubility. Among all these partners MBP and Trx were by far the most promising. They were able to solubilise the whole C-terminus of K_v7.2 channels (Appendix P1). Therefore we decided to scale up the production of Q2ABCD and Q2ABc with MBP and Trx. Purification in all cases were successfully, reaching yields up to 5 mg/ml. However, we were still not sure whether these proteins were properly folded and able to interact with CaM. In order to verify we carried out a **fluorimetric assay** with Dansyl-Calmodulin and GST-Q2ABc, MBP-Q2ABc and Trx-Q2ABc proteins (Figure 2.8.A and B)

When 12.5 nM of D-CaM was alone in the fluorescence buffer, the collected spectra showed a pike around 510 nM. But when we added saturating $CaCl_2$ (9,7 mM) (2 µM of free Ca^{2+}), the emission peak increased and shifted to the left, around 485 nm (Figure 15.A). In the absence of Ca^{2+} (10 mM EGTA), when we loaded 100 nM of our positive control (GST-Q2ABc) D-CaM's

fluorescence increased due to the protein-protein interaction (Figure 2.8.C). This happened because when Q2ABc is present it binds to D-CaM and as a consequence D-CaM suffers a conformational change which exposes dansyl residues to the media. When Ca^{2+} was loaded, the increment of fluorescence was bigger, and it underwent a shift to shorter wavelengths as before. In fact, Dansyl group is a chromophore which increases the fluorescence in hydrophobic environments.

Regarding MBP-Q2ABc and Trx-Q2ABc, the assay revealed that the proteins were not be properly folded (Figure 2.8.D and E). When we added 100 nM of the protein, we could observe a possible interaction because a fluorescence enlargement was visible. The same happened when we loaded 9,7 mM of $CaCl_2$. However, this escalation was not proportional in the case of GST-Q2ABc which demonstrated that either our proteins were not properly folded or that MBP or Trx might disturb the interaction. Moreover, some kind of dispersion was also observed at lower nm. After making DLS measurements of the protein samples, we were able deduce that this dispersion was produced by the presence of big proteins or soluble aggregates.

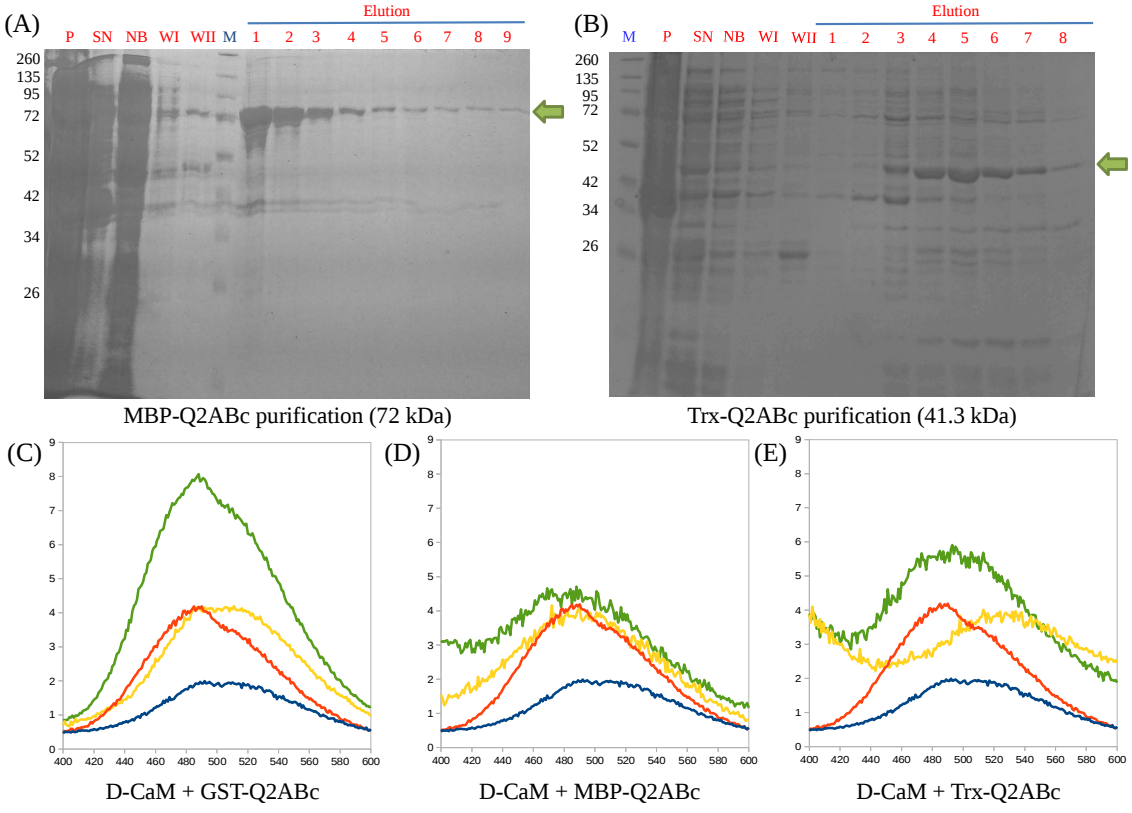


Figure 2.8. Protein purification and fluorometric assay. (A)(B) MBP-Q2ABc and Trx-Q2ABc protein purification in Ni-Sepharose step. M, Marker; P, pellet; SN, supernatant; NB, not bound to the column; W1, First was with 40 mM imidazole; W2, second wash with 60 mM imidazole; and Elution were done increasing imidazole concentration up to 500 mM. (C)(D)(E) D-CaM tritation with GST-Q2ABc, MBP-Q2ABc and Trx-Q2ABc in absence and presence of calcium. In *blue* D-CaM (12.5 nM) alone in absence of (10 mM EGTA); in *red*, D-CaM (12.5 nM) with Calcium (10 mM EGTA + 9.7 mM $CaCl_2$); in *yellow*, D-CaM (12.5 nM) with 100 nM of fusion protein in absence of Ca^{2+}(10 mM EGTA); and in *green*, D-CaM (12.5 nM) with 100 nM of fusion protein in presence of Ca^{2+} (10 mM EGTA + 9.7 mM $CaCl_2$). The response of MBP-Q2ABc and Trx-Q2ABc is different when we compare to our positive control GST-Q2ABc

After this fiasco we decided to go back and focus on **CaM fused constructs**. As shown in supplemental data (Appendix P1) calmodulin resulted to be a very good solubiliser when the linker between helix A and B was removed (Figure 2.9). By Native-PAGE gels protein oligomerization state was analysed, indicating that proteins were correctly folded. (Chapter 3).

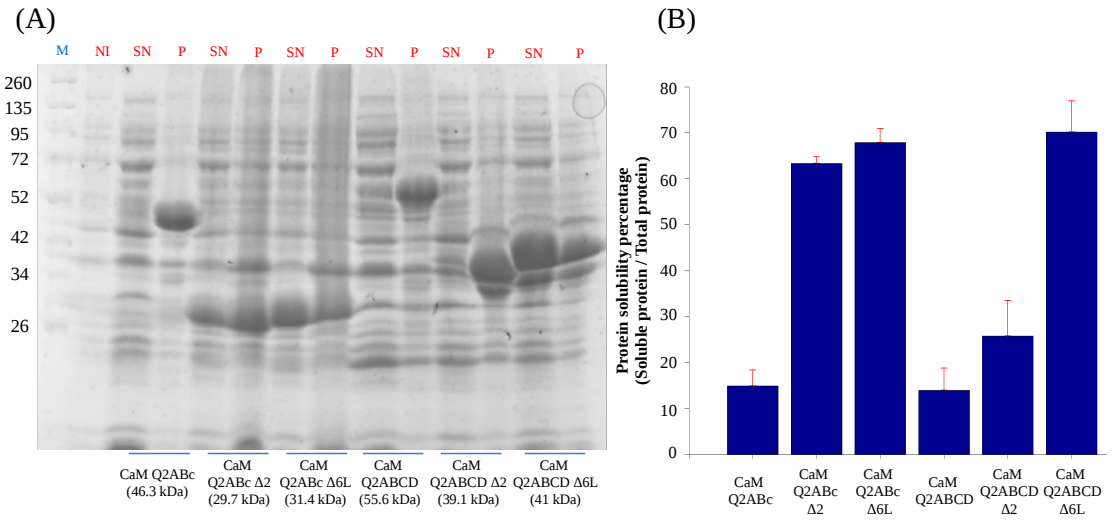


Figure 2.9. Example of CaM-fused proteins. (A) SDS-PAGE gels are shown where M, Marker; NI, not induced; SN, supernatant; and P, Pellet. Protein construction and size is defined on the bottom. (B) Protein solubility percentage of each construct. Deletions between helices A and B yield more soluble protein. Even if they do not appear here, more CaM fused proteins were analysed (Appendix P1).

First, we wondered if CaM could share some of its most important characteristics such as the thermostability when it was fused to different constructs to the carboxy termini (Figure 2.10). We took Q2ABc Δ6L and Δ2 and Q2ABCD Δ6L and Δ2 constructs. After expressing these proteins we separated the soluble and insoluble fractions. We heated the crude supernatant at 95 ºC for 5 min, taking every minute a sample. Eventually samples were centrifuged again to get rid of aggregates. By this way we were able to confirm that the thermostability property was shared in most of the cases except for CaM-Q2ABCD Δ2, which could be too large and not properly bound to the CaM.

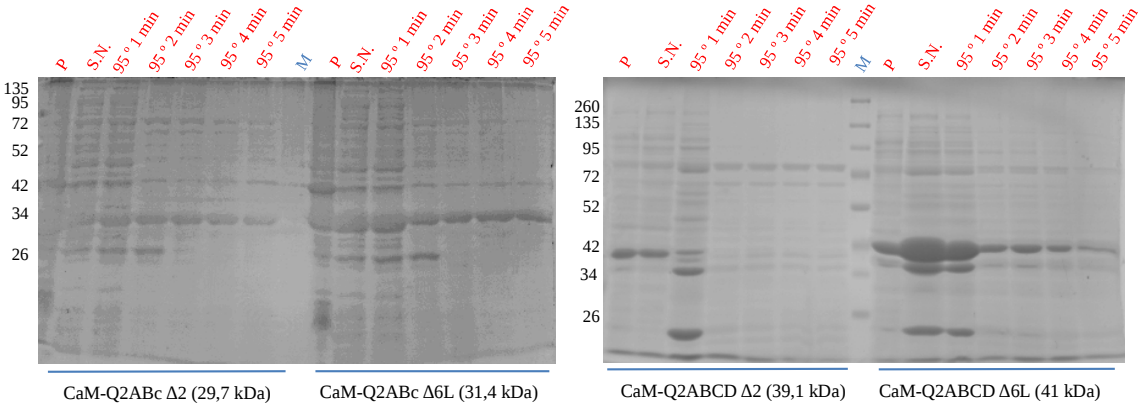


Figure 2.10. Thermostability assay of CaM-fused proteins. SDS-PAGE gels are shown where supernatant fraction of these constructs were heated at 95 ºC for different times. These samples were then centrifuged and loaded. All the proteins except CaM-Q2ABCD Δ2 faced the heat shock thanks probably to calmodulin thermostability. However, larger proteins with helices C and D did not behaved as well as ABc version.

We also checked whether the Poly-His tag location could influence or not protein expression or solubility. In order to do so we decided to take Q2ABc Δ6L and Δ2 and we cloned them using two different vectors, pET28a which adds a Poly-His at the C-terminus and pProEX-HTc which adds a Poly-His at the N-terminus. After protein expression and lysis of bacteria, we separated the supernatant from the pellet fraction. SDS-PAGE gels were loaded and protein was quantified using ImageJ. Surprisingly, we saw a noticeable decrease when Poly-His was located in the C-terminus (Figure 2.11) reinforcing the idea that this kind of test would be done more often.

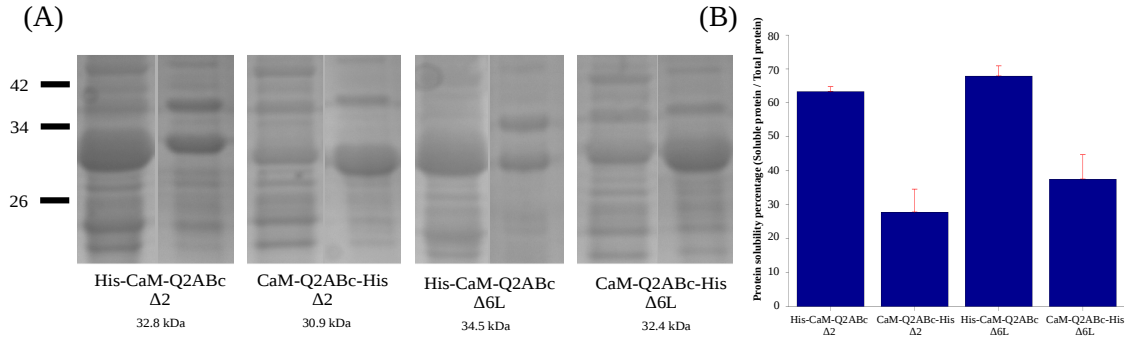


Figure 2.11. His-tag location effect in protein solubility. (A) CaM Q2ABc Δ6L and Δ2 constructs were expressed placing the His-tag in the N- and C-terminus. (B) Solubility chart. When His-tag was located in C-terminus the solubility was lower.

In order to make sure that the interaction of Calmodulin with the C-terminus of K_v7 channels was real, we decided to introduce some point mutations that could affect CaM binding to the target protein. Some well-known mutations that affect binding are I340E and S511D (Alaimo, Alberdi, Gomis-Perez, Fernańdez-Orth, et al., 2014). When protein solubility was tested the results revealed the importance of CaM binding to the C-terminus in order to get soluble protein (Figure 12). In fact, these two mutations do not produce soluble protein even if we combine them with a deletion between helices A and B. As a positive control, we also tested S342D mutation which *in vitro* give rise to a complex formation. As expected, in this case we obtain soluble protein.

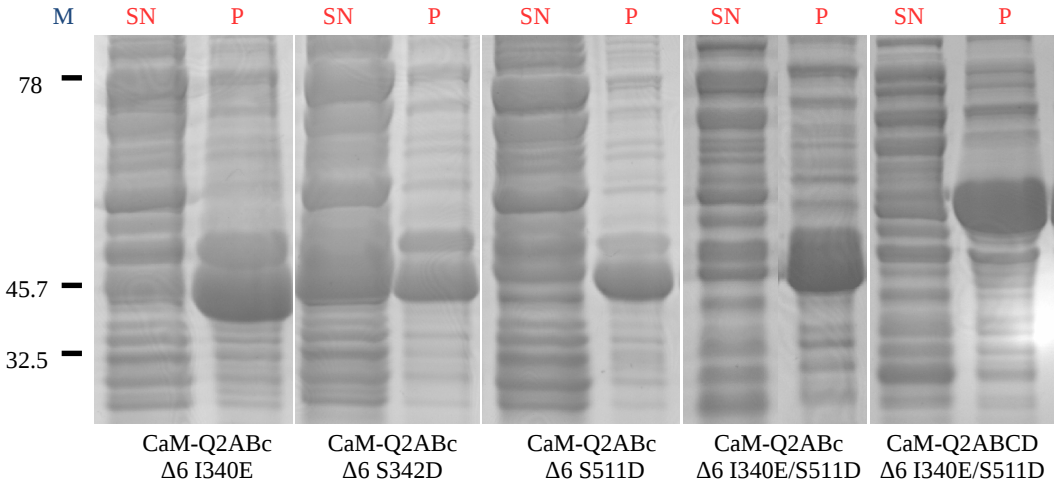


Figure 2.12. Screening of protein solubility introducing point mutations. Three point mutations were analysed. Among them two are well-known to jeopardise CaM binding to the C-terminus (I340E and S511D). S342D is known to be neutral. For the case of the double-mutant I340E/S511D either ABc and ABCD constructs with deletion 6 were tested.

2.3.6 None Of The Selected K_v7 Channel Orthologous Shows Improved Solubility Profile

As mentioned before protein complexity and function is the result of evolution. Based on this, we have tried to look back on the evolutionary tree of K_v7 channels. In fact, we have picked very simple organisms that express voltage gated potassium channels that are considered to be orthologous K_v7 channels, trying to find out a better candidate for protein purification. At the beginning we were not sure how to test the intrinsic solubility properties, but finally we decided to check it with CaM as a fusion partner (Figure 2.13).

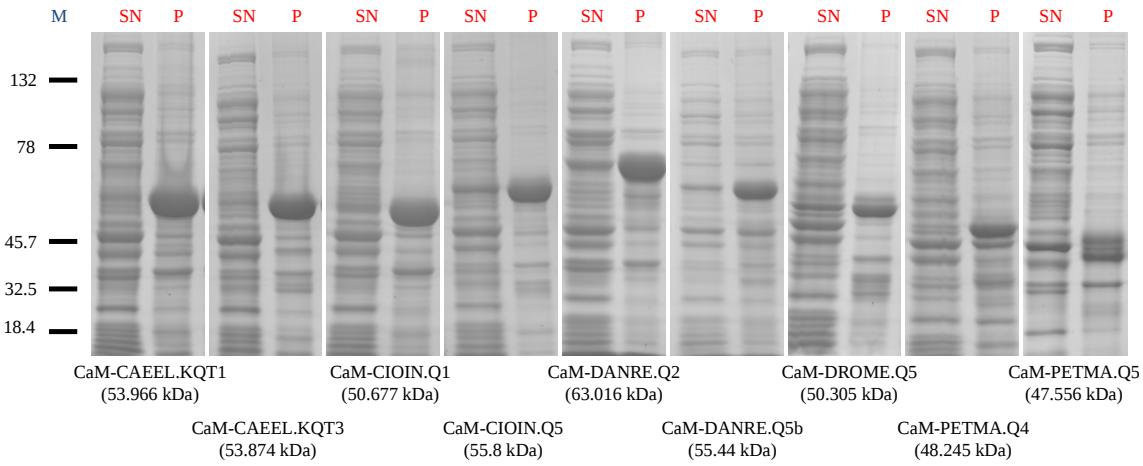


Figure 2.13. Solubility screening of K_v7 channels orthologous fused to CaM. SDS-PAGE gel where SN and P were loaded from each construct. CAEEL, *Caenorhabditis elegans;* CIOIN, *Ciona intestinalis;* DANRE, *Danio rerio;* DROME, *Drosophila melanogaster;* and PETMA, *Petromyzon marinus*. None of the studied orthologous showed a better solubility. Q refers to KCNQ, which is the gene encoding for K_v7 chanels

2.4 DISCUSSION

One of the biggest limiting factors of working with proteins *in vitro* is the enormous demand of purified protein. Thereby in science there have been two phenomena: on one hand, scientific tools developers have created more sensitive tools that requires lower amounts of sample to run experiments. On the other hand, scientist have also improved the protein production and extraction by improving and developing several strategies to obtain higher amounts of protein. But, in general, the more protein we get the more experiments we can run.

The first step of improving protein expression is gaining more information about the protein we are working with: Which is the molecular weight of the protein? Do I know the amino acid sequence? What is the isoelectric point of the protein? Is it a membrane protein? Does it interact with any other protein or cofactor? How many helices it may have? Wait... Perhaps someone has faced this problems before.

Indeed, the best way of acquiring the knowledge of a specific protein is the literature reading. Usually we can find more people who have been dealing with the same protein or with the same problem and for sure that this is the best and the most reliable way of learning. Nevertheless, we have seen that there are a lot of *in silico* tools that are specialised in predicting different facts of a target protein based on amino acid sequence comparison and alignment in a specific database. these programs give an estimation, a general idea of how a protein may be, but we shouldn't trust them completely without checking what they say. In our case, most of the programs behaved quite well calculating the secondary structure and disordered regions but solubility predictors failed.

Regarding which strategies should be followed in order to improve protein solubility, we have to admit that is a difficult task to address. All proteins behave in a different manner and what it is beneficial for one could be harmful. However, apparently there are some rules of thumb that often work. In case of insoluble proteins, first and easiest way of improving protein solubility would be either lowering the expression temperature or lowering the inductor concentration. Doing this bacterial protein synthesis slows down giving more time to the recombinant protein to fold appropriately. Nonetheless, for the C-terminus of K_v7 channel do not work perhaps due to the high hydrophobic residues found in the sequence and also due to the high amount of disordered regions which may interact with the membrane.

In this we have obtained soluble expression of recombinant proteins using the most recent improvement strategies in recombinant expression of proteins in *E. coli.*

Besides changing expression condition we have also done some changes at host level that have been beneficial for solubility improvement. After testing six different *E. coli* strains, we

conclude that for our recombinant proteins the most suitable strain has been demonstrated to be BL21(DE3)CodonPlus. In this case it corresponds to a humanized strains that has proven to express more bulk protein when we compare it with the most basic competent *BL21 E. coli* strain. Furthermore, working with this strain also has another advantage.

Between the strategies for sparing "rare" codons, our data suggest that humanized strains are more advisable alternative to codon-usage optimization. The same goal can be achieved by both of them, but the DNA sequence-optimization by site-directed mutagenesis is complicated and time-consuming, and the *de novo* synthesis is rather expensive yet. Nevertheless, the money investment to acquire humanized strains could be a little bit expensive, but this tool could be used for more than one construct. Moreover, these competent bacteria could be stored at -80 ºC with glycerol, and whenever they were necessary they could be regrowth unlimitedly.

Target engineering has demonstrated to be really useful too. Fusion protein technology has improved the solubility and the purification of plenty of all-kind of proteins. In this chapter we have conduct a comparison of six commonly used fusion proteins with regard to their relative effect on solubility of the proteins of the K_v7 channels C-termini. *A priori*, MBP and Trx tags were the most solubility enhancers. However, although we may show very promising results for the solubility of our recombinant proteins, it should be underlined that it does not *per se* mean that the proteins are properly folded or soluble when the fusion partner is cleaved off. In fact, in case or MBP it has been described that this tag may create "soluble aggregates" (Zanier et al., 2007), which may also happen to Trx and other type of tags.

On the other hand, CaM (which is the natural interacting partner of K_v7 channels), has amusingly been able to solubilise K_v7 constructs, specially these who has the linker between helix A and B removed. Moreover, some properties such as thermostability were tested indicating that CaM is able to protect the protein is wrapping against a heat shock. As a interaction partner calmodulin is able to bind the C-terminus of K_v7 channels. In the case where CaM is not been co-expressed the hydrophobic residues of the C-terminus in which CaM interacts are exposed and thereby unspecific interactions begin to occur leading to a protein aggregation. Nonetheless, when CaM is been co-expressed, CaM binds to these hydrophobic residues hiding them of the solvent. The fact that the protein solubility improves when the linker between helices A and B is removed only strengths the idea that protein aggregation occurs due to a non-specific hydrophobic interactions. Furthermore, when point mutations which are known to jeopardise CaM binding to the C-terminus (either helix A or B) are introduced either in helix A, helix B or in both, it has been shown that protein solubility was smaller, which highlights the idea that CaM is wrapping helices A and B and that this binding is necessary for the protein solubilisation.

To finish with, we tested protein solubility of K_v7 channel orthologous from 5 different species co-expressed with CaM without any success. In some case this strategy could give rise to a huge improvement in protein obtaining such as the case of $K_v7.4$ C-terminus (Xu et al., 2013), but in our case did not work, probably because the linker between helices A and B was present in all the cases.

2.5 APPENDIX

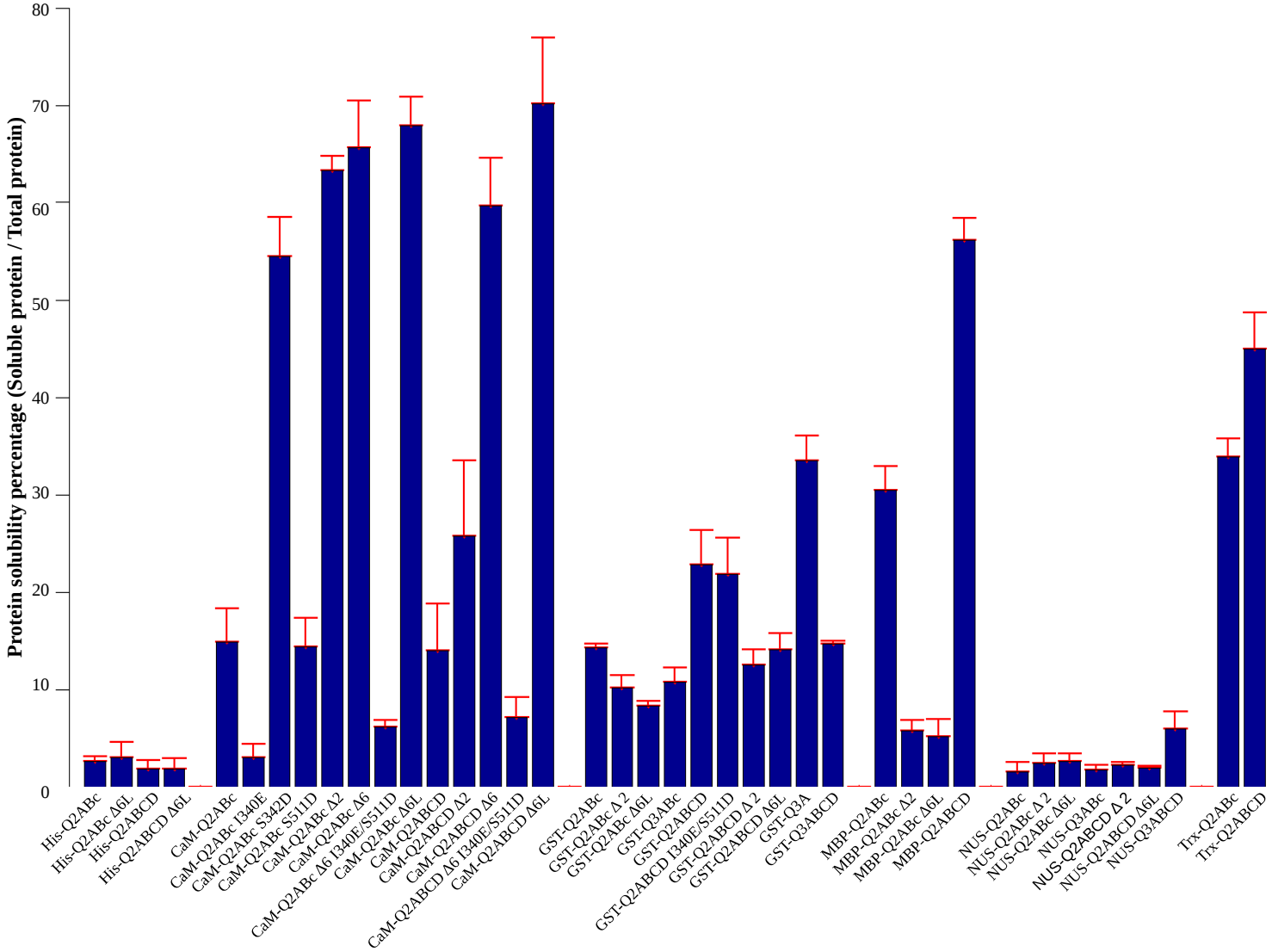


Figure 2.14 $K_v7.2$ and $K_v7.3$ C-terminus solubility using different fusion proteins. Proteins were expressed at 37 ºC for 3 h with 0.5 mM of IPTG. These data is based on all the SDS-PAGE gels run in order to determine protein solubility. Six tags have been used: His, Poly Histidines; CaM, Calmodulin; GST, Glutathione S-transferase; MBP, Maltose-Binding Protein; NUS, N-Utilization Substance; and Trx, Thioredoxin.

3 STRUCTURAL CHARACTERIZATION OF CAM/K_V7.2_AB COMPLEX

3.1 INTRODUCTION

3.1.1 ATOMIC PROTEIN MODELS DEFINITION TECHNIQUES

Protein structure determination has been crucial in the biomedical development in the last decades. Thanks to the structural information at molecular or atomic level different improvements have been achieved in the understanding of molecular mechanisms governing different biological processes. Moreover this valuable information has provided new clues on the comprehending the origin of different diseases and the "intelligent-drug-design" by pharmaceuticals to avoid them. This advancement has also benefit industry where protein engineering has leaded to better efficiency of, for example, enzymatic catalysis.

Nowadays three main techniques are used to determine high-resolution structural models.

3.1.1.1 X-Ray Crystallography

According to Protein Data Bank almost 90 % of the structures deposited have been resolved using this technique. By X-ray crystallography structures have been released where it has been enabled de distinction of two points in space as close as 1 Å resolution, or even less.

In order to used this method, proteins have first to be purified (purity > 90-95 %) and then crystallized. The idea of the crystallization consist on taking a soluble protein to a its less solubility state. When a soluble protein reaches the degree of supersaturation nucleation and crystal growth occur. Crystals are formed by periodically ordered molecules that are organized by translational repetitions in the three dimensions. Crystallization step often is the most time consuming one. There is not a comprehensive theory in this task and will absolutely depend on the protein. There are several commercial crystallization screenings available and different types of robots that make simpler this process.

Once the crystal is formed, then it is subjected an intense beam of X-rays (Figure 3.1.A). If the crystallization process is good enough X-rays will diffract due to the interaction of the electrons in the atoms from the protein and the X-ray beam. The diffraction is collected to determine the distribution of electrons in the protein. This gives rise to a electron density map which is used to determine the position of each atom in the space. All the coordinates are into a single structural model which fits all the collected data. The correlation between the atomic model and the experimental data is calculated as R-value.

3.1.1.2 NMR Spectroscopy

Nuclear magnetic resonance can provide high resolution models too, but it uses a very different strategy. Nevertheless this technique has some limitations so as to get a structural models, specially these which are bigger than 10 kDa as we will see later. This could explain why only the 9 % of the structures in PDB are provided using this method.

In order to used this technique, proteins have to be purified as before (purity > 90-95 %) and then they are placed in a strong magnetic field where are irradiated with radio waves. The strong high-frequency magnetic field stimulates the atomic nuclei of the isotopes that have a magnetic spin, and when the irradiation finishes it measures the frequency of the magnetic field of the stimulated atomic nuclei during the oscillation period back to the initial state (Figure 3.1.B).

After analysing a set of distinctive set of observed resonances spectroscopist have a list of restrictions where the distances between same or different atomic nuclei is represented. These restrains are used then to build up a structural model which shows the location of each atom in the space. The closest parameter to measure the accuracy of the model is express by the root-mean-square deviation of the atomic coordinates.

3.1.1.3 Cryo-Electron Microscopy

This method is gaining popularity in structural biology in the last years. Protein samples are placed on an electron microscopy grid and later are rapidly frozen by liquid nitrogen or ethane. In this case the method is based on a beam of electrons that passes through the sample change the rotational angle of the incising beam. The information is collected as a 2D images and by different softwares the reconstructions of 3D models can be achieved.

There are some advantages of using this technique. For example, the amount of protein required for this technique is very low compared to X-ray crystallography and NMR. Moreover, proteins were in solution just before freezing them which means that can be assume that they were in their wild-type conformation, not like in crystallography.

One limitation of the cryo-electron microscopy is that proteins are required to be identical so as to perform the image averaging. Therefore only homogeneous proteins are accepted. However, the utilization of the technique is increasing as more three dimensional structures of small proteins or big complexes have been released recently at good atomic resolutions.

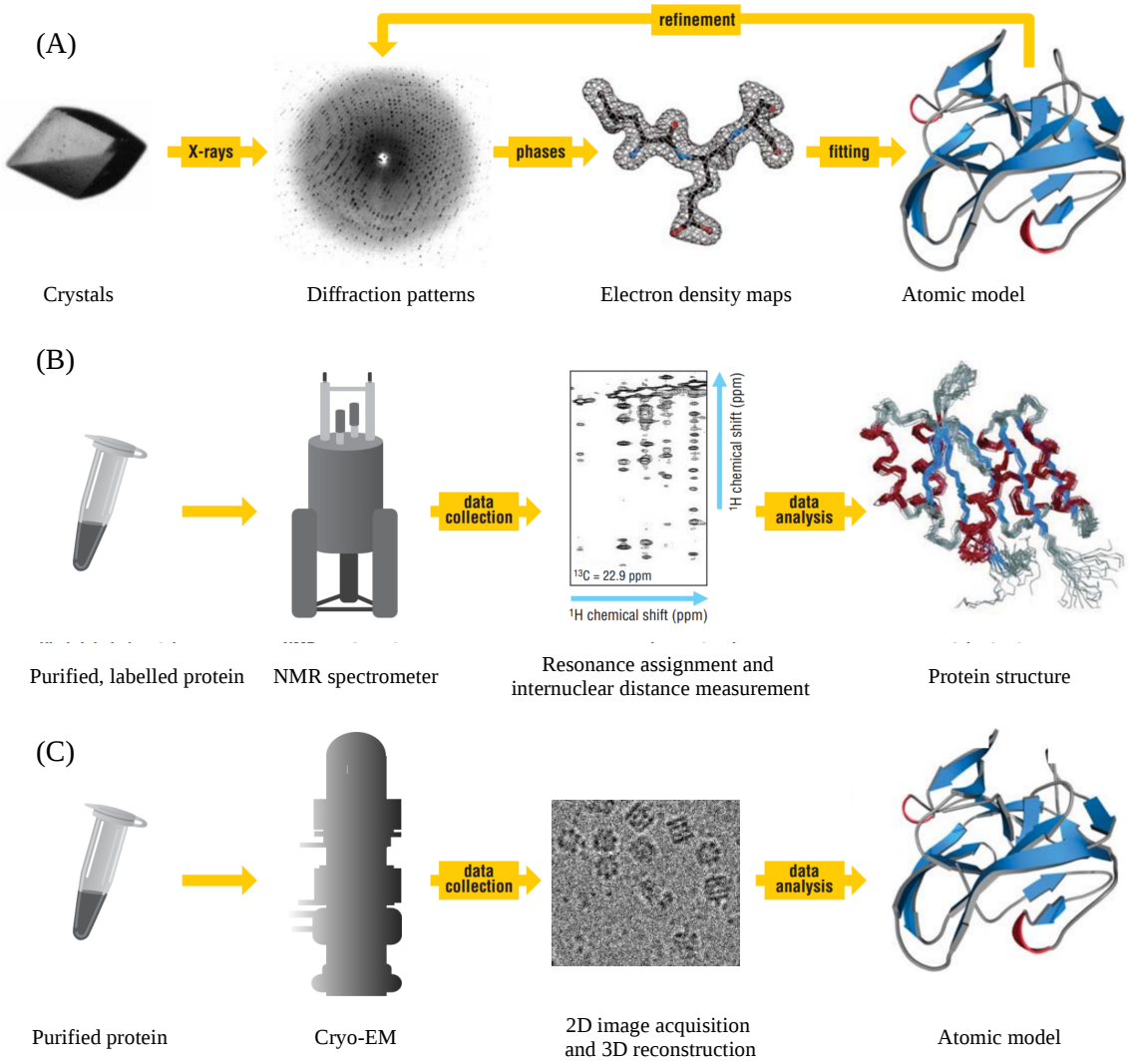


Figure 3.1. Schematic view of different structure determination techniques. (A) Structure determination by X-ray crystallography. After protein crystallization a source of X-rays bombards the protein causing a scattering which is collected. From the diffraction patter a electron density maps are computed. Finally a atomic model is adjusted to the data. (B) Structure determination by NMR. Often labelled protein is concentrated and placed in the magnetic field. Radio frequency pulses are applied perturbing atoms nuclei. Relaxation to the original state emits radio frequency radiation which will depend on the atomic environment. Radiation is recorded and process in comparison to a reference signal to give a chemical shift measurement. Depending on the list of restrictions obtained in this process a protein structure model is built. (C) Structure determination using Cryo-EM. Protein sample is diluted to a determine concentration around 20-50 μM. Afterwards the sample is cryogenated rapidly with N_2 or ethane and placed inside the microscope. X-ray beam radiates the sample and images are collected from different angles. Finally a software takes together all the pictures and reconstitutes a 3D structure of the protein. Adapted from (Petsko & Ringe, 2004)

3.1.2 PROTEIN STRUCTURE DETERMINATION BY NMR

NMR is a well-established method in many different fields such as physics, medicine, biology and chemistry. Such is the importance of the applications and development of this method that in recognition, several Nobel Prizes have been awarded in the last 60 years: Otto Stern (1943), Isidor Rabi (1944), Felix Bloch and Edward Purcell (1952), Richard Ernst (1991), Kurt Wüthrich (2002) and Paul Lauterbur and Peter Mansfield (2003).

For this thesis, Nuclear Magnetic Resonance has been chosen as the main protein structure determination technique as it is a very robust tool for studying structure, function and dynamics of biological macromolecules in solution. This small part of the introduction will be focused on providing the reading the basic knowledge of NMR bases so as to understand NMR spectroscopy under a non-spectroscopist point of view. For more detailed description of NMR processes and phenomena we recommend the reading of some interesting books such as *Protein NMR Spectroscopy: Practical Techniques and Applications* (Roberts, 2011) or *Protein NMR Spectroscopy* (Keeler, 2010)

3.1.2.1 NMR Theoretical Bases

NMR is a specific spectroscopy technique which detects energy states transitions (**FID**) of certain nuclear **spin** states after applying a **electromagnetic pulse** or pulses in the presence of a **magnetic field**. Perhaps this is too much information in just a sentence, so lets disassemble it.

The magnetic resonance phenomenon occurs as a result of the nuclear ***spin*** (I), which is a quantum mechanical property of atomic nuclei. Nuclei have positive charges which are spinning, and as any rotating charged particles, they generate a magnetic field or nuclear spin. For a single nucleus, its spin determines the magnitude and orientation of magnetic angular momentum (μ). When we introduce a nuclei with a spin quantum number different from zero into a bigger magnetic field, it adopts one of the possible angular momentums which is determined by its magnetic quantum number (*m)*. Magnetic dipole is then generated in all these nuclei with an odd mass number such as ^{1}H, ^{13}C, ^{15}N and ^{17}O (I=1/2) because they have an unpaired proton. In these cases nuclei will only split in two energetic states in the presence of de magnetic field (Zeeman effect), one defined as m = +1/2 (lower energy state) and m = -1/2 (higher energy state). As a result only one type of transition is allowed.

$$\mu = \gamma \cdot I \cdot \hbar$$

Equation 3.1. Magnetic angular momentum (μ), constant gyromagnetic ratio (γ), spin number (I) and ħ = h/2Π, where h is the Planck's Constant ($6.626 \cdot 10^{-34} \cdot J \cdot s$)

It is recommendable to keep in mind that in nature the most abundant nuclei with an odd mass number is hydrogen. Unfortunately, other biological nuclei such as ^{12}C ^{14}N or ^{16}O have I=0, so do not give NMR spectra. Thus, nuclei isotopic enrichment is necessary to study study big proteins by NMR.

Previously described energy levels are quantized and the difference between both states will be proportional to the magnetic field (B_0) and the constant gyromagnetic ratio (γ) as defined in Equation 3.2.

$$\Delta E = h \cdot \nu = \hbar \cdot \gamma \cdot B_0$$

Equation 3.2. Nuclear energetic transitions. The energetic transitions are proportional to Planck's constant ($h = 6.626 \cdot 10^{-34} \cdot J \cdot s$) and the Larmor frequency (ν) for each nucleus. Energetic transition could be define too as proportional to the gyromagnetic ratio constant of each nucleus (γ) under the influence of the magnetic field (B_0) and $\hbar = h/2\Pi$

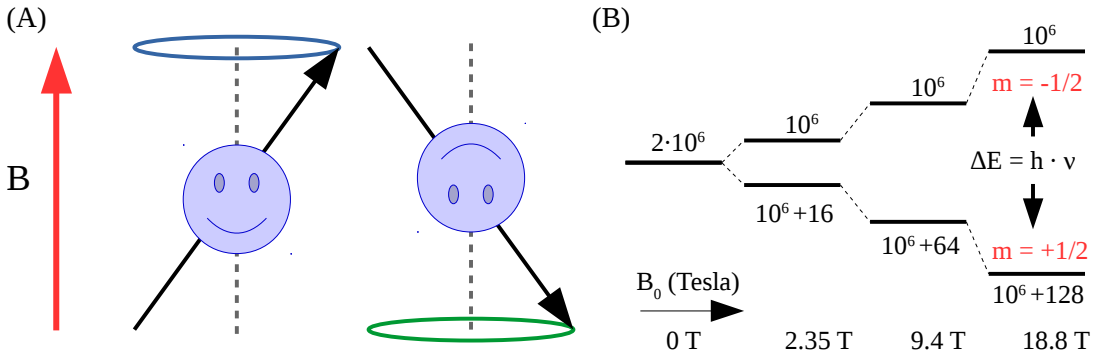


Figure 3.2. **Spin precession and energetic diagram.** (A) In the case of I=1/2 protons, the z component of the magnetic moment of the nucleus is aligned along the direction of the external magnetic field B. Two orientations are possible, the more stable one with lower energy (m = +1/2) and the less stable one with the highest energy (m = -1/2). (B) The energetic differences will be proportional to the magnetic field. Moreover, the energetic difference corresponds to a photon of a electromagnetic radiation at Larmor frequency. Assuming that the initial nuclei population is 2 millions the population differences between both energetic states variates depending on the magnetic field.

The larger the magnetic field the more expanded is the frequency range and the higher is the resolution. Even if several improvements have been done in order to develop more powerful **magnets**, the population differences between both energetic states is still very low (Figure 3.2). Thereby, NMR is a low-sensitivity technique that requires a high repetition series of the same experiment so as to increase the signal-to-noise ratio. Another way of improving the sensitivity is raising the concentration of nuclei or increasing the sample's volume.

Nuclear spin transitions in NMR are located in the range of 10-1000 MHz, therefore **electromagnetic pulses** are used to create transitions between these energetic level states. Radiofrequence pulses (RF) are electromagnetic radiation associated to an oscillating magnetic field. When a RF pulse with the appropriate energy is employed (when equals ΔE), the nuclear

spin system absorbs the radio frequency photon and resonates allowing energetic transition either from lower to higher or the opposite. Again, as mentioned in Equation 2, this pulse will strongly depend on the external magnetic filed (B_0) and the Larmor frequency (v).

As stated by quantum physics, we can not measure the spin rotation components in the x and y axis when we are measuring the z component. In fact, the magnetic moments oscillate around the z-axis depending on the resonance frequency. This phenomena is name as *spin-precession*. The origin of this net magnetization around the z-axis comes from the addition of all the small magnetic moments of the protein. When we apply a radiofrequency pulse as explained before, populations of the different energetic levels are altered and in consequence the orientation of the net magnetic moment changes too (Figure 3.3).

When the oscillation frequency of the spin-precession matches the Larmor frequency, net longitudinal magnetization along z-axis is transformed into a transversal magnetization on the *xy* plane, where the magnetization oscillates at the resonating frequency of the nuclei with an *offset* with respect to the Larmor frequency. When the radiofrequency pulse finishes sample magnetization remains in the *xy* plane and continues precessing about the static field until it recovers the spin system thermal equilibrium in the *z* plane generating radio signals that are called *free induction decay* (**FID**). However, the FID is a time-domain signal with contribution from different nuclei. In order to create a frequency-domain spectrum Fourier transformation of the signal-average FID is computed. This is what we call the NMR spectrum.

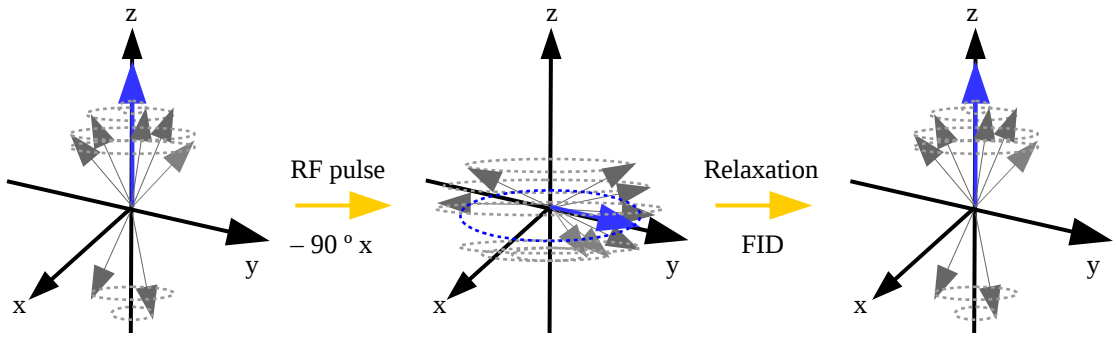


Figure 3.3. **Magnetic moment perturbation and relaxation.** Individual magnetic moment (grey arrows) oscillates in precession around z axis at Larmor frequency. Net magnetization (blue arrow) is aligned to the field. After a radiofrecuency pulse at Larmor frequency along x-axis shift z and y component of individual magnetic moments to the y axis, where it still precesses around the z-axis at Larmor frequency. After RF pulse, spin system recovers the thermal equilibrium in the z plane by emitting radiofrecuency signals (FID) which is recorded and processed later.

Resonance frequencies (v) are extremely sensitive to steric and electronic effects. Thereby, they depend either on the local magnetic field influenced by the external magnetic field (B_0) and by surrounding the observed nuclei (B_{local}) which in case of proteins will be related to their location in the three dimensional structure (secondary and tertiary structure) and also exposure to the solvent. Therefore we would expect that observed magnetic field ($B_{local} = B_0 + B_{induced}$) and the

characteristic resonating frequency (ν) is singular for each nuclei under study. In consequence, nuclei in different chemical shifts environments will resonate at different frequencies with are more popularly known as **chemical shifts (δ)**.

Chemical shifts depends on the magnetic field and they are expressed as MHz. However, in order to make easier a homogeneous comparison MHz are expressed as parts per million (ppm) (Equation 3.3).

$$Chemical\ shift, \delta(ppm) = \frac{\nu_0 - \nu_{REF}}{Spectrometer\ frequency(MHz)} \cdot 10^6$$

Equation 3.3. Chemical shifts in ppm. Obtained from the subtraction of the resonance frequency from that of a reference compound (tetrametylsilane (TMS) or similar) and normalized by the frequency of the spectrometer.

Protein NMR often multidimensional experiments are employed because, even if every nucleus experiences a distinct chemical environment and thus, chemical shifts, in large molecules where thousand of signals are present signal overlapping occurs. Multidimensional experiments correlate the chemical shifts of distinct nuclei adding more dimensions to the spectra and therefore decreasing the chance of overlaps. This type of experiments is possible due to a particular feature which differentiates NMR from other spectroscopy methods: excited states are relatively long lived (milliseconds to seconds). By this phenomena we we transfer the magnetization from one nucleus to other using pulses of electromagnetic energy and delays. Depending on the type of information we want to obtain, different type of magnetization transfers could be used:

- Scalar coupling or J-coupling: Magnetization is transferred through bond or electron cloud interactions between two nuclei with non-zero spin. The number of bonds by which the magnetization is transferred from 1 to 3. Neighbours magnetically active nuclei can be from the same type (homonuclear coupling) or from different types (heteronuclear coupling). Therefore this type of couplings are very useful in the backbone assignment of proteins.

- Dipolar coupling: This type of coupling allows magnetization transfer though space. Each nuclei can sense the presence of other nuclei that are located up to 5 – 6 Å. Protons are particularly sensitive to this type of magnetization transfers. Dipolar coupling is the basis for the Nuclear Overhauser Effect (NOE), which allows the detection of short to long range structural restrains which are used in the structure calculation by NMR.

3.1.2.2 Parameters Affecting Quality Of NMR Spectra

Following parameters have been already described previously but in this section we are going to empathise the main implications in the spectral acquisition.

Magnetic field has a compelling effect in the quality of the recorded data. Accordingly, when magnetic field is higher population differences between different energetic states increase too. Thereby either sensitivity and resolution are generally enhanced at higher magnetic field. Proteins stability do not depend on magnetic field.

Temperature changes can produce a variety of effects in the spectra. Higher temperatures lead to two effects: On one hand, temperature increment causes an acceleration in the protein tumbling (rotation correlation time)which in consequence chemical shift signals become narrower. On the other hand, protein may have less stability at higher temperature with a final result of protein aggregation. Besides, labile amide protons' exchange rate is increased which could lead to a signal lost.

Molecular weight is related to the molecular tumbling and to signal line width. When a protein is too big (> 30 kDa) molecular tumbling is reduced which causes line broadening. In this regard **molecular shape** also has a big influence in the line broadening of signals as globular proteins present less resistance to the media, and therefore to the molecular tumbling, than extended proteins.

Buffer composition and concentration can also affect spectral quality.

- High salt concentrations affects NMR acquisition itself. High salt concentration may lead to a reduction of the sensitivity of the 50 % as it becomes more difficult the tunning and matching and also result in bigger pulse duration with all the secondary effects concerning heat load, dephasing, etc. Reader should not be concerned about these concepts as they were not explained before. For further information it is recommended to check cited books.

- High pH also reduces signal sensitivity. 1H-15N exchange is base catalysed. Highest sensitivity to NH protons is around pH 4.5. For pHs from 4.5 to 7.5 is very recommendable but sensitivity drops as the pH goes up. For pH above 7.5 data collection is still possible but more protein will be needed.

- Protein stability: There are no guidelines to know *a priori* which type of buffer is more convenient for a certain protein. Protein samples are required to be into the magnet for long period where they are constantly suffering radiofrequency pulses which in a consequence heat them. Magnets are equipped with temperature regulators but protein stress leads to protein aggregation. Buffer screening is recommended so as to improve protein stability.

3.1.2.3 ^{15}N-HSQC Spectrum and Interpretation

Heteronuclear Single-Quantum Coherence (HSQC) spectrum is by far the most used experiment in protein NMR. In fact, this NMR experiment is refereed as a quality control method to be passed before further structural analysis of proteins. This spectrum shows a signal for each covalently bonded ^{1}H-^{15}N group (J-coupling). Therefore it provides a very valuable information in which a peak is represented for each backbone amide proton (except for Pro), a peak for each NH group in tryptophans and also a pair of peaks for the side chain amide groups of each Asn and Gln residues. Accordingly, this spectrum provides a excellent high-resolution "fingerprint" of the protein (Figure 3.4).

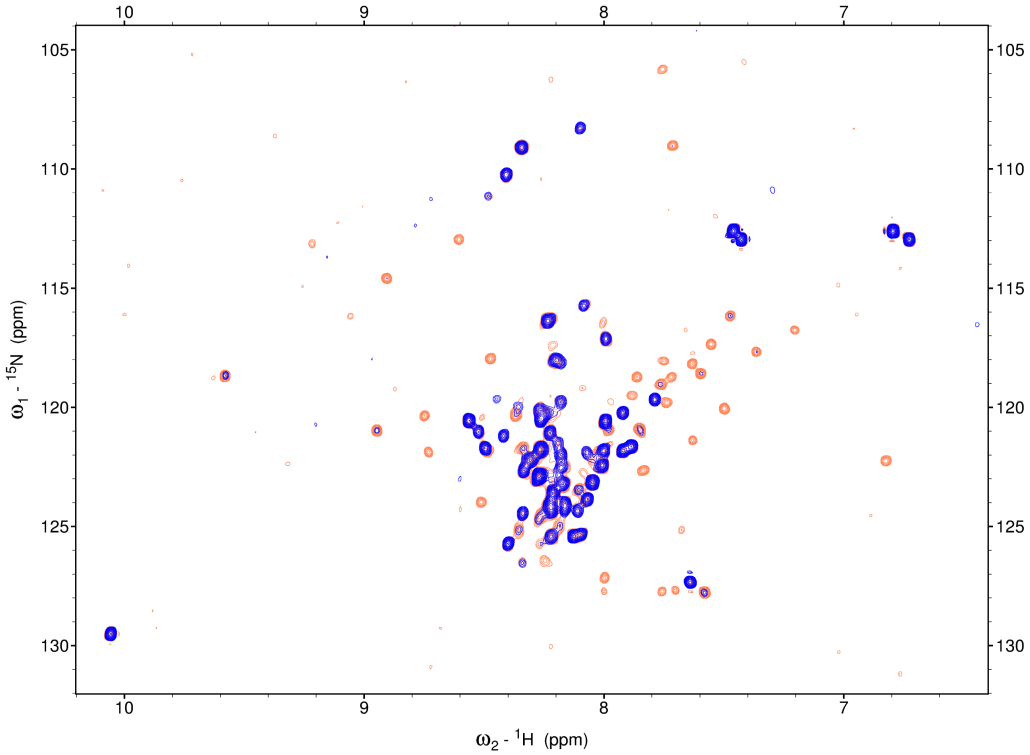


Figure 3.4. ^{15}N-HSQC spectra of COX17 protein. ^{1}H-15N correlation of all amide groups are shown. In blue the unfolded state and in orange the folded state of the protein in the presence and absence of DTT.

There are a plenty of regards that this simple experiment can address.

3.1.2.3.1 Protein folding state

Spectra from well-structured proteins which has a high secondary structure (α helix or β sheets) are more dispersed than unfolded protein. The reason is that amide groups are in similar chemical environments in the case of unfolded proteins, and in wider variety of chemical environment in structured regions (Figure 3.4). Thus, HSQCs can help in the identification of disordered regions which could be very useful for X-ray crystallographers so as to remove these parts that might prevent crystallization.

3.1.2.3.2 Protein aggregation state

Self-association will broaden almost all signals because molecular tumbling is slower in such cases. However, if we are suffering a conformational exchange rather than a protein aggregation only signals from the nuclei that are changing their chemical environment will change. Often some signal disappears and other appears. Nevertheless, in some cases is difficult to differentiate between them and it is a better option to use another technique to determine protein aggregation such as gel filtration or DLS.

3.1.2.3.3 Protein dynamic state

Comparing the theoretical number of signals with the observed ones may lead to mismatches. Subsequent analysis have to be done in order to now what happens to these missing signals. Nonetheless, it has been reported for several loops that they may undergo μs to ms conformational changes. Thus, when these kind of dynamics takes that long signal reduction is expected as this concrete region is changing its structure changing the amide protons' chemical environment every time.

3.1.2.3.4 Protein stability state

If we collect different HSQCs along a time period we can see whether or not protein's structure changes. When a protein departures from a structural state to a unfolded state dispersed signals decrease and more compacted signals in the middle of the spectra arise. For the case of degradation, this process is the same but signals often appear in the middle-bottom part of the spectra.

3.1.2.4 Chemical Shift Assignment

Each one of the signals detected by NMR should be assigned to its correspondent nucleus so as to be able to process correctly the structural information provided by them. Protein structures can be solved thanks to the isotopic labelling (typically ^{1}H, ^{15}N and ^{13}C) and multidimensional NMR experiments where magnetization can be transferred through different nuclei in protein.

Backbone assignment is done due to the sequential connectivity of the signals between the preceding and the subsequent amino acid. For this case Scalar coupling are necessary which allows the magnetization transfer along the backbone of the protein to the side chain atoms. Typically, 3D multidimensional heteronuclear experiments are HNCA / HNcoCA / HNcaCB / HNcocaCB / HNCO / HNcaCO / HNcaHA and HNcocaHA. This names represents how the magnetization transfer occurs where CA corresponds to alpha, CB to beta, Co to carbonyl carbons and HA to alpha proton.

HNCA, HNcaCB, HNCO and HNcaHA experiments give information about the CA, CB, CO and HA chemical shifts of the amino acid we are exciting (intra residue (i)) and CA, CB, CO and HA chemical shifts' preceding amino acids (inter-residue (i-1)). Usually these two signals are differentiated easily because intra-residue signals are more intense than inter residue signals as they are located farther. HNcoCA, HNcocaCB, HNcaCO and HNcocaHA however only provides information about preceding amino acids. By these two types of experiments intra- and inter-residue connectivity is established along all the backbone chain (Figure 3.5).

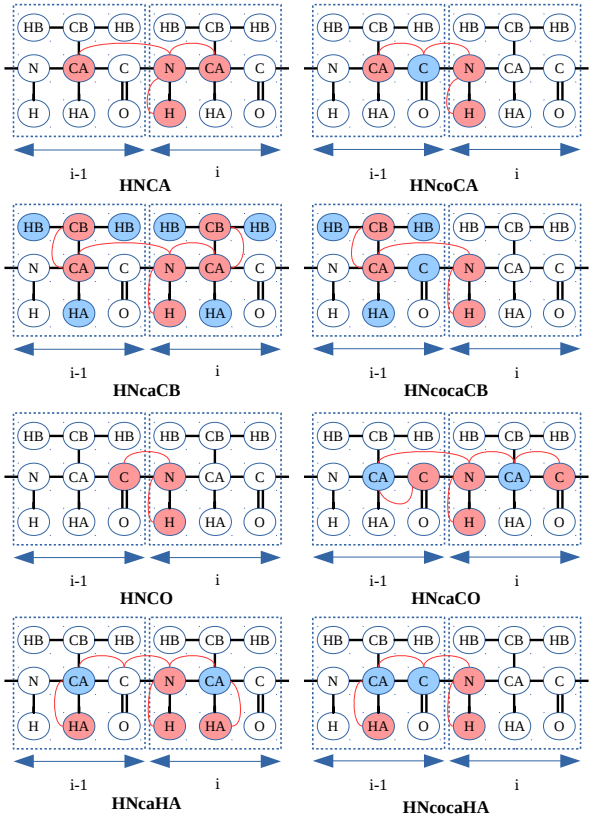


Figure 3.5. Magnetization transfers in three dimensional spectrum. Thin red lines corresponds to how the magnetization transfer occurs. In red they are highlighted the chemical shifts that appear in the spectra. In blue they are represented these nucleus that take part in the process but do not appear in the spectra.

Chemical shift of each amino acid can be compared with the average chemical shifts of each amino acid type (Figure 3.6). Among them Ala, Gly, Ser, Thr are the easiest to assign as they have particular chemical shift. However this is not enough to know the amino acid number of the protein that we are talking about. In order to do that we need to identify a the identity of a consecutive amino acids.

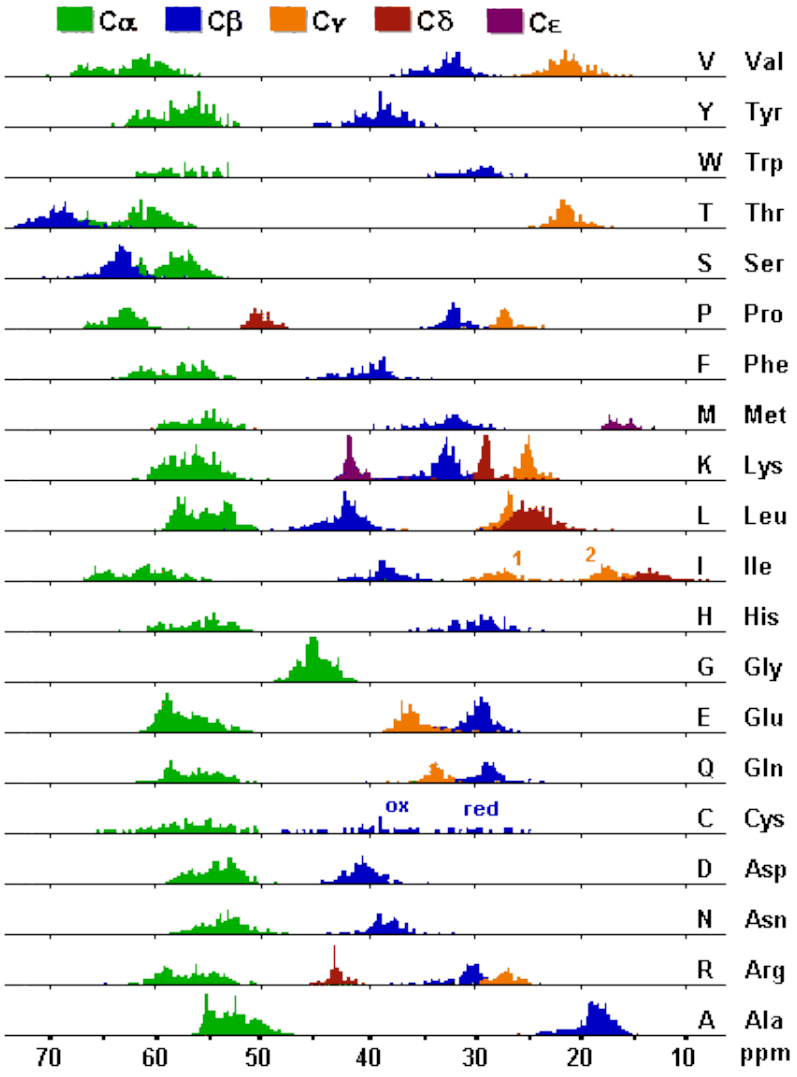


Figure 3.6. **Average values for each residue.** Statistical average values from BMRB database.

If we are interested in assigning the protons, carbons and nitrogens which are farther from the backbone we need to perform other type of experiments such as HNCcoNH, which shows all the carbon chemical shifts from i-1, and HccoNH, which connects all the proton chemical shifts from the i-1. These two are interesting also for side-chain's chemical shift identification.

Just with the backbone assignment huge valuable information can be obtained:

3.1.2.4.1 Secondary structure

Among other techniques to elucidate protein's secondary structure such as UV-Raman or Circular Dichroism, NMR based secondary structure characterization is residue-specific. For each type of amino-acid, chemical shifts could be interpreted in terms of secondary structure if we subtract our observed data to a reference values of same amino-acids in random coil structures. Moreover, this information gains reliability when nearest-neighbour effects are taken into account.

Besides, reference chemical shift values for every amino acid type in α-helices and β-sheets have been calculated (Wishart, 2011). These references are the bases of the Chemical Shift Index (CSI), a method which identifies and defines secondary structured regions. Moreover, CSI method has demonstrated to be easy to use, fast, and notably accurate. However, chemical shifts have to be first reference them properly otherwise the method fails.

There are a big number of programs so as to determine the secondary protein structure based on chemical shifts such as CSI (Berjanskii & Wishart, 2006), PSSI (Y. Wang & Jardetzky, 2002), PsiCSI (Hung & Samudrala, 2003), PLATON (Labudde et al.,, 2003), PECAN(Eghbalnia et al., 2005) and 2DCSi (Chuang et al., 2007)

3.1.2.4.2 Torsion angles

Torsion angles restraints are also very useful restrains in order to determine protein structure. During the structure refinement, torsion angles are used to define secondary structures within the protein. This information can be directly obtained using scalar coupling or cross-relaxation experiment but they can derived from the chemical shifts too in a very confident way as they are very sensitive to indicator of dihedral angles.

Backbone conformation is defined two dihedral angles angles for each amino acid (Φ and Ψ). Different programs are able to calculate Φ and Ψ such as TALOS+ (Shen et al., 2009), SHIFTOR (Neal et al., 2006), REDCAT (Schmidt et al., 2013) and PREDICTOR (Berjanskii et al., 2006). Often, these programs used their own databases where defined proteins with their corresponding heteronuclear shifts and dihedral angles are known. Then these programs try to find similar amino acid sequences with similar chemical shifts. Usually, they take into account the neighbouring amino acids as well as the PDB extracted amino acid references to build up an output file where the estimated torsion angles are defined for each amino acid. Nevertheless these programs also detect extreme torsion angles and they warm the user about their reliability. Side chain torsion angles can also be predicted by more specialized programs but they are less used.

3.1.2.4.3 *De novo* protein modelling from shifts

Lots of efforts has been done to compute protein structures only chemicals shifts information so as to evade the time consuming process of NOE assignment. Due to a rapid increment of protein's chemical shifts deposited in BMRB several bioinformatics have tried to developed algorithms to predict proteins structures. *Ab initio* structure predictors are based in the use of fragment-based assembly methods. Well known structures are fragmented in small segments in a database and when we submit our chemical shifts programs tries to find small fragments of the query that matches with similar fragments in the database. Actually this strategy it was first established in dihedral angles predictors. High scored fragments are fused together a initial structure is form. Afterwards this structure is readjusted and refined using heuristic energy functions.

Most popular programs are CS-ROSETTA (Vernon et al., 2013), CHESHIRE (Cavalli et al., 2007) and CS23D (Wishart et al., 2008). These programs for protein prediction based on chemical have improved a lot in the last period and they are a reality in terms that they can predict fairly protein structures. Nevertheless, several developments should be done to combine backbone chemical shift information with other type of informations such as residual dipolar couplings, J-couplings or NOEs.

3.1.2.5 Long-Range Contacts

Typically, through-space connectivity is detected and measured by two dimensional homonuclear NOESY experiments. For big proteins, routinely a third dimension is increased by the addition of either ^{15}N or ^{13}C in order to increase the resolution of the signals and to avoid the overlapping. This experiments include either the information of at short-range (sequential inter-intra- residues), medium range (intra-helix or intra-beta sheet) or long-range interactions. Thereby, the information gathered by this method is very completed where first we have to assigned local contacts before assigning long-range contacts.

The intensity of cross-peaks in NOESYs is proportional to the inter-proton distance (d) ($1/d^6$). As mentioned before these distances are below 5 – 6 Å. NOEs are the most useful restrains in order to determine protein structure as they reflect the distances through space between a specific nuclei with the rest of the protein. In fact, these contact are not limited only to the intra-molecular signals: in case of protein complexes they can be determine in the same way in these regions where protein-protein interactions are happening. It is highlighted that in this case more important the quantity of the contacts rather than the quality of them.

As mentioned above, NOESY spectra may include a huge quantity of signals derived with all these short- medium- and long-range contacts. In order to do the assignment easily, several softwares have been developed in which not only the automatic assignment is done but they also create the structural model based on all the provided restrictions.

NOEs are the only NMR- derived structure restrains that can be utilized alone and would still be able of producing reliable high resolution atomic models. Nevertheless, NOE experiment results are combined with other experiments in order to be more accurate in the structural model.

3.1.2.6 Residual Dipolar Couplings (RDCs)

Dipolar couplings depend on the orientation of the protein in solution towards the magnetic field (B_0). In isotropic solutions dipolar coupling are averaged to zero due to the Brownian motion of the nuclei.

If we take our protein into an anisotropic media which contains molecules that are oriented relative to B_0, a degree of the alignment is transferred to the protein recovering dipolar coupling. Alignment media must fulfil some criteria so has to be employed: It has to be aqueous, uniformly anisotropic over the whole sample, stable at different pH, ionic and temperature conditions and it shouldn't interact significantly with the protein (Figure 3.7). This steric orientation restriction has been accomplished using bicelles, pilamentous phages, polyacrilamide gels and also synthetic reagents such as DHPC (hexanoyl-phosphatidyl-choline), DMPC (dimyristoyl-phophatidyl-choline) or C12E5/hexanol (dodecyl-pentaethylene-oxide/hexanol) also referred in this thesis as Otting media.

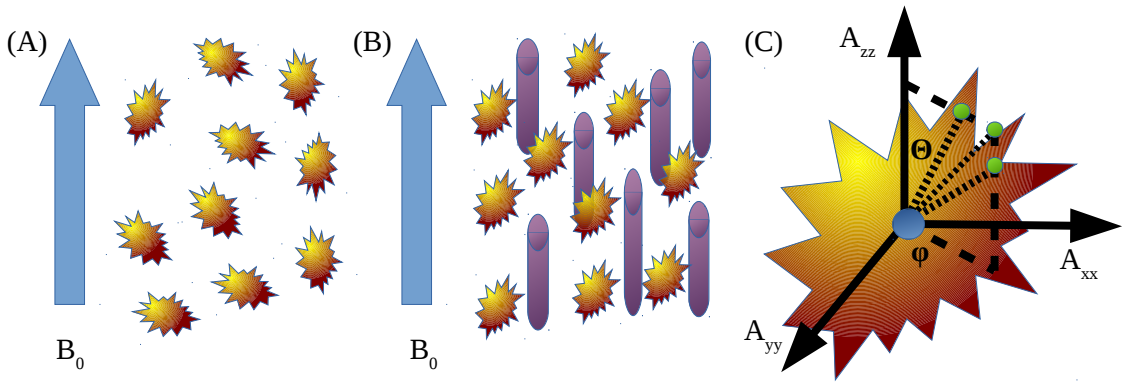


Figure 3.7. **RDCs graphical representation.** (A) In isotropic media protein tumbles freely. Dipolar couplings cancel each others because all the protein orientations towards the magnetic field have same probability. (B) In anisotropic media proteins acquired preferential orientation towards the magnetic field. The incomplete averaging of all protein's possible orientations leads to RDCs. (C) Protein preferential orientation can be described by an alignment tensor. For example, measured RDC between ^{1}H - ^{15}N depends on the the orientation of the internuclear vector with respect to the alignment tensor, which is described by polar coordinates θ and Φ.

As the scalar coupling, the dipolar coupling will be manifested as a splitting of the resonance lines. Imagining that we are interested in ^{1}H - ^{15}N chemical shifts, two measurements have to be done before extracting RDCs values: First we have to measure the splitting lines of the signals in a isotropic media and then measure the splitting signals in a anisotropic media. The difference between them is define as the residual dipolar coupling.

In order to use RDCs for *de novo* protein structure refinement two parameters have to be known: the axial component of the alignment tensor (Da), which is the defined as a traceless matrix (equivalent to Saupe matrix) which concretes the preferential orientation of the protein in the anisotropic media towards the magnetic field, and the Rhombicity (R) of the tensor which is

another parameter of the that defines the alignment tensor. The way in which these values are measured without a protein structure is by plotting a histogram of the observed couplings. The magnitude of the alignment (Da) and R is related to the extrema (A_{zz} and A_{yy}) and mode (A_{xx}) of the histogram using equations (Figure 3.8). Note that at least five independent dipolar couplings are needed so as to determine the alignment tensor, where independent means coming from non-parallel vectors.

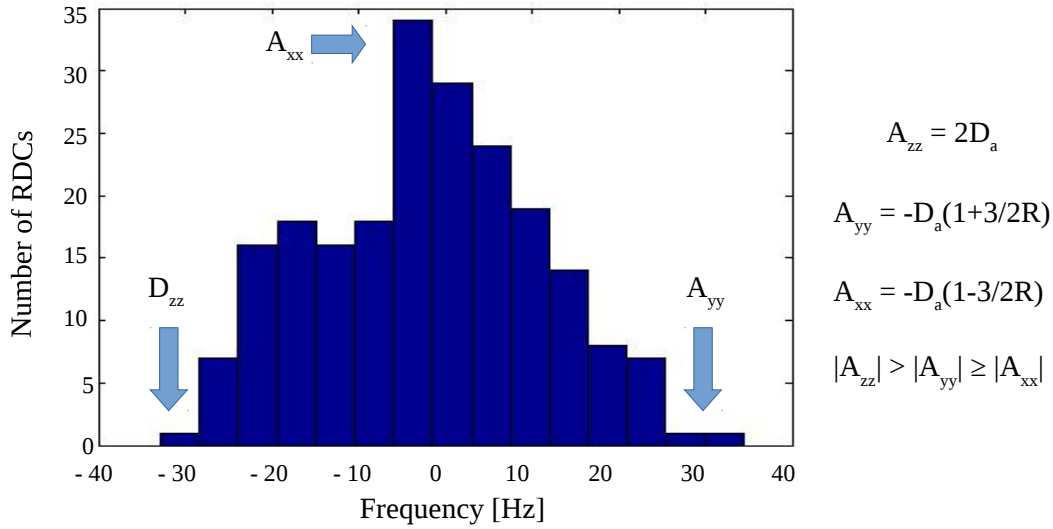


Figure 3.8. ***De novo*** **stimation of the axial component and the rhombicity value of the alignment tensor.** After plotting all the dipolar couplings in a histogram Da and R values can be calculated using the three equations in the right side of the figure. related to the extrema (A_{zz} and A_{yy}) and mode (A_{xx}).

When the protein structure is known is much easier to measure either Da or R fitting RDC values to a PDB file. PALES (Zweckstetter, 2008), MODULE (Dosset et al., 2001) and REDCAT (Schmidt et al, 2013) are most used programs to predict alignment tensors.

What kind of information provide RDCs? In case where protein's secondary structure is mainly alpha helices, all N-H dipoles found within the same helix will point in the same direction as a vector. The relative orientation of different helices or vectors can be determine when the alignment tensor, the RDC value and the distance between these two nuclei (N-H) is known. The orientation of distal domains can be measured in the same way too. However, even if RDCs determines the relative orientation of the elements, distances can not be determined. Therefore this information is usually combined with RDCs in order to improve protein structure.

3.1.2.7 3D Structure Calculation

After a really rough time assigning chemical shift and gathering information it is time to create a structural model with all of it (Figure 3.9).

Basically only two types of information are required from the experimental data in order to determine a protein structure: backbone chemical shifts and NOE information. With only these restrains we would be able to create a 3D model with a nice precision. However, incorrect chemical shift assignments or NOE assignments can lead to a wrong structure determination. In this regard, it is worth to mention that it is the quantity rather than the precision of these restrains what really matters for NMR structure determination. Thereby, not only previously mentioned two data sets are used: Dihedral angles restrains, hydrogen bond restrains and RDC restrains are often added so as to over-determine the structure. As a rule of thumb the more restrains we have the more accurate protein we will determine. But again, over-retraining distances or angles may lead to errors. Therefore we have to find a equilibrium between the amount of restrains we use and the results we obtain.

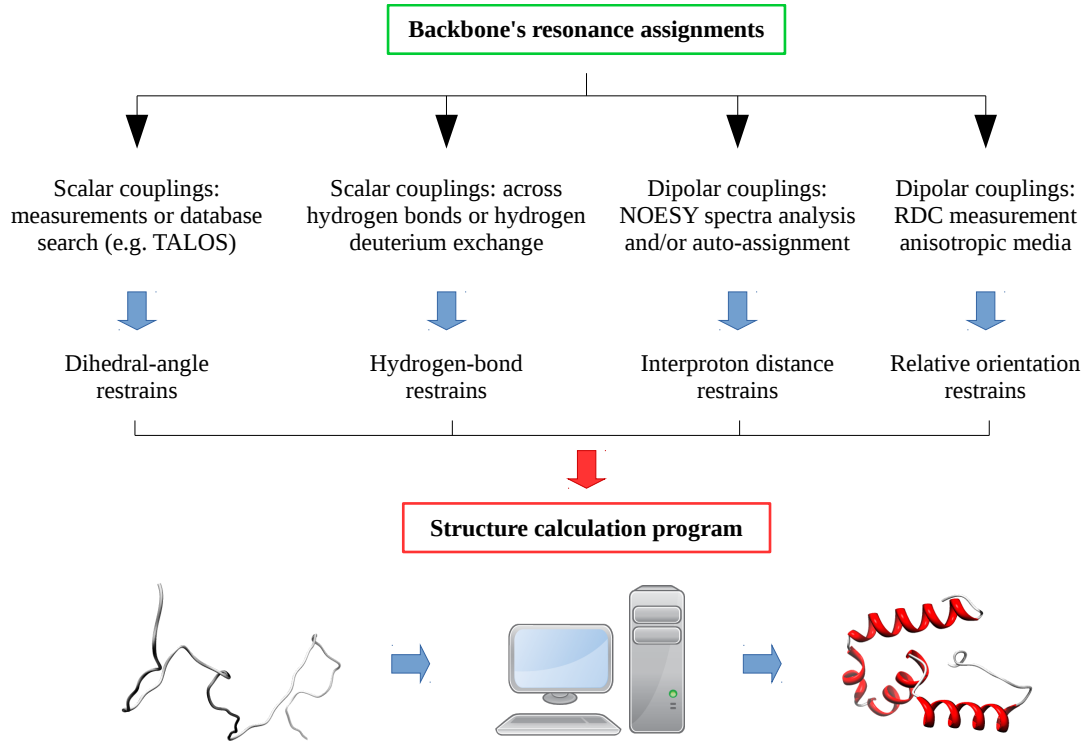


Figure 3.9. **Overview of macromolecular structure determination by NMR.** Analysis and assignment of backbone amino acids' chemical shifts lead to a four dataset of structural restrains, which are dihedral angles, hydrogen bonds, interproton distances and relative orientations. These are used as input to a computer program that using algorithms calculates a model which fits the experimental data.

Most of the computer programs utilized to solved NMR structures are based on molecular dynamics (MD) simulation in the presence of the experimental restrains. In classical molecular dynamics Newton's equation for motion are solved at the same time using the influence an empirically derived physical force field. Adding NMR restrains to this force field allows the structural refinement: At the beginning of simulation high temperatures are used so as to avoid any kind of non-specific interaction of the protein with itself. As the calculations continues simulation temperatures are reduced slowly integrating the restrains into the force field. These programs are able to do NOE auto-assignments and in every step where a transitional structure is calculated they evaluate the accordance between the restrictions data and the possible NOE assignments. Energetically unfavourable conformations which are not excused by restrains are discarded as well as these NOE “extreme” assignments. After several iterations decreasing temperature and restrain numbers, structures with good covalent geometry, minimal structural and experimental constrains violations and favourable non bounded interactions are selected.

Nowadays, there is a worldwide initiative to improve a NMR structural biology called WeNMR (Wassenaar et al., 2012) which offers and connects different tools that aid scientists in every single step of the structural elucidation: NMR data processing, resonance assignment, restraint definitions, structure calculation, protein docking and structure validation. Therefore is a very recommendable web portal in which different kind of programs will be found that could be very useful. In this regard, most successful programs for protein determination are CYANA (López-Méndez & Güntert, 2006), ARIA2 (Rieping et al., 2007), UNIO (Guerry & Herrmann, 2012) and NIH-Xplor (Schwieters et al., 2003).

3.1.3 NMR SAMPLE PREPARATION

When we are interested in elucidating the atomic structure of a certain protein first we have to adapt our sample to the experimental requirements. We have to keep in mind that data acquisition can be last for several days or even week and that our protein is going to be spinning for all that time inside the NMR tube at room or even higher temperatures depending on the case. Therefore it is worth to employ our time in designing a good NMR sample where our target protein and buffer selection is the optimum to ensure the protein survival to the data collecting process (Acton et al., 2011).

3.1.3.1 Protein quality and concentration

Broadly, even if greater purity is preferred sample's purity should be at least around 90 %. Besides protein sample should be homogeneous to make protein assignment easier. However, one of the main interesting point about NMR is that more than one protein conformation or even more than one protein can be studied at the same time. Moreover, this purity requirement is just to be sure that the only thing that we are measuring is our protein and nothing else. Nevertheless, it has been proven that this purity minima is not valid for all the cases as protein structure determination has been done directly in living cells by in-cell NMR spectroscopy (Sakakibara et al., 2009).

Depending on the type of experiment, different concentration amounts are required. For example, ^{15}N-HSQC spectra are routinely collected at 50-100 μM. This amount of protein is not too high compared to other techniques such as isothermal calorimetry, far-UV circular dichroism or surface plasmon resonance. Nonetheless, backbone assignment and other type of experiment need more sensitivity, or in another words, protein sample. For these cases protein concentration should be around 200 μM or more. People may think that the higher the better but always should be taken into account that some protein aggregates when they are concentrated close to their maximum solubility concentration.

3.1.3.2 Buffer selection

As mentioned before, in principle all buffers are compatible with NMR. However minimizing buffer concentrations is recommendable as well as using pH ranges from 4.5 to 7.5. Buffers may have many protons that interfere with proton NMR. Fortunately deuterated versions of many buffers are commercially available.

Protein compatibility with different buffers is more important than buffer and magnet compatibility. Taking this into account several approaches should be done in order to find a buffer in which our target protein will be comfortable if protein aggregation is wanted to be avoided.

Lamentably, there is no easy way to find which buffer is the best for a given protein. Several additives have been reported to improve protein solubility and stability such as glutamate and arginine mixture (Golovanov et al, 2004), triton, glycerol or even sodium sulphate (Ducat et al., 2006). Later we will discuss the way in which buffers are tested rapidly.

3.1.3.3 Protein alternative constructs selection

Big proteins (over 30-50 kDa) are not recommendable for NMR structure determination basically for two reasons. The first one is that their rotation correlation time is bigger (protein tumbling) which causes signals line broadening. The second one is that as they have more amino acids more signals will appear in the spectra, and even if we do multidimensional recordings overlapping may occur.

NMR is a very versatile technique which in contrast with crystallography it can deal perfectly with disordered segments or proteins. Nevertheless, if we have a protein which contains a big number of disordered regions and we are not interested in resolving that concrete part of the overall structure of the protein, it is recommendable to delete these parts from the target protein as disordered regions often promote aggregation and in this case deleteriously affect NMR spectral quality. Moreover, proteins with a significant amount of disordered regions are often degraded by *E. coli*. As mentioned in chapter 2, there are available a bunch of softwares specialized in the prediction of disordered regions. Alternative protein constructs can be design based on a consensus analysis of several disorder predictors, but in any cases the solubility and stability of all of them have to be tested.

If we are working with a big protein which case many multidomains, the most successful strategy has been the “divide and conquer” strategy.

3.1.3.4 Solubility and stability assays

When we are dealing with the screening of different protein constructs and different buffers in order to find the most favourable combination of both of the often we can end up with a large numbers of possibles solutions. This problem is similar to the X-ray crystallography where a hundred of thousands of crystallization buffers are tested with the aiming of obtaining crystals. In this case robotization has increased a lot the number crystal structures. Instead, NMR do not posses a rapid and simple screening protocols for NMR.

The best way of improving NMR acquisition is by getting better protein stability outside the magnet. Protein solubility and stability could be tested by doing pilot experiments of every construct in many different buffers and then analysing the result with SDS-PAGE gels. This process would be tedious so we can take advantage of several tactics developed by our crystallographers friends for out benefit.

Protein solubility can be easily monitored if we fused a fluorescent protein to our target protein (Drew et al., 2006). Traditionally green fluorescent proteins (GFP) have used for this purpose. GFPs have a exceptional stability and can report protein overexpression and solubility since the synthesis. In fact, after transforming bacteria with a plasmid carrying our recombinant protein sequence with GFP we can see in the colonies formed in the Petry dishes if the protein is been expressed and whether or not the protein has chances to be soluble. When protein is not soluble inclusion bodies are formed where the brightening of the GFP inside the bacteria is much lower compared to the soluble proteins.

After selecting putative soluble proteins a stability test has to be done to discard between all the alternative constructs which one is more likely to be more stable. Even if the following thermostability assay could be done with untagged proteins, we still can take advantage of the GFP fused protein as a solubility reporter (Hattori et al., 2012). This assay could be done with low protein concentrations (nanograms to micrograms). Purified or unpurified proteins are incubated over a range of temperatures for distinct periods of time. Then are applied to a gel filtration column with a fluorescence detector. The result provide a denaturation or “melting” temperature as a indicator of stability. The bigger the melting point is the better stability.

Buffer selection can be easily done by GFP tagged proteins too. Some groups have used 96-well assay plates to perform this test (Cabantous & Waldo, 2006). In each well different buffers are placed. Concentrated GFP fused proteins are then diluted in every well. Afterwards we can monitor the protein aggregation by centrifuging the plate and transferring all supernatants to a new 96 well-plate. Using a Fluorescence Plate Reader, protein brighting can be detected and quantified. Thus, best buffer can be easily chosen using this technique.

3.1.3.5 Protein isotopic labelling

After making all the efforts to select the best protein and buffer candidate it is time now we can proceed with the next step. Apart from hydrogen, natural abundances of isotopes which gives NMR spectra are really low. Therefore protein NMR is almost exclusive of recombinant proteins that have been isotopically labelled.

This isotopic labelling is made by a using a growing minimal medium in which the source of nitrogen and nitrogen is controlled and enriched in these isotopes by adding $^{15}NH_4Cl$ or ^{13}C-glucose. H_2O can also be substituted by D_2O which is mainly used for large proteins. By this relaxation properties of big proteins are improved as most of the ^{1}H atoms are changed to ^{2}H. Nevertheless, NH groups will exchange again to ^{1}H when the protein is purified in a H2O based buffer. Deuterium water has been related to protein aggregation (Lee & Berns, 1968).

Apart from these labelling modes there are more advances ways of selectively labelling specific amino acids in the protein. In the case of Ile, Val and Leu side-chain methyl groups labelling, a

triple labelling is done first with ^{2}H, ^{13}C and ^{15}N but one hour before induction α-ketobutyrate and α-keto-isovalerate are added. By this all protons will be deuterated but NH groups and the side-chains of Ile, Val and Leu. Moreover Thr, Met and Ala can be selectively labelled.

3.1.3.6 Final additives

Depending on the NMR tube we use different sample volumes are recommended. For a regular 5 mm tube 500 µl, for a Shigemi tube around 300 µl and for a 3 mm tube (which is often used to avoid high salt effects) 270 mm.

Apart from the protein, final sample should have 5 to 10 % of D_20 in order to do properly the "locking" and to be able to see labile proton exchanges. Moreover, some other additives are recommendable to avoid protein degradation such as DTT, DMSO, or PIC. Besides, the use of 2 to 10 µM of NaN3 (Sodium Azide) avoids microorganisms growth.

3.2 MATERIALS AND METHODS

In this chapter following codon optimized version of the $K_v7.2$ C-terminus were done using molecular biology techniques previously described in the materials and methods section of chapter 2. However there are some methods that are used in this chapter:

3.2.1 MOLECULAR BIOLOGY TECHNIQUES

Following **plasmids** were used:

Name	Provider	Size (bp)	Resistance	Tags	Cleavage
Easy-T	Promega (A1360)	3015	Amp		
pProEX-HTc	Invitrogen (10711018)	4780	Amp	His	TEV
pProEX-CaM (Homemade)	Based on pProEX-HTc (69864-3)	5211	Amp	His-CaM	Double TEV
pOKD4	(Dzivenu et al., 2004)	6473	Kan		
pCDNA3.1	ThermoFisher (V790-20)	5428	Amp Neo		

Following ***E. coli*** strains were used:

Strain	Provider	Humanized?	Extra plasmid resistance?
BSJ	SUNY Stony Brook	No	No
BL21(DE3)	Novagen (69450-3)	No	No
CodonPlus(DE3)(RIL)	Agilent (#230240)	Yes	Cam

Following **eukaryotic cells** were used:

HEK293T cells

3.2.1.1 Bacterial co-transformation

Bacterial co-expression have been used to express together CaM with different constructs of the C-terminus of K_v7.2 channels (not in tandem as before). In order to do that CaM was cloned in pOKD4 vector and the rest of K_v7.2 constructs in pProEX-HTc vector. Once our protein genes are cloned in the co-expression compatible vectors, they are co-transformed by electroporation. The machine used is the ECM630 (BTX A division of Genetronics) and the following values are set:

V = 2500 V
R = 0.200 Ω
Capacitance = 0025 µF

First, at the bottom of a pre-chilled electroporation cuvettes, 0.5 µl of each plasmid are added. Next, 50 µl of electrocompetent BL21(DE3) bacteria are added and everything is gently mixed and spread all the volume on the bottom to maximize the surface. Once the cuvette in adjusted in the holder or electroporation cell, a pulse of 5.9 µs is sent and immediately 1 ml of LB without antibiotics is added. Bacteria are then incubated for recovery no longer than 40 min at 37 °C. Afterwards, bacteria are then centrifuged for 1 min at maximum speed, we get rid of the 950 µl of the supernatant, we resuspend the bacterial pellet with the remaining LB and they are plated in a Petri with the appropriate antibiotics (ampicillin and kanamycin in this case).

Protocol 3.1: Co-transformation

- Thaw one vial (or more) of electrocompetent cells.
- In pre-chilled electroporation cuvette add 0.5 µl of each plasmid from a miniprep (1 µg/µl)
- Add 50 µl of BL21(DE3) electrocompetent cells and mix and spread gently.
- Configure electroporation machine as V =2500 V, R = 0.2 Ω and Capacitance = 0025 µF
- Adjust the cuvette in the holder and send the pulse (around 5.9 µs).
- Add 1 ml of LB without antibiotics.
- Take it to the shaker at 37 ºC for 40 mins.
- Spin down the cells for 1 min at maximum speed.
- Pour out 950 µl of the supernatant.
- Resuspend the pellet and seed a Petry dish with the corresponding double antibiotics.
- Incubate the plates inverted at 37 ºC until colonies have grown (6 – 12 h)
- Plates can be stored at 4 ºC for weeks.

3.2.1.2 Eukaryotic cell transfection

All DNA used to transfect cells were purified by alkaline lysis and affinity columns from a 100 ml of bacteria culture (Roche, Genopure Plasmid Midi Kit, Ref.: 3143414). DNA was quantified using a NanoDrop 2000 Spectrophotometer, (Thermo Fischer) and diluted to 1 µg/µl solution. The purity of the DNA was estimated by the 260/280 ratio.

Lipofectamine transfection was used for electrophysiological experiments. Lipofectamine (Lipofectamine 2000, Invitrogen) is based on liposomes that can be easily merged with cell membranes since they are both made out of phospholipid bilayers. Lipofectamine generally uses cationic lipids to form an aggregate with the anionic genetic material. While the main advantage of this method is its high transfection efficiency, the main disadvantage is its high price and toxicity.

It is a quite simple protocol. First, prepare two tubes; one is a mixture of the DNA with uncompleted DMEM (without FBS) and the second one a mixture of lipofectamine also with uncompleted DMEM. Incubate both mixtures for 3-5 minutes and mix both of them together. Let everything incubate for 20 min. In the meanwhile, remove media from cells, clean once with PBS 1 X, to get rid of any trace of serum, and add half of the volume of uncompleted medium. This way, when after the incubation the mix of lipofectamine + DNA is added it would be more concentrated. Less than six hours later, remove the media of the cells and add complete medium supplemented with gentamycin.

3.2.2 CONSTRUCT SELECTION

3.2.2.1 Thermostability Assay Based on Venus Fluorescence

In order to screen rapidly which deletion between helices A and B is most favourable for the [CaM-K_v7.2 C-terminus] complex stability, Venus fluorescent protein was fused to the C-terminus of the following constructs.

- His-CaM-Q2ABc-Venus
- His-CaM-Q2ABc-Venus, with Δ2 (ΔT359_T501)
- His-CaM-Q2ABc-Venus, with Δ6 (ΔR374_K493)
- His-CaM-Q2ABc-Venus, with Δ6L (ΔR374_T501)
- His-CaM-Q2ABCD-Venus
- His-CaM-Q2ABCD-Venus, with Δ2 (ΔT359_T501)
- His-CaM-Q2ABCD-Venus, with Δ6 (ΔR374_K493)
- His-CaM-Q2ABCD-Venus, with Δ6L (ΔR374_T501)

Venus is a improved Yellow Fluorescent Protein which has a maximum excitation at 515 nm and a maximum emission at 528 nm (Nagai et al., 2002). Quantum yield is around 0.57 and the Epsilon coefficient at 515 nm is 92982.45 $M^{-1}cm^{-1}$. As described by authors Venus stability is even higher than traditional GFP which makes it very interesting reporter of protein solubility and stability. Venus sequence was optimized for *E.coli* and it was *de novo* synthesized by ShineGene enterprise with the desired restriction sites to make cloning process easier.

Pilot experiments were done (see Chapter 2 Materials and Methods). Protein lysis was done in 10 mM Tris pH 7.4, 120 mM KCl, 10 mM NaCl, 2 mM $MgCl_2$ and 10 mM EGTA. SN was collected and protein concentration was measured based on Venus absorption. Also Denaturing-SDS, Not-Denaturing-SDS and Native-PAGE gels were run to detect protein fragmentation.

Protein thermostability was tested with a alternative version of the method described by Gouaux (Hattori et al., 2012). All proteins were diluted or concentrated to 15 μM depending on the case. From the SN of each construct 50 μl were transferred to new 6 PCR Eppendorfs. A Eppendorf of each construct was then placed in a thermal cycler which are often used for PCR. Every sample was heated at different temperatures (4, 30, 60, 70, 80 and 90º C) for 10 minutes. Protein aggregates were centrifuged at maximum speed for 10 minutes and SNs were collected. Afterwards samples were loaded into Not-Denaturing-SDS-PAGE gels which basically consist on running a normal SDS-PAGE gel but without boiling or heating the sample. By this the Native-State of the protein as well as the Venus brightening remains. Gels without staining were pictured with VersaDoc using a 488 nm light in which Venus is visible. Moreover same gels were stained with Coomassie and pictures were taken again with VersaDoc white light. Band

densities were measured with ImageJ. Values were referenced with the density obtained from the first sample at 4 ºC.

3.2.2.2 Q2ABc Δ6L Alternative Constructs With Fused CaM and Venus

Intrinsically disordered regions of Q2ABc construct (residues 310-548 from human K_v7.2) were detected using PONDR VL-XT software. Afterwards two synthetic sequences were design adding some deletions in the N- and C-terminus of the protein. Moreover, specific restriction sites were introduce to make the combination of these residues easier. Mutations in the N-terminus are ΔF316_R325 and ΔF316_N334. Mutations in the C-terminus are ΔP533_H546 and ΔM518_H546. Designed DNA was optimized and synthesized by *ShineGene*. These four mutations were combined among them with the original Q2ABc Δ6L (Figure 3.10).

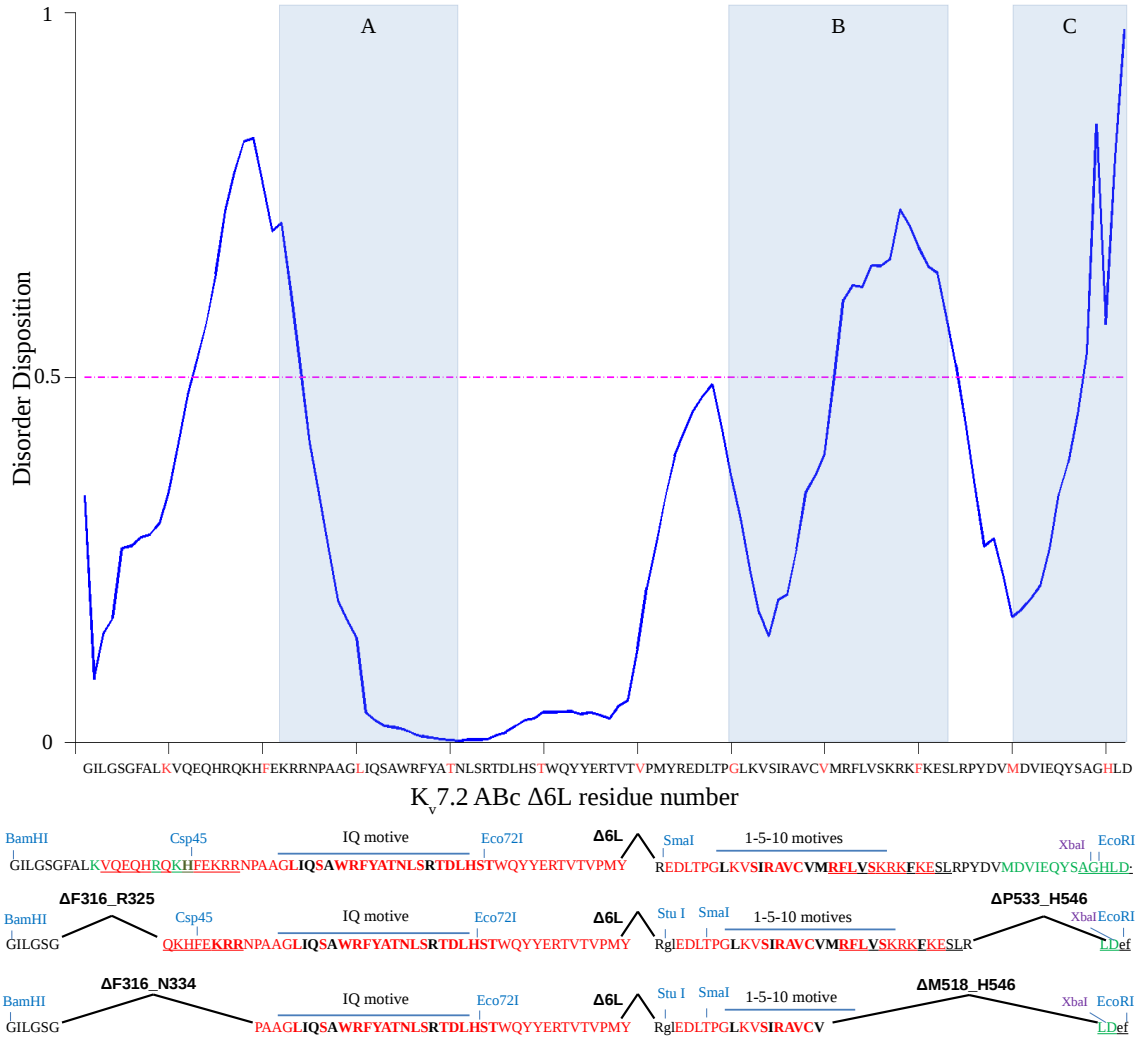


Figure 3.10. PONDR VL-XT disorder prediction and designed amino acid sequences trying to minimize these regions.

Eventually from the 9 combinations only the following 6 (Figure 3.11) were cloned as expressed with fused CaM and Venus as His-CaM-(ALTERNATIVE)-Venus:

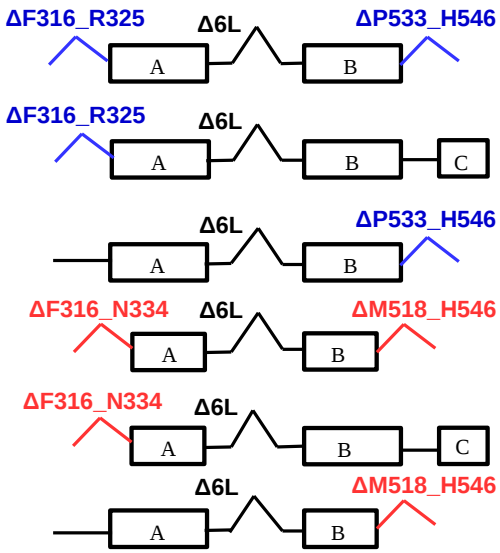


Figure 3.11. Combinations of tested deletions.

Pilot experiments were done (see Chapter 2 Materials and Methods). Protein lysis was carried out in 10 mM Tris pH 7.4, 120 mM KCl, 10 mM NaCl, 2 mM $MgCl_2$ and 10 mM EGTA. Supernatant was collected and protein concentration was measured based on Venus absorption. Nevertheless Not-Denaturing-SDS and Native-PAGE gels were run to detect protein fragmentation if needed.

3.2.2.3 Q2ABc Δ6L Alternative Constructs Co-Expressing CaM

CaM was cloned in pOKD4 vector and previously done 6 alternative contracts plus Q2ABc and Q2ABc Δ6L were cloned in pProEX-HTc vector, which has a 6 histidines in the N-terminus. Pilot experiment were done as described before with the only difference that in LB media either Ampicillin and Kanamycin were added. SN and P fractions were load in SDS-PAGE and Native-PAGE gels and protein solubility was calculated.

3.2.2.4 Electrophysiological recordings

The patch-clamp technique (Ogden & Stanfield, 1981) uses a single microelectrode with a relatively large tip opening of 1-2 µm for stimulation and acquisition of the electrical signals. Measurements of the plasma membrane current were performed with the whole-cell patch-clamp technique. HEK293T cells were obtained at RT (21-25°C) 48 h after transfection using lipofectamine 2000. Cells were bathed in the following solution: 140 mM NaCl, 4 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 10 mM HEPES (Na), 5 mM D-glucose, adjusted to pH 7.4 with NaOH. The osmorality was adjusted with mannitol to ~315 mOsm. Pipettes were pulled from borosilicate glass capillaries (Sutter Instruments) using a Narishige micro-pipette puller (PC-10; Narishige Instrument). Membrane currents were measured using an EPC-8 amplifier (HEKA) with pipette and membrane capacitance cancellation.

Pipettes were filled with an internal solution containing 125 mM KCl, 10 mM HEPES (K), 5 mM $MgCl_2$, 5 mM EGTA, 5 mM Na_2ATP, adjusted to pH 7.2 with KOH and the osmolarity adjusted to ~300 mOsm with mannitol. The amplitude of the K_v7 current was defined as the peak difference in current relaxation measured at -30 mV after 500-1,500 ms pulses to -110 mV (all channels closed) and to +30 mV (all channels opened). The data were acquired and analysed using pCLAMP software (version 8.2), normalized and plotted using OppenOffice Calc.

These experiments were performed with the help of Dr. Carolina Gomis-Perez.

3.2.3 BUFFER SCREENING

Venus protein was fused to the C-terminus of Q2ABCD Δ6L construct in pProEX-HTc vector. CaM was co-transformed as before. Complex production was carried out in 1 L of LB media with 0.5 mM of IPTG for 6 h at 37 ºC. Bacteria were collected and resuspended in 40 ml of *Lysis Buffer* (120 mM KCl, 10 mM K-HEPES pH 7.4, 3 mM Imidazole, 1 mM PMSF, 1 tablet of Protease inhibitor without EDTA). Afterwards, 3 passes through the Emulsiflex were done and, later bacteria were centrifuged in the JA25.50 rotor at 25000 g for 30 min at 4 °C. The rescued supernatant was filtered with 0.22 µm filter before running the ÄKTA.

First, ÄKTA tubes were washed with degassed ddH_2O and afterwards the column HiTrap Talon Crude 5 ml (GE Healthcare, Ref: 28-9537-67), was connected to the ÄKTA with slow flow to take out the bubbles. The ethanol from the columns was washed out with 5 column volumes (CVs) of ddH_2O at 5 ml/min flow. After equilibrating the column with 10 CVs of *Washing Buffer* (120 mM KCl, 10 mM K-HEPES pH 7.4), at 5 ml/min flow, the sample was injected at 1 ml/min flow up to 50 ml. 5 ml fractions were collected.

Next, the column was washed with different combinations of *Wash Buffer* and *Elution Buffer* (120 mM KCl, 10 mM K-HEPES pH 7.4, 300 mM Imidazole): W0 = 10 CV of buffer A (0 mM imidazole) - 2 ml/min; W1 = 4 CV of 3 % mixture (9 mM imidazole) 2 ml/min; W2 = 8 CV of 6 % mixture (18 mM imidazole) 2 ml/min; and W3 = 4 CV of 8 % mixture (24 mM imidazole) 2 ml/min. Finally the complex was eluted with a gradient from 30 to 300 mM of imidazole in 8 CVs at 2 ml/min flow. 1 ml fractions were collected and these with high 280 nm absorbance were picked. A 12 % acrylamide SDS-PAGE was run to check the presence of the protein in the fractions. To regenerate the column apply 10 CVs of 20 mM MES pH 5.0, afterwards 10 CVs of ddH_2O and finally 10 CVs of EtOh 20 %.

All the positive fractions were pooled together and dialysed against 1 l of *Fluorescence Buffer* (120 mM KCl, 5 mM NaCl, 1 mM $MgCl_2$, 10 mM Tris pH 7.4 and 10 mM EGTA) O/N at 4 ºC in dialysis tubing cellulose membrane (Sigma-Aldrich , Ref: D927-100FT). Protein complex was concentrated until 500 µM using Amicon Ultra-15 centrifugal units with a 3kDa cut-off and protein concentration was measured using Nanodrop.

Empty 96-well plate was filled with 90 µl of buffers in each well as follows (Table 3.1):

	MES B.			$K_X(PO_4)$ B.			TRIS B.			HEPES B.		
	pH 5	pH 6	pH 7	pH 5	pH 6	pH 7	pH 7	pH 8	pH 9	pH 7	pH 8	Controls
	1	2	3	4	5	6	7	8	9	10	11	12
A	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	-
B	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	*Fluo. B.*
C	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	Protein complex in *Fluo B*
D	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	Venus protein
E	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	MES pH 6 25 + 50 KCl + AA
F	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	K_X(PO4) pH 6 25 B + 50 Kcl + AA
G	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	TRIS pH 9 25 B + 50 Kcl
H	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	HEPES pH 8 25 B + 50 Kcl

Table 3.1. Buffer screening preparation table. Four different buffers (MES, Kx(PO4), TRIS and HEPES) were used at two different concentrations (25 and 50 mM) at different pHs taking into account four different ionic strengths (50, 100, 150 and 200 mM of KCl). Besides, a number of controls were used as well as a combination of 50 mM of L-Glu and L-Arg (AA). All the units inside the box are in mM.

Afterwards, 10 µl of concentrated protein was added to each well. Mixture was incubated at room temperature for a week. Fluorescence measurement were taken every day after centrifugation the well-plate and transferring all the SN to a new plate using Synergy HT plate reader (BioTek , Ref.: 7091000).

Differences were seen in this process but protein remain stable in most of them. Therefore, in order to stress protein stability the plate was incubated at 50 ºC for 1 h and fluorescence was measured. Output data were then analysed with LibreOffice Calc.

3.2.4 Protein Complex Purification for NMR

3.2.4.1 Whole complex labelling

CaM and His-Q2AB with deletions ΔF316_R325, Δ6L and ΔP533_H546 (GB116) were co-transformed by electroporation as before. From the Petry dish, 10 ml of LB media with antibiotics were inoculated O/N at 37 ºC. Next day, two different dilutions: on one hand a 1:200 dilution was done to fresh 10 ml sterile Falcon in order to do a rapid solubility and expression test, and on the other hand a 0.2 µl were inoculated in another fresh 10 ml sterile Falcon. After checking protein expression, in the evening new dilutions were done using different bacterial volumes (5, 10, 25 and 50 µl) into 4 Erlenmeyers with 200 ml of ^{14}N-^{12}C-M9 media (Protocol 3.2)

Next day, OD_{600} was measured from each Erlenmeyer. The volume from the Erlenmeyers needed to inoculate the new labelled media to a starting OD_{600} 0.05-0.1 was calculated. Afterwards this volume was collected and centrifuged at 2000 g for 30 mins at RT. SN was discarded and bacterial pellet was resuspended with the M9 labelled media. Bacteria were growth at 37 ºC until OD_{600} reached 0.3, then temperature was reduced to 25 ºC and when OD_{600} reached 0.5 protein expression was induced with 0.5 mM IPTG for 12 h. Following day bacteria were centrifuged with JLA 8.1 rotor at 4 ºC with 5000 g for 30 mins. SN was discarded and P was resuspended with pre-chilled *Wash Buffer* (120 mM KCl + 20 mM $K_X(PO_4)$ pH 7). If needed, bacteria were once again centrifuged at 5000 g at 4 ºC in a Falcon. SN was discarded and bacterial pellet was frozen with liquid N_2 and store in the freezer until purification.

For HSQCs, T1 relaxation and N-NOESY experiments 3 litres of ^{15}N-^{12}C-M9 was used. For backbone assignment and C-NOESY experiments 3 litres of ^{15}N-^{13}C-M9 prepared in 50 % of D_2O.

Protocol 3.2: M9 media preparation

- Prepare a 10X M9 stock:
  - 60 g Na2HPO4
  - 30 g KH2PO4
  - 5 G NaCl
- Dilute to desired volume with ddH_2O to 1X.
  - * For **50 % D_2O** media dilute 10X stock to 2X with ddH_2O)
- Autoclave previous liquids.
- For 1 litre of media add (everything sterile!)
  - * For **50 % D_2O** media add now corresponding volume to reach 1X M9 and filter with 0.22 µm
  - 1 g $^{14}NH_4Cl$ or **1 g $^{15}NH_4Cl$**
  - 3 g ^{12}C-Glucose or **2 g ^{13}C-Glucose**
  - 1 mM $MgSO_4$
  - 0.1 mM $CaCl_2$
  - 1 mg/ml Thiamine
  - 1 mg/ml Biotin
- Add corresponding antibiotics
- Inoculate the cultures with the desired amount of bacteria

3.2.4.2 Whole protein purification

Frozen bacterial pellets were thawed in ice and after resuspended with 40 ml of *Lysis Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7, 1 mM PMSF, 2 mM Imidazole and 1 tablet of Protease inhibitor without EDTA). Afterwards, 3 passes through the **Emulsiflex** were done and, later bacteria were centrifuged in the JA25.50 rotor at 25000 g for 30 min at 4 °C. The rescued supernatant was filtered with 0.22 µm filter before running the ÄKTA.

First, ÄKTA tubes were washed with degassed ddH_2O and afterwards the column **HiTrap Talon Crude** 5 ml (GE Healthcare, Ref: 28-9537-67), was connected to the ÄKTA with slow flow to take out the bubbles. The ethanol from the columns was washed out with 5 column volumes (CVs) of ddH_2O at 5 ml/min flow. After equilibrating the column with 10 CVs of *Wash Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7), at 5 ml/min flow, the sample was injected at 1 ml/min flow up to 50 ml. 5 ml fractions were collected.

Next, the column was washed with different combinations of *Wash Buffer* and *Elution Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6, 300 mM Imidazole): W0 = 10 CV of buffer A (0 mM imidazole) - 2 ml/min; W1 = 4 CV of 3 % mixture (9 mM imidazole) 2 ml/min; W2 = 8 CV of 6 % mixture (18 mM imidazole) 2 ml/min; and W3 = 4 CV of 8 % mixture (24 mM imidazole) 2 ml/min. Finally the complex was eluted with a gradient from 30 to 300 mM of imidazole in 8 CVs at 2 ml/min flow. 1 ml fractions were collected and these with high 280 nm absorbance were picked. A 12 % acrylamide SDS-PAGE was run to check the presence of the protein in the

fractions. To regenerate the column apply 10 CVs of 20 mM MES pH 5.0, afterwards 10 CVs of ddH_2O and finally 10 CVs of EtOh 20 %.

All the positive fractions were pooled together and **dialysed** against 1 litre of *Running Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6 and 5 mM EGTA) O/N at 4 ºC in dialysis tubing cellulose membrane (Sigma-Aldrich, Ref: D927-100FT). Protein complex was concentrated until 10 to 13 ml using Amicon Ultra-15 centrifugal units with a 3kDa cut-off in a fixed angle rotor at 3500 g if needed to inject the sample.

A second purification step was done by **gel filtration**. First ÄKTA tubes were washed with degassed ddH_2O. Afterwards, before connecting the Superdex 75 pg 26/60 (GE Healthcare, Ref.: 28-9893-36) column to the ÄKTA, pressure alarm was set at 0.5 MPa. Slow-flow was allowed to pass to get rid of the bubbles in the system. To wash out the ethanol from the system, 1 CV (320 ml) of ddH_2O was passed at 0.5 ml/min approximately. Next the column was equilibrated with 1 CVs of degassed *Running Buffer* at slow-flow (to be done O/N).

Next up to 13 ml of dialysed and 0.22 µm filtered sample was injected. The flow was speed up to 1.5 ml/min. Elution started after 100 ml approximately (void volume) and 4 ml samples were collected. The ones with high 280 nm absorbance were picked up and checked in a 12 % SDS-PAGE. The column was washed with 1 CV of ddH2O at 0.5 ml/min and then with 1 CV of 20 % EtOh at the same rate to store it again.

Positive fractions were pooled together and the protein was concentrated using a to 1 ml with a 3 kDa cut-off Amicon ultra-15 concentrator in a fixed angle rotor at 3500 g. When reached this point 5 % of D_2O was added to the sample as well as a NaN_3 and PIC to a final concentration of 3 µM and 1 µM respectively. Afterwards protein concentration was again continued until a final volume of 500 µl was achieved. Later protein concentration was measured using NanoDrop and concentrated protein was transferred to a 5 mm NMR tube.

3.2.4.3 Independent labelling of each protein from the complex

Protein production was done separately. CaM and GB116 were independently transformed by heat shock technique in BL21(DE3)CodonPlus. Unlabelled protein was produce in normal M9 media as mentioned in 3.2.4.1 section of this chapter. When CaM was labelled with ^{15}N-^{13}C-M9 media. In the case of Q2AB protein ^{15}N-^{13}C-M9 with 50 % of D_2O media was used.

3.2.4.4 Complex reconstitution on-column

Channel peptide tends to aggregate easily and the refolding by itself is impossible. However K_v7.2 C-terminus refolding in the presence of CaM leads to complex formation even if the yield is low.

There are other three ways to refold the protein:

1. Temperature-leap tactic. Protein is allowed to refold at low temperatures, to minimize aggregation, and then the temperature is rapidly raised to promote fast folding after the intermediates responsible for aggregation have been depleted. (Xie & Wetlaufer, 1996)

2. Refolding by dilution to final denaturant concentrations that are high enough to solubilize aggregates but low enough to promote proper folding. (Maeda, Ueda, & Imoto, 1996)

3. On-column refolding. Denatured protein is bound the HiTrap column and denaturing agent is drastically removed (Lemercier et al., 2003). Nevertheless protein should be diluted in order to avoid the aggregation.

The method explained here is a our modified version of on-column refolding. Our tactic is to first make sure that the peptide is bound to the column in presence of 7.5 M urea. As it should be diluted we will use 3 HiTrap columns in tandem. CaM is folded below 2 M of Urea, (Guerini & Krebs, 1983). Therefore we will prepare a CaM with 1 M urea and then we will perform the rapid refolding of the K_v7 protein in presence of CaM.

3.2.4.4.1 Calmodulin purification

CaM purification was done independently as it was labelled or not, following the instructions in 2.2.4.1 section. Nevertheless, when protein was eluted from the Phenyl-Sepharose, instead of dialysing against ddH_2O it was dialysed with 120 KCl, 20 mM $K_X(PO_4)$ and 1 M Urea. Protein was later concentrate until a final volume of 20 ml reaching a CaM concentration around 3 mg/ml.

3.2.4.4.2 K_v7.2 alternative construct resolubilisation from inclusion bodies

Frozen bacterial pellets were thawed in ice and after resuspended with 40 ml of *Lysis Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7, 1 mM PMSF, 2 mM Imidazole, 1 % Triton and 1 tablet of Protease inhibitor without EDTA). Afterwards, 3 passes through the **Emulsiflex** were done and, later bacteria were centrifuged in the JA25.50 rotor at 25000 g for 30 min at 4 °C. In this case SN was taken apart and we got focused on the pellet where the inclusion bodies are.

Pellet was resuspended and washed with *Wash Buffer I* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7 and 2 mM Imidazole). Afterwards resuspended media was centrifuges again in the JA25.50 rotor at 25000 g for 30 min at 4 °C. SN was discarded. This process was repeated with *Wash Buffer II* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7, 2 mM Imidazole and 1 M Urea). Next pellet was

resuspended in the *Resolulizing Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7, 2 mM Imidazole and 7.5 M Urea). This time after centrifugation SN was collected and filtered with a 0.22 µM filter before injecting to the HiTrap columns.

3.2.4.4.3 Complex formation on-column

ÄKTA tubes were washed with degassed ddH_2O and afterwards three HiTrap Talon Crude 5 ml columns were connected in tandem. The ethanol from the columns was washed out with 5 CVs (75 ml) of ddH_2O at 5 ml/min flow. After equilibrating the column with 10 CVs of *Equilibrium Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7 and 7.5 M Urea), at 5 ml/min flow. Then, $K_v7.2$ channel sample was injected at 1 ml/min flow up to 50 ml. Afterwards 5 CVs of *Equilibrium Buffer* were used to clean non-specific proteins.

CaM was now injected decreasing the urea concentration from 7.5 to 1 M, which was visible in the flow-through density. Next, column was washed with *Wash Buffer III* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 7 and 1 M Urea) and Elution Buffer (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6, 300 mM Imidazole): W0 = 10 CV of buffer A (0 mM imidazole) - 2 ml/min; W1 = 4 CV of 3 % mixture (9 mM imidazole) 2 ml/min; W2 = 8 CV of 6 % mixture (18 mM imidazole) 2 ml/min; and W3 = 4 CV of 8 % mixture (24 mM imidazole) 2 ml/min. Finally the complex was eluted with a gradient from 30 to 300 mM of imidazole in 8 CVs at 2 ml/min flow. 1 ml fractions were collected and these with high 280 nm absorbance were picked. A 12 % acrylamide SDS-PAGE was run to check the presence of the protein in the fractions. To regenerate the column apply 10 CVs of 20 mM MES pH 5.0, afterwards 10 CVs of ddH_2O and finally 10 CVs of EtOh 20 %.

All the positive fractions were pooled together and **dialysed** against 1 litre of *Running Buffer* (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6 and 5 mM EGTA) O/N at 4 ºC in dialysis tubing cellulose membrane. Protein complex was concentrated to 10 - 13 ml using Amicon Ultra-15 centrifugal units with a 3kDa cut-off in a fixed angle rotor at 3500 g if needed to inject the sample.

Following gel filtration and sample preparation steps were done as described previously in 3.2.4.2. section.

3.2.4.5 Anisotropic media for RDCs recording

After protein labelling, purification and concentration to 225 µl at 300 µM. Taking into the account that the final volume of the sample would be 450 µl Otting phase media was prepared (Rückert & Otting, 2000).

180 µl of sample's buffer (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6 and 5 mM EGTA) was prepared with the 8 % of C12E5 (dodecyl-pentaethylene-oxide)(Sigma-Aldrich, Ref: 76437-1G). To this mixture, 1-hexanol (Sigma-Aldrich, Ref: 471402-1G) µl were added, vortex and centrifuged in every microlitre addition. The anisotropic media formation was monitored as follows:

- 0 µl → we could observe a biphasic separation, cloudy.
- 1 µl → The mixture became really cloudy, milky, opaque.
- 2 µl → Less cloudy. It was becoming more and more clear with the time.
- 3 µl → After waiting a bit transparent it became transparent.
- 4 µl → Still transparent and it was becoming less dense.
- 5 µl → Still cloudy.
- 6 µl → Cloudy again and by the time it became biphasic again.

This process was done in order to determine the optimal 1-hexanol volume addition to the 8 % C12E5 so as to form a homogeneous anisotropic media.

Again 180 µl of sample's buffer (120 mM KCl, 20 mM $K_X(PO_4)$ pH 6 and 5 mM EGTA) was prepared with the 8 % of C12E5 (dodecyl-pentaethylene-oxide). 4 µl of 1-hexanol were added to the mixture microlitre by microlitre vortexing and centrifuging the mixture. At this point Otting media should be transparent and homogeneous. Afterwards 0.45 µl of PIC and NaN_3 were added as well as 22.5 µl of D_2O. Sample's buffer was added adjusting the final volume to 225 µl. The molar ratio between C12E5/1-hexanol was 1.23.

Eventually, labelled protein sample and Otting media sample were mixed leading to a final product of 150 µM of labelled protein, 4 % of C12E5 with 1-hexanol, 120 mM 120 mM KCl, 20 mM $K_X(PO_4)$ pH 6, 5 mM EGTA and 5 % D_2O.

TROSY and semi-TROSY experiments were run in the isotropic media and in the anisotropic media for ^{NCD}GB116 and for ^{NCD}Complex proteins. The first experiment provides the scalar coupling “J” which for N-H should be around -92 Hz. The second experiments provides the the total coupling “T” which with is defined by the summary of the J coupling with the residual dipolar coupling (T=J + RDCs). Therefore, in order to extract the information we do the following:

$$\text{J coupling} = \text{Isotropic TROSY} - \text{Isotropic semi-TROSY};$$
$$\text{T coupling} = \text{Anisotropic TROSY} - \text{Anisotropic semi-TROSY};$$
$$T = J + RDCs;$$
$$RDCs = T - J$$

For the RDCs representation, we modelled the the crystal structure PDB: 4UMO in which helices AB from the $K_v7.4$ are bound to CaM with the sequence of $K_v7.2$ channel using SWISS-MODEL (Guex & Peitsch, 1997).

3.2.5 SIDE-CHAIN AND NOE ASSIGNMENTS

Chemical shifts for ^{13}C and ^{1}H were assigned for each amino acid's side-chain nuclei. HccoNH for ^{1}H assignment and HNCccoNH for ^{13}C assignment were used for this purpose. Nevertheless, the assignment were not done for all the nuclei in the side-chain as in some cases the signal-to-noise ratio resulted to be really low. Unfortunately, affected residues were specially these that were in a shielded environments. In another words, these amino acids which were making contacts had more relaxation ways an signal sensitivity was decreased in these regions.

After side-chain assignments, ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC experiments were collected in order to obtain short-, media- and long-range restrains. For the short- and media-range NOEs manual assignments were done for most of the amino-acids. Afterwards, an automatic peak picking was done in both spectra using sparky so as to create a list of ambiguous contacts.

3.2.6 STRUCTURE CALCULATION

For each protein, ARIA2 program was used to create atomic model in which experimental data was fitted. The number of structures per iteration was adjusted to 50 and the number of structures in the last cycle to 100. The final number of refines structures were 10. Torsion angles were uploaded using XPLOR-NIH format. Chemical shifts list were adjusted to XEASY/CYANA format. Assigned and ambiguous NOE restrains were added in XEASY/CYANA format too. For the automatic assignment, chemical shift tolerances were established as follows: for ^{1}H 0.02 ppm, for ^{13}C 0.2 ppm and for ^{15}N 0.2 ppm. NOESY mixing time was 120 ms and the rotation correlation time of the complex was calculated to be 16 ns. Spectrometer frequency was defined as 800 MHz. and tolerance the which an automatic NOE assignment using the ambiguous NOE list. Besides, RDCs were loaded in SANI format. The first iteration in which RDCs were used was the 6^{th}. Rhombicity coefficient was determined to be 0.2218 and the axial coefficient -18.9.

3.2.7 PROTEIN DOCKING

When the structure of every member of the complex was known, information derived from the inter-protein contacts was adjusted to HADDOCK2.2 program as well as other parameters:

For GB116:

- Direct contacts: 44, 45, 47, 48, 49, 51, 95, 96, 99, 100, 102 and 103.
- Passive residues: Automatically defined around active residues.
- Flexible segments: 1 – 32 and 75 – 84.
-Semi-flexible segments: Manually: 33 – 35, 57 – 63, 73 – 74, 85 – 86 and 113 – 115.

For CaM:

- Direct contacts: 19, 51, 54, 71, 76, 87, 88, 89, 92, 109, 112, 113, 114, 115, 116 and 124.
- Passive residues: Automatically defined around active residues.
- Flexible segments: 76 – 81
- Semi-flexible segments: Manually: 1 – 4, 20 – 29, 39 – 43, 55 – 63, 76 - 77, 93 – 100, 111 – 117, 128 – 137, and 147 – 148.

Number of structures for rigid body docking was 1400.

Restraints for energy constants, at energy constants per stage was configured to hot 0.1, cool1 1, cool2 5, cool3 5.

Force constant for surface contact restraints was 2.

Centre of mass restraint was established to 2.

Scaling of intermolecular interactions for rigid body EM was fixed in 0.1.

3.2.8 NMR ACQUISITION AND PROCESSING

All the experiments above were acquired on an 800 MHz Bruker Avance III spectrometer always at 303 K.

For the whole complex labelled following TROSY experiments were done:

- Backbone assignment: ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNcoCA, HNcaCB, HNcocaCB, HNCO, HncaCO, HNcaHA and HNcocaHA.
- Side-chain assignment: HccoNH and HNCccoNH
- RDCs: Isotropic TROSY ^{15}N-HSQC and Semi-TROSY ^{15}N-HSQC and Anisotropic TROSY 15N-HSQC and Semi-TROSY 15N-HSQC.
- NOE assignment: ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC.

For the complex with only CaM labelled with ^{15}N and ^{13}C:

- Backbone confirmation assignment: ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNCO and HNcaHA.

For the complex with only K_v7.2 labelled with ^{15}N and ^{13}C and 50 % of D_2O:

- Backbone confirmation assignment: ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNcaCB, HNCO, HNHA, and HNcaHA.
- RDCs: Isotropic TROSY ^{15}N-HSQC and Semi-TROSY ^{15}N-HSQC and Anisotropic TROSY 15N-HSQC and Semi-TROSY 15N-HSQC.
- NOE assignment: ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC

All the spectra were processed with Topsping 3 (BRUKER) and they were analysed with Sparky (T. D. Goddard and D. G. Kneller, University of California) and NMRPipe (Delaglio et al., 1995).

WeNMR platform was used as a starting point for structural elucidation from the obtained data:

- Chemical shift were referenced and protein flexibility was calculated using Random Coill Index (Berjanskii & Wishart, 2005).
- Dihedral angles were predicted using TALOS+ (Shen et al., 2009).
- RDCs were analysed by home made MATLAB scripts by which Da and R factors were estimated. Nevertheless, MODULE (Dosset et al., 2001) and PALES (Zweckstetter, 2008) were used to check the reliability of the data.
- ARIA2 (Rieping et al., 2007) and was used to perform the NOESY auto-assignment of the ambiguous restrictions and structural models of each protein were obtained using all the information above.
- Quality of single proteins was calculated using PROCHEK software.
- Protein-protein docking was performed with HADDOCK2.2 (de Vries et al., 2007).

3.3 RESULTS

3.3.1 "GB116" K_V7.2 HELICES A-B ALTERNATIVE CONSTRUCT IS SELECTED FOR STRUCTURAL ANALYSIS

3.3.1.1 Deletion ΔR374_K493 (Δ6L) between helices AB has high thermostability

Based on Venus fluorescent protein the thermostabillity effect of Δ2 (ΔT359_T501), Δ6 (ΔR374_K493) and Δ6L (ΔR374_T501) in His-CaM-Q2ABc-Venus and His-CaM-Q2ABCD-Venus. After lysis, protein concentration in crude supernatants was measured taking advantage of Venus properties for each construct (Figure 3.12.A). By this procedure we tested again that Δ6L constructs were producing higher soluble protein than the others. Afterwards, protein concentrations were matched to 15 µM the and SNs were then incubated at different temperatures for 10 minutes. After discarding any protein aggregate by centrifugation samples were run in non-denaturing SDS-PAGE gels (Figure 3.12.B).

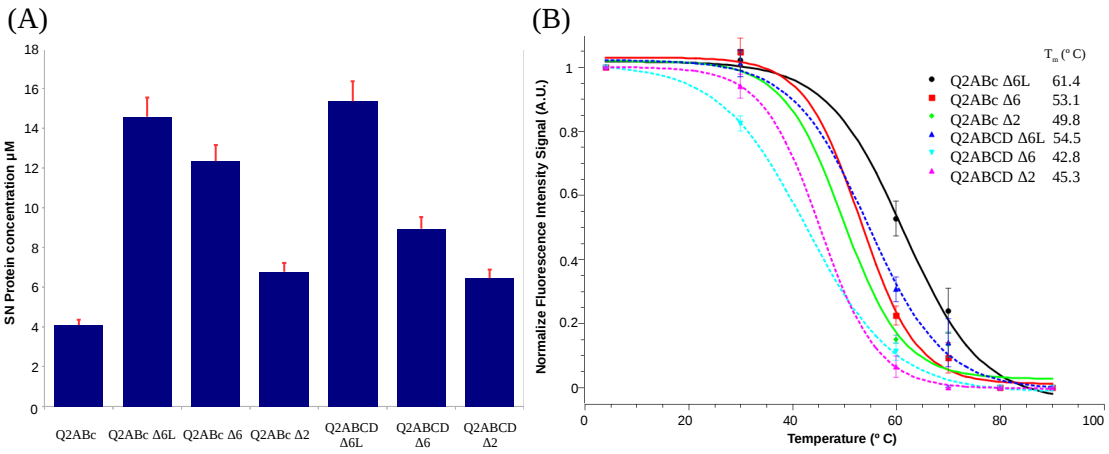


Figure 3.12. Thermostability assay for different deletions. (A) Protein concentration based on Venus absorption at 515 nm. (B) Δ2 (ΔT359_T501), Δ6 (ΔR374_K493) and Δ6L (ΔR374_T501) deletions were tested. Boltzmann equation was adjusted to the data. In thin line ABc constructs were represented and in dotted lines. All constructs were expressed fused between CaM and Venus.

As anyone would expect, ABCD constructs were less stable compared to ABc alternatives as they have more intrinsically disordered regions and flexible linkers. Regarding deletions between helices AB we found that it was Δ6L (ΔR374_T501) the one with more stability towards the heat-shock. Melting temperature (T_m) for this deletion was around 10 ºC higher than Δ6 and Δ2. For this reason Δ6L was chosen as a starting point deletion.

3.3.1.2 ΔR374_T501 deletion between helices AB is compatible with channel function in HEK293T cells

Wild-type full length $K_v7.2$ and full length $K_v7.2$ with Δ6L were cloned in pCDNA3.1 with a fluorescent tag in the C-terminus of the channel. 24 h after cell transfection electrophysiological recording were done in order to determine whether or not Δ6L was affecting channel's function. We took advantage of the presence of the fluorescent tag to select cells with similar levels of protein expression. Therefore, possible differences of the current size could not be attributed to changes in the total numbers of channels in the cell (Figure 3.13). Nonetheless, the average current density tended to be a bit larger in Δ6L but the difference was not significant. Furthermore, the impact of the deletion on the Boltzmann parameters was not significant either. Therefore Δ6L deletion results in functional $K_v7.2$ channels. For the record, it was found that there was no correlation between subunit expression based on fluorescent signal and functional electrophysiological detection.

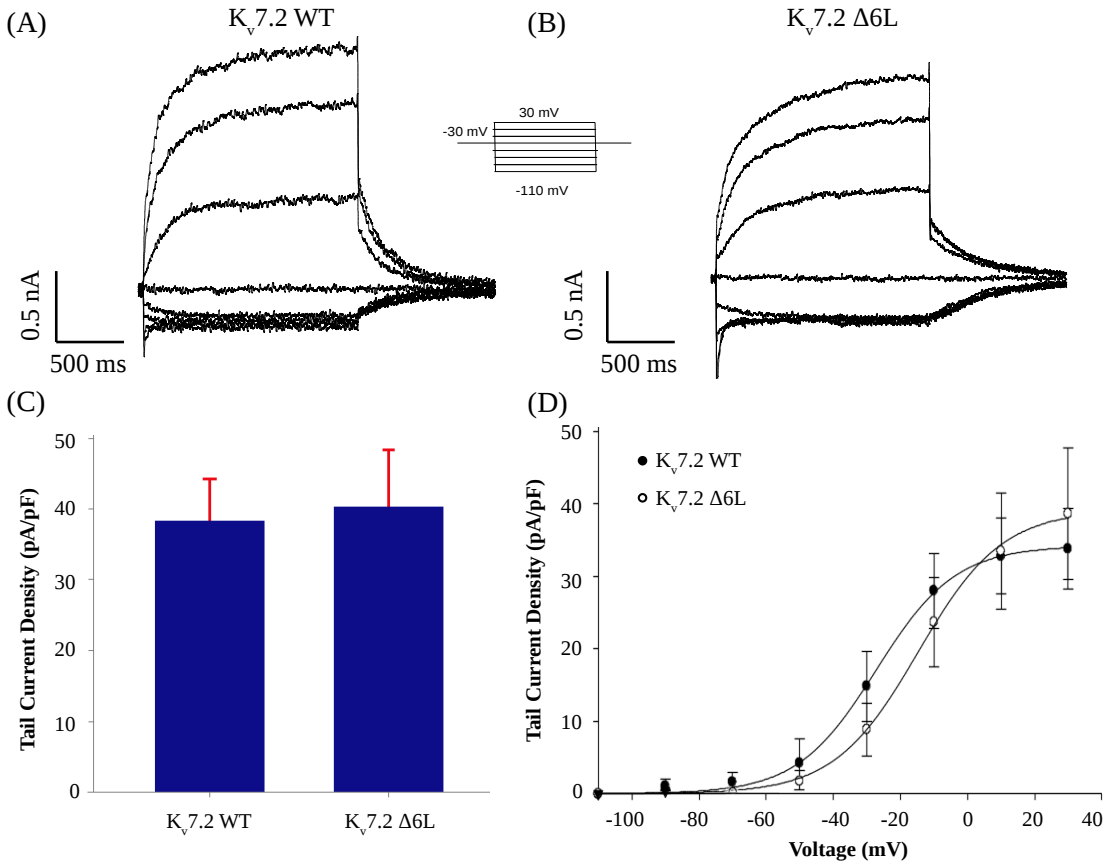


Figure 3.13. Δ6L deletion give rise to functional $K_v7.2$ channels. (A)(B) Representative currents recording after $K_v7.2$ WT and $K_v7.2$ Δ6L expression in HEK293T cells. Channels were activated from a holding potential (Vh) = -30 mV after 1500 ms steps to the indicated voltages. (C)(D) Current density-voltage relationship from tail currents of WT (n = 11) and Δ6L (n=8). Boltzmannn equation was fitted to data.

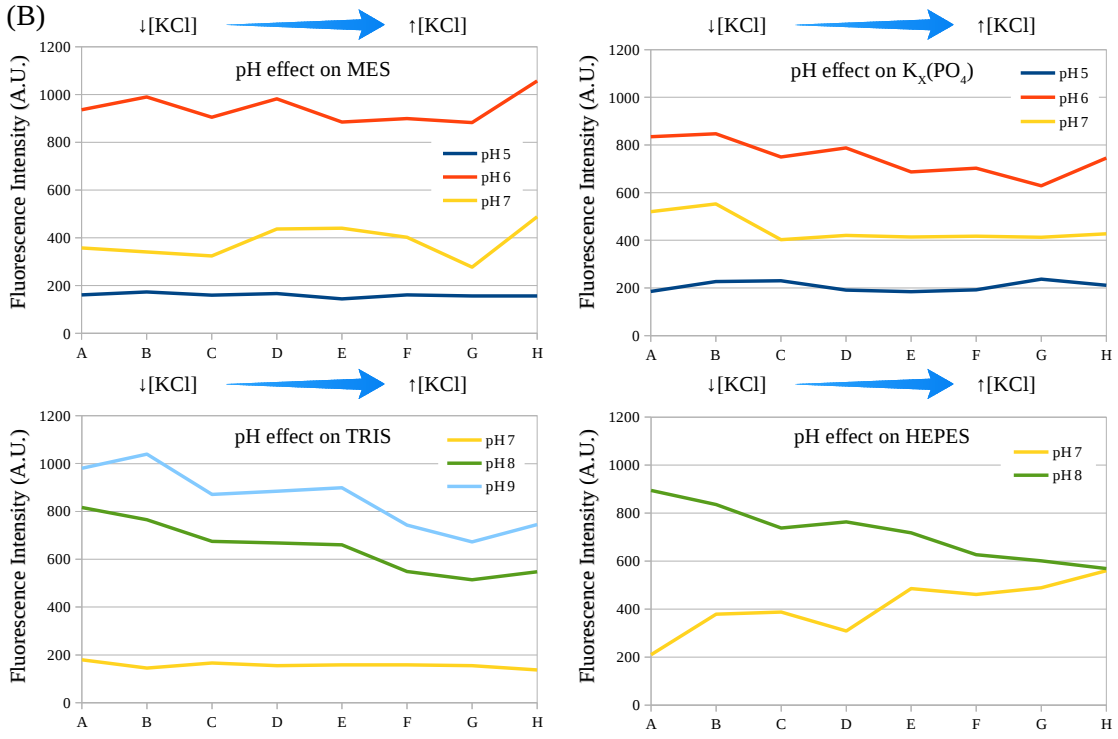
3.3.1.3 Low ionic strength, basic or acid pHs and moderate buffering agent concentrations are optimal for buffering K_v7.2 C-terminus

(A)

	MES B.			$K_x(PO_4)$ B.			TRIS B.			HEPES B.		
	pH 5	pH 6	pH 7	pH 5	pH 6	pH 7	pH 7	pH 8	pH 9	pH 7	pH 8	Controls
	1	2	3	4	5	6	7	8	9	10	11	12
A	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	25 B + 50 KCl	-
B	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	50 B + 50 KCl	*Fluo. B.*
C	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	25 B + 100 KCl	Protein complex in *Fluo B*
D	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	50 B + 100 KCl	Venus protein
E	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	25 B + 150 KCl	MES pH 6 25 + 50 KCl + AA
F	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	50 B + 150 KCl	K_x(PO4) pH 6 25 B + 50 Kcl + AA
G	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	25 B + 200 KCl	TRIS pH 9 25 B + 50 Kcl
H	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	50 B + 200 KCl	HEPES pH 8 25 B + 50 Kcl

	MES B.			$K_x(PO_4)$ B.			TRIS B.			HEPES B.		
	pH 5	pH 6	pH 7	pH 5	pH 6	pH 7	pH 7	pH 8	pH 9	pH 7	pH 8	Controls
	1	2	3	4	5	6	7	8	9	10	11	12
A	161	936	357.5	185	834.5	519.5	179.5	816	980	209.5	895	32
B	173	990	340.5	226.5	847.5	552.5	145	765	1039	379	835.5	65
C	159.5	905	323.5	229.5	749.5	402.5	166	674.5	871.5	387.5	738	292
D	167	982.5	436.5	190.5	788	420	155.5	668	884.5	309	764	591.5
E	144.5	885	440	184	686.5	413	158.5	660	899	485.5	718	785.5
F	160.5	899	402	192.5	703	417	158	548.5	743	461	626.5	610
G	156	883	277	237	628.5	412.5	155.5	514	673	489	601.5	782.5
H	156.5	1057.5	488	210.5	745	427	137	547.5	745.5	560	569	490

(B)



(C)

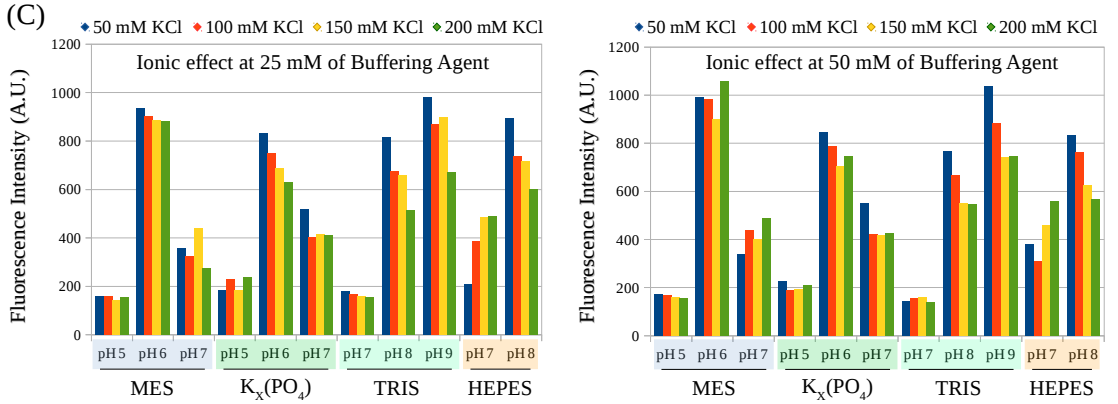


Figure 3.14. Buffer screening for K_v7.2 C-terminus with Δ6L fused to Venus and co-expressed with CaM. (A) 96-well plate buffers organization and fluorescence intensities obtained after a week incubation. (B) pH effect on protein stability depending on the buffer agent used. Higher ionic strength is affecting solubility. (D) Ionic effect either at 25 mM or 50 mM of Buffering Agent.

We decided to do a rapid buffer conditions' screening based on Venus fluorescence. In this time the construct we used was the one with four helices plus Δ6L. Moreover, for this approached CaM was co-expressed as a independent protein rather than fused to the helices. After co-expression and purification using a HiTrap column purified protein was concentrated. 96-well was prepared as in Figure 3.14.A. Afterwards protein was diluted each well and fluorescence "survival" was monitored along time.

Taking into account the **pH effect** (Figure 3.14.B), mayor decrease in fluorescence was seen at pH 7, independently of the buffering agent or ionic strength. On the contrary, basic and acidic pHs are better preservatives as more farther are from pH 7. Nevertheless, at pH 5 we observed a dramatic decrease which was related to Venus protein denaturation rather than protein complex aggregation (Campbell & Choy, 2001). The **ionic effect** was also an important factor towards protein stability (Figure 3.14.C). The output suggested that independently of the pH or Buffering Agent concentration more salt concentration was counterproductive towards the protein complex solubility. Finally the more **Buffering Agent** concentration was in the condition tested the better the stability was (Figure 3.14.C). In parallel some preservatives were tested such as 50 mM of L-Glu / L-Arg but no beneficial effects were appreciable in any of the cases.

Considering this information, we decided to work at acidic pHs (6-7) for two reasons: On one hand the protein stability was higher at this pH range, and on the other hand acidic pHs are more recommendable for NMR acquisition as amide NH sensitivity is higher at lower pHs. Regarding Buffering agent we decided to work with $K_X(PO_4)$ at 20 mM even if higher concentrations were not damaging. Salt concentration used in our experiment was 120 mM of KCl as this is the ionic strength in neuron cells and our purpose was to mimic as much as possible this conditions.

3.3.1.4 "GB116" alternative construct with deletions ΔF316_R325, Δ6L and ΔP533_H546 is the most convenient for further NMR analysis

The constructs which encloses helices A and B with ΔF316_R325, Δ6L and ΔP533_H546 deletions showed interesting characteristic to be chosen as the construct to study by NMR.

Based on the PONDR VL-XT intrinsically disorder regions predictions for Q2ABc Δ6L constructs, we designed two different deletions in the N-terminus (ΔF316_R325 and ΔF316_N334) and other two in the C-terminus (ΔP533_H546 and ΔM518_H546). We combined them and we fused Venus protein in the most distal part of the C-terminus in order to monitor protein solubility. CaM co-expression was done in this section. After pilot experiments we measured protein concentration in the SN based on Venus absorbance at 515 nm (Figure 3.15).

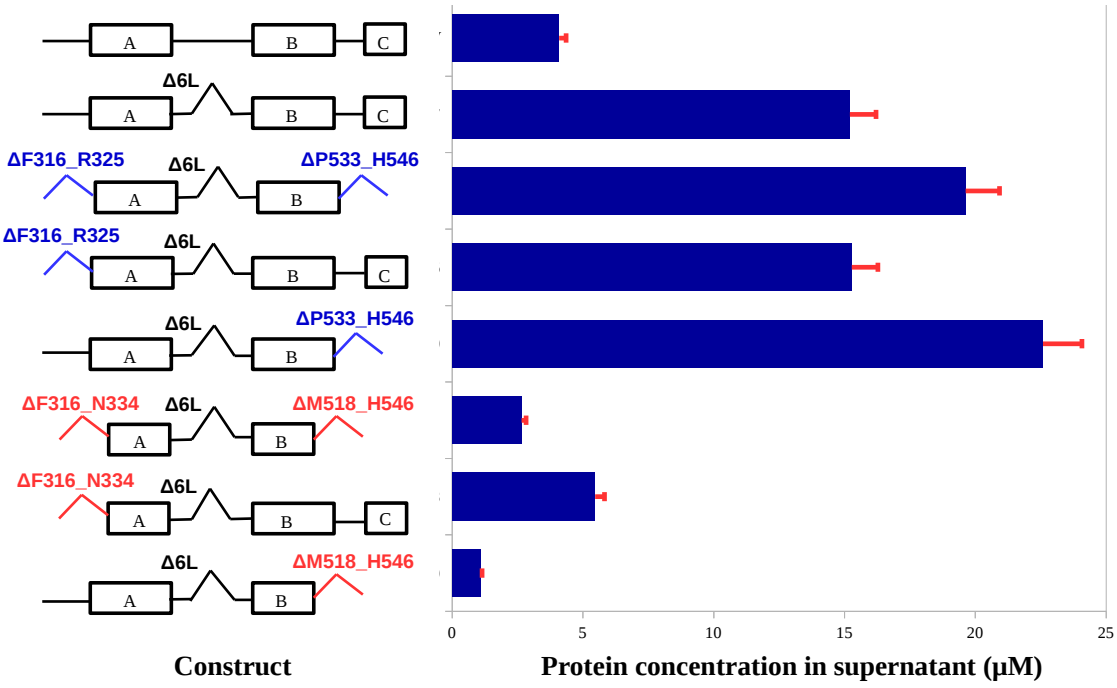


Figure 3.15. Alternative constructs' concentrations. All constructs were cloned between a His-tail in the N-terminus and Venus in the C-terminus. After protein co-expression with CaM soluble protein concentration was measured using Venus absorbance at 515 nm.

The output of the experiment was very interesting. Bigger deletions (in red) ended up in poor solubility possibly derived from the reduction of the helices' areas responsible of CaM binding. Shorter deletions (in blue) showed better solubility profiles, specially these with ΔP533_H546 deletions. In these cases the small region from the helix C was removed, which, in fact, was a very disordered region according to PONDR VL-XT. However, ΔF316_R325 deletion was decreasing a bit the protein solubility but not jeopardizing CaM interaction as its solubility is comparable to the best behaved construct.

For further analysis we decided to pick these constructs that were showing good solubility and we took out Venus protein to make sure that this protein was not interfering in the experiment's result. In order to do that His-tagged constructs were co-expressed with CaM and after running SDS-PAGE gels protein solubility percentage was calculated quantifying protein bands' intensities using ImageJ (Figure 3.16).

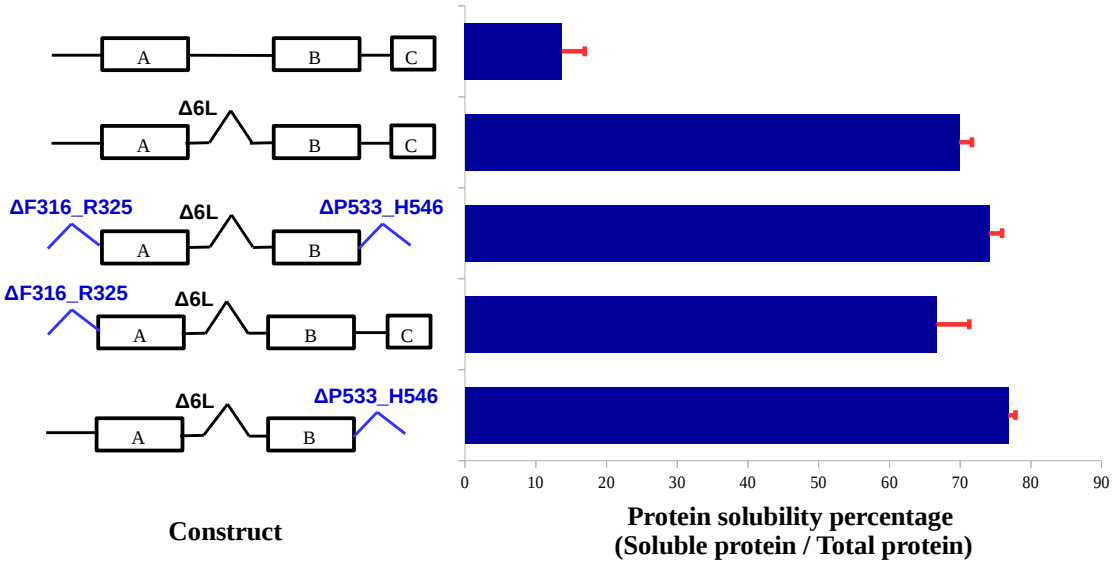


Figure 3.16. Protein solubility percentage of selected constructs. All constructs were co-expressed with CaM and relative solubility percentage was measured as previously described.

After inspecting the solubility percentage of each construct without Venus protein we can first observed that solubility tendency is proportional as the one described by Venus protein absorbance. Thus, best behaved was the alternative construct with ΔP533_H546 deletion. Nevertheless the difference among them were not statistically relevant. Therefore, in order to make chemical shift assignments easier we elect the smaller version as it has less amino-acids. This clone was named has GB116 which has helices A and B with ΔF316_R325, Δ6L and ΔP533_H546 deletions.

3.3.2 ^{15}N-HSQC Spectrum Encourages For Further Structural Analysis

GB116 and CaM were co-expressed in a ^{15}N enriched media. After complex purification it was concentrated till 240 µM in 120 mM KCl, 20 mM $K_X(PO_4)$ pH 6, 5 mM EGTA, 5 % D2O, PIC and NaN_3. Spectra was collected on an 800 MHz Bruker Avance III spectrometer at 303 K (Figure 3.17).

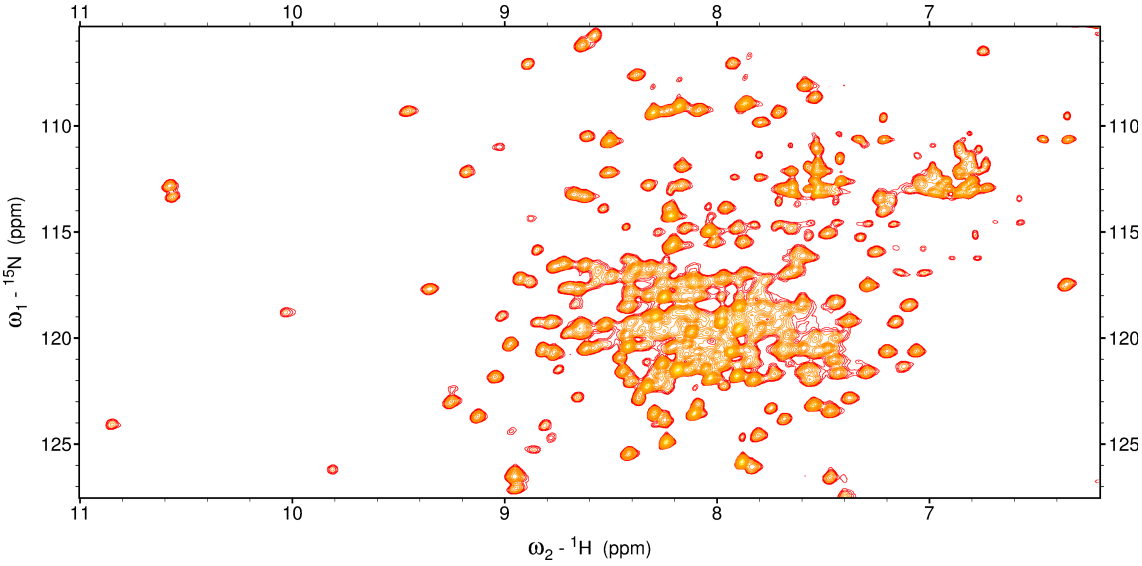


Figure 3.17. Protein complex ^{15}N-HSQC. First NMR Spectra obtained from a labelled complex. Folded signals are not shown.

The protein complex formed by CaM (149 amino-acids)(16.7 kDa) and GB116 (115 amino-acids)(13.57 kDa) has in total 264 amino acids and molecular weight 30.3 kDa which is on the edge for NMR structural determination. After doing a peak picking we detected up to 310 peaks. In the optimal conditions for the complex where the peculiarities of certain residues such as Pro, Trp, Asn and Gln are taken into account the ideal spectra of this protein should demonstrate 330 peaks. Therefore, even if some peaks are missing we could confirm that the spectra corresponds to our protein in which perhaps some labile regions are not seen in the HSQC.

Moreover, the overall shape of the spectra shows that the protein complex is generally structured as seen by the dispersity of the signals. However, some non-structure or random-coil regions are expected as we observed a crowed region in the centre of the image.

Due to this good result we proceeded to perform further NMR experiment in order to get as much structural information restrains as possible. However, been aware of the high overlapping signals in the region that would make proteins' chemical assignments tough, we decided as well to tackle the strategy of labelling proteins individually.

3.3.3 ^{15}N-HSQC Spectrum Of Individually Labelled Proteins Confirms Correct Complex Formation On-Column

Protein complexes consisting of different members are better analysed by labelling one of the subunits at a time. By this, only labelled part would be visible for NMR but would also maintain the original structure when forming the protein complex in bacteria. ^{15}N-HSQC spectrum were collected with the complex formed by labelled GB116 (^{NCD}GB116) and with the labelled CaM (^{NC}CaM). Both spectrum were superimposed with previous ^{15}N-HSQC where the whole complex was labelled (Figure 3.18).

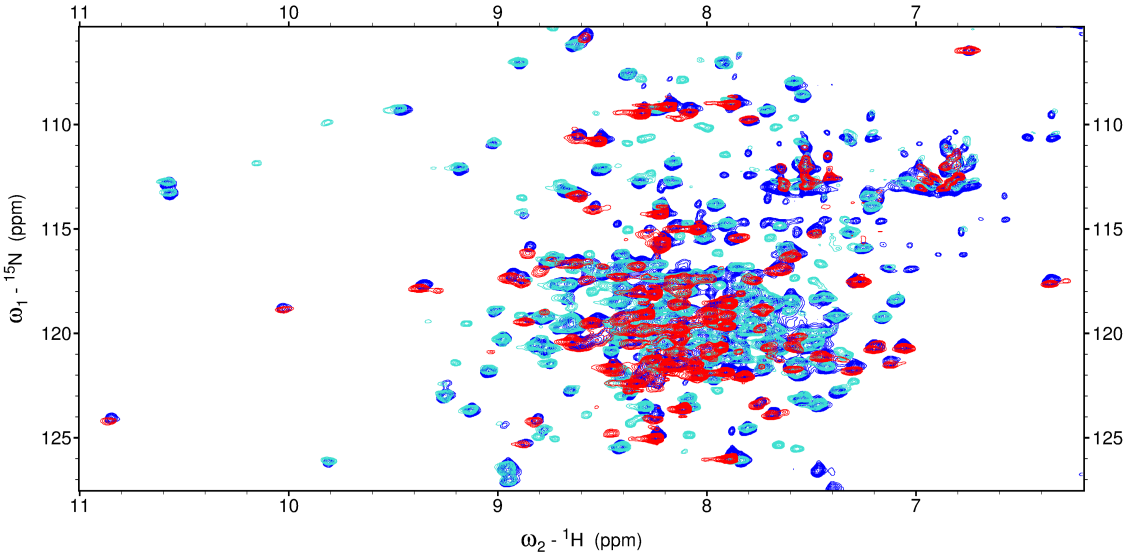


Figure 3.18. HSQC overlay of independently labelled proteins forming CaM/K$_v$7.2_AB. In dark blue both proteins are labelled. In cyan only CaM was labelled and was assembled with unlabelled GB116. In red GB116 protein is labelled and forming the complex with unlabelled CaM.

To begin with, either ^{NC}CaM complex or ^{NCD}GB116 complex were showing that protein was folded as the chemical shifts in the spectra were dispersed indicating at least for GB116 that protein refolding did work as suspected. Moreover, when superimposition of the three spectrum was done it was visible that matches between independently labelled protein with the whole labelled complex. In fact, signals between ^{NCD}GB116 complex and ^{NC}CaM complex were not overlapping meaning that this strategy could help us in the backbone assignment.

Nevertheless, specially for ^{NC}CaM complex we could also observed that besides the signals corresponding to CaM in complex with K$_v$7.2 AB, other signals arose which were not overlapping. For this sample, as we were able to confirm later, protein aggregation occurred inside the NMR tube. During this process helices AB tend to precipitate as unspecific interaction occur when CaM is not bound. As a consequence, new signals could perfectly come from free-CaM which remains soluble when helices AB precipitates. Although we would have preferred to have more homogeneous sample regarding signal sensitivity, thanks to NMR versatility we were able to use this ^{NC}CaM complex data set for further analysis.

3.3.4 PROTEIN-COMPLEX BACKBONE ASSIGNMENTS

Resonance assignment either for the whole labelled complex or for each individual components are done in a similar manner to the approaches used for single chain proteins. Thanks to the independently labelling strategy we were able to assign completely 99 residues from GB116 and partially assign 6 more residues. First 10 residues from the His-tag site of the protein were not able to be assigned (Annexed). Therefore, most of the chemical shifts from the backbone of this part of the complex were assigned. On the other hand, we were able to assign completely 145 residues from CaM and partially assign 3 more residues (Annexed). Thus almost the complete chemical shifts from the backbone were assigned.

Following information was extracted from the backbone's chemical shifts of both proteins.

3.3.4.1 Random coils detection

A simple method called the random coil index (RCI) was used to quantitatively measure protein flexibility. This approach developed by Dr. D. S. Wishart and colleges extracts information on protein motions by comparing experimental backbone's chemical shifts the random coil index. This final value can be translated to standard measures of protein flexibility such as S2, NMR-RMSD or crystallographic B factor. RCI server was used for this calculation for both proteins. Accordingly to the developers, there is also a tight relationship between RCI values and the protein dynamics suggesting that higher limits in RCI values could correspond to microseconds to milliseconds time-scales and lower RCI values to femtosecond. Nevertheless authors warn RCI users about this qualitative approach as it may depend on a big number of factors such as labelled nucleus, conformational exchange rate, molecular weight, temperature and media.

For the case of GB116 protein (Figure 3.19.A), the RCI values for the atomic displacement shows that there are three main regions with high flexibility (or dynamics): in the N-terminus from the 10^{th} to the $38t^{h}$ amino-acid, in the middle from the 73^{th} to the 88^{th} and finally in the C-terminus from the 111^{th} to the 115^{th}. Moreover, from amino-acid acids 59 to 62 also shows a small flexibility. The lack of RCI values in the graph is due to the lack of chemical shift assignments in this region. Amino acid numbering is according to the amino acid sequence of the construct and does not correspond to the amino acid numbering from $K_v7.2$ channel.

In general, for the case of CaM protein (Figure 3.19.A) RCI, and in consequence, RMSD values are smaller than GB116. Moreover, nine flexible regions are observable which perfectly overlaps with the flexible linkers that appears in CaM structure: From 1 to 5, from 19 to 24, from 39 to 44, from 55 to 60, from 73 to 81, from 92 to 98, from 110 to 115, from 127 to 136 and finally from 146 to 148 amino-acids. The amino acid counting starts with the first Ala, not with Met.

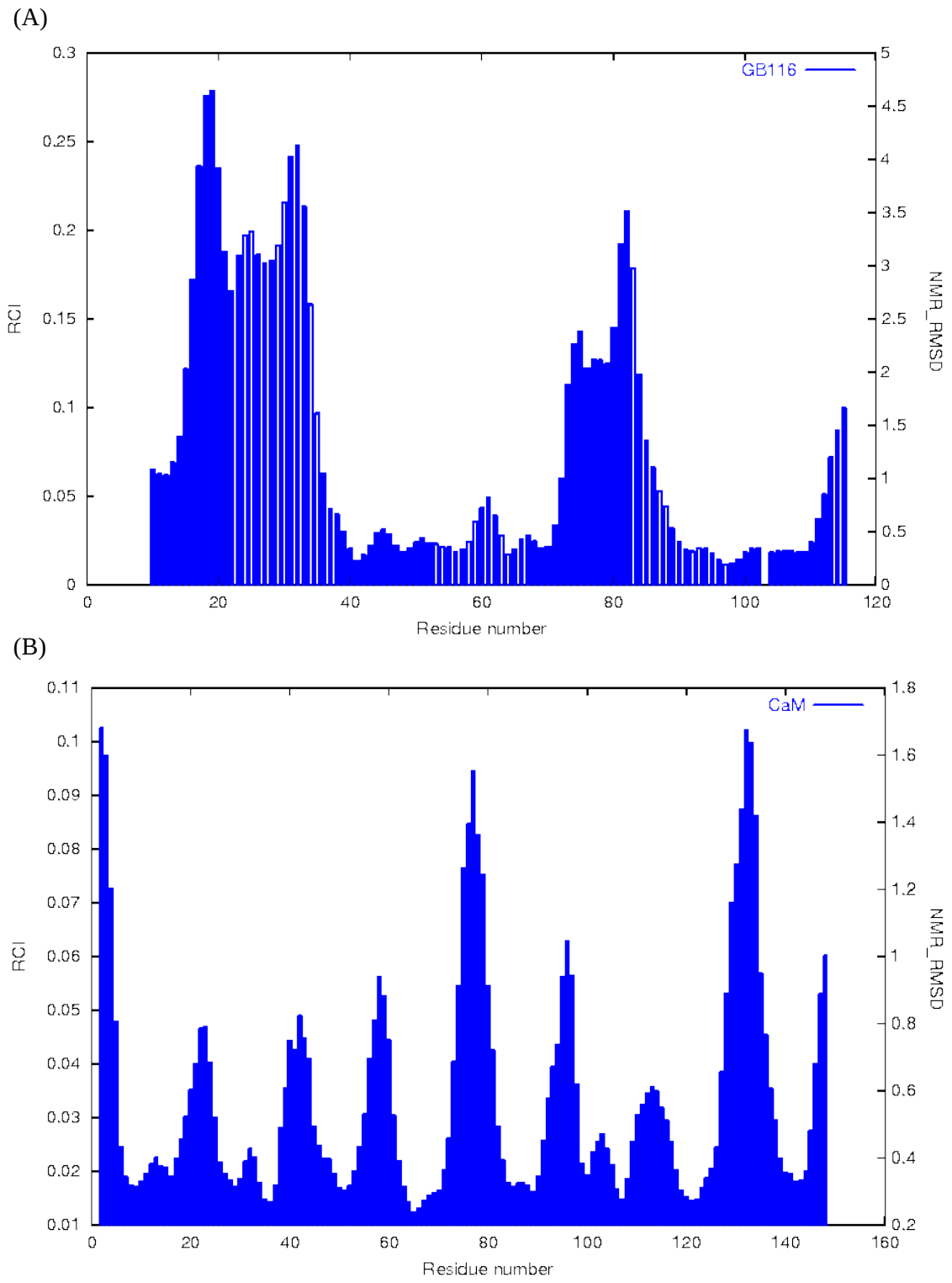


Figure 3.19. Random Coil Index (RCI) and Root Mean Square Deviation (RMSD) values for GB116 and CaM. In case of GB116 three flexible or dynamic regions are detected as well as a small semi-flexible region around the 60th amino-acid. For CaM, nine flexible or dynamic regions are observable.

3.3.4.2 Secondary structure detection

As for the flexible regions characterization, secondary structure content is defined at a residue-specific level using Random Coils Indexing. In order to do that we used the same RCI server as in the previous result. After correlating our backbone's chemical shift values for $^{1}H\alpha$, ^{13}CO, $^{13}C\alpha$ and $^{13}C\beta$ nuclei with a set of upper and lower limit cut-off values described by Dr. D. S. Wishart and colleges, we were able to identify secondary elements in the proteins under study.

Regarding GB116's secondary structure (Figure 3.20.A), we were able to observe that three alpha-helices were defined. This result was a bit shocking as at the beginning, taking into the account the information in literature and the secondary structure predictor, we were expecting just at most two helices, A and B. Consensus data from the RCI server claim that the helix A is formed by the amino acids located from the His-35 to Ala-55 of GB116 sequence, which in other words, they would be integrated of the amino acids from His-328 to Ala-348 from the original K_v7.2 WT sequence. Helix B is composed by the amino acid string from G88 to S111 in GB116 sequence, Gly-497 to Ser-531 of K_v7.2 WT sequence. The small helix in between, which from now on we will refer to is at TW Helix, is build with the amino acids from His-64 to Arg-72 in GB116 sequence or His-357 to Arg-365 of K_v7.2 WT sequence.

In CaM's case eight alpha helices were distinguished in harmony with literature and PDB information (Figure 3.20.B). These alpha helices were formed by Glu-6 to Lys-21, Thr-29 to Ser-38, Glu-45 to Asn-60, Phe-65 to Lys-75, Glu-82 to Val-91, Ala-102 to Val-108, Asp-118 to Glu127, and finally Asn-137 to Met-145. Besides alpha helices also four beta-sheets were detected from Thr-26 to Thr-28, Thr-62 to Asp-64, Tyr-99 to Ile-100 and finally Thr-110 to Leu-102. When we compared the location of these β-sheets with previously published structures in PDB we were able to recognize that the first three sheet were in agreement with CaM structures as they are forming the EF-hands where Ca2+ interacts. However, the last sheet does not agree with previous data as we would expect to appear in the next linker around Ile-130.

Nevertheless, we were very confident with the results as we knew that these secondary structure predictor have an accuracy of <90 %. Therefore, we assumed that the GB116's secondary structure was well-defined as it did make sense. For CaM's secondary structure the secondary structure definition was believable even if it had a small disagreement with the location of the last β-sheet.

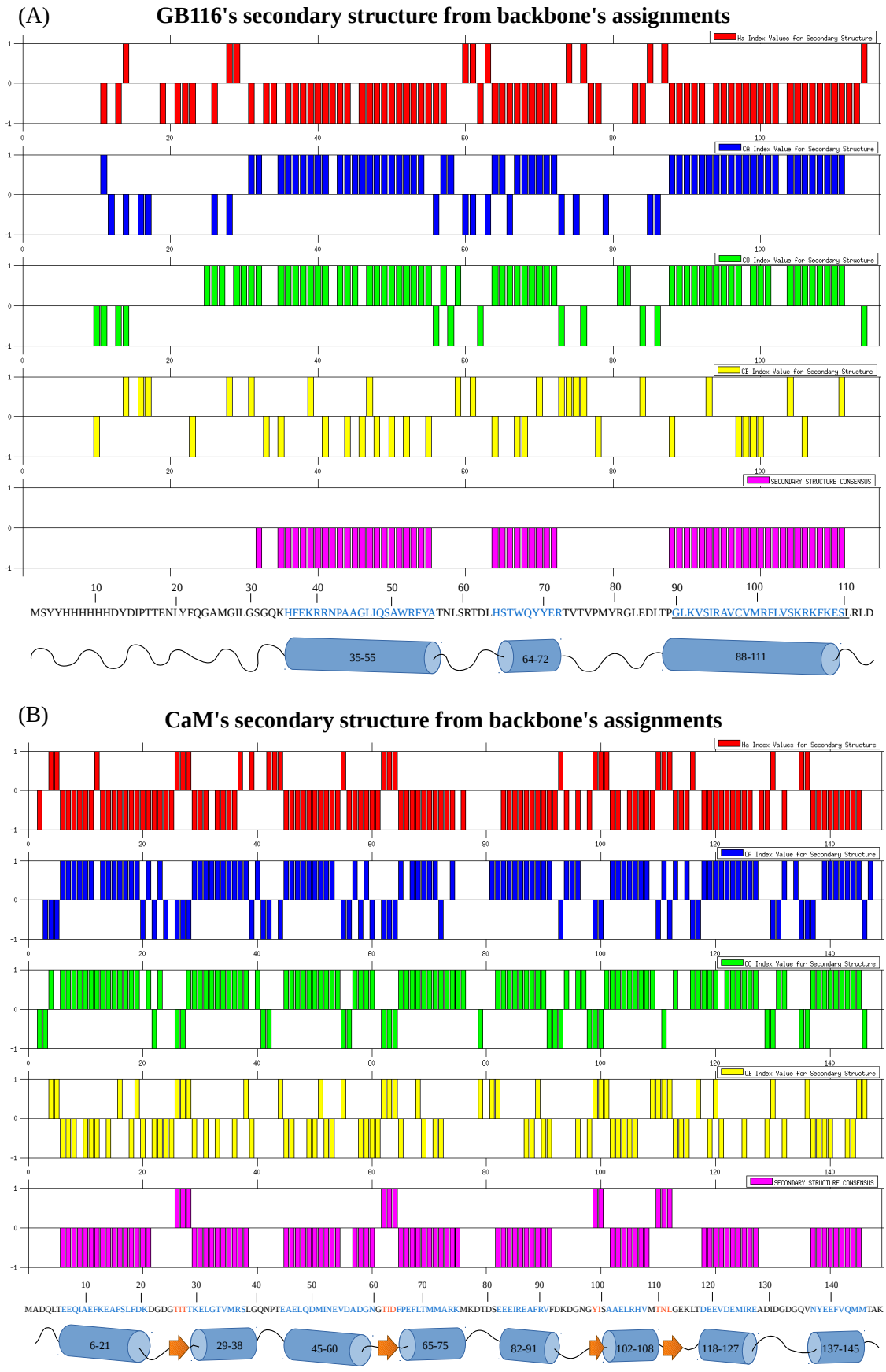


Figure 3.20. Secondary structure of GB116 and CaM. In red, Hα; in blue, Cα; in green, CO; in yellow, Cβ; and in purple, consensus values. For Cα, CO and Consensus charts positive values mean α-helix and negative values β sheet propensities. For Hα and Cβ is the opposite. Cβ is only useful for β-sheet prediction.

3.3.4.3 Dihedral angles prediction

Torsion angle restrains are perhaps one of the most useful restrains for the NMR determination. Therefore, the introduction of these restrains for the final structure calculations are really advisable even if not required. In our case dihedral angles were predicted using backbone's assignments and TALOS+, which generates useful dihedral angles restrains (Figure 3.21).

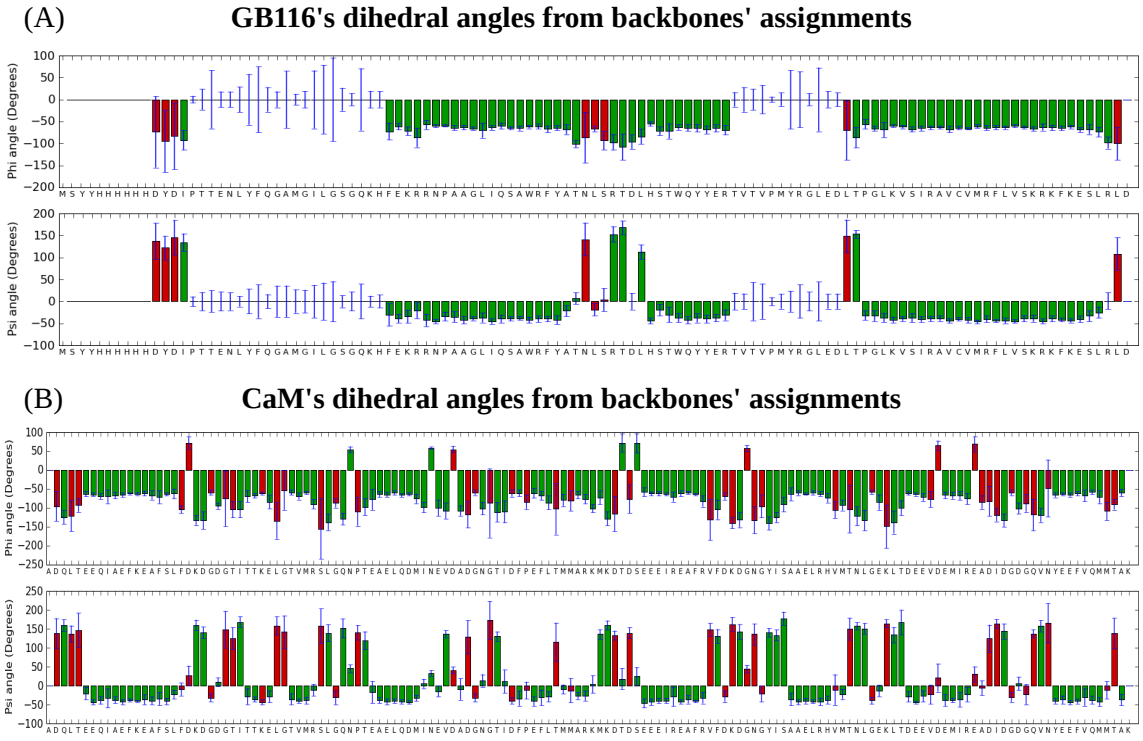


Figure 3.21. Phi (Φ) and Psi (Ψ) torsion angles restrains for GB116 and CaM. In this graphs are represented the phi and psi angles for the residues and the uncertainty of the prediction. In green, accurate predictions are symbolised and in red unreliable predictions. If the bar is absent, residue could be dynamic or prediction was made due to lack of information.

In theory, the average phi and psi values for alpha-helices should be around -57 and -47 degrees. On the contrary, beta-sheets should have phi and psi values around -80 and +150. If we have a look in GB116's prediction (Figure 3.21.A) we can rapidly observe that the general shape of the graphic adds up with the previous data extracted from the chemical shifts. Phi and psi values describe three alpha helices in the same amino-acid positions as in the secondary structure definition by RCI values. There are two highly dynamic regions, one in the N-terminus and another one in between TW and B helices. Moreover, it seems that it could be some kind of β-sheet in the middle of Helices A and TW. Seven torsion angles were not reliable and others were not visible as chemical shifts were missing.

In case of CaM the predicted torsion angles looks different from the GB116's one as it has β-sheets in some parts of the structure. Nevertheless there are much more not reliable torsion angles which could be due to chemical shifts changes due to the complex formation.

3.3.5 RDCs Measurements shows Relative Orientations of Complex forming Proteins

Residual dipolar coupling have proven to be particularly useful to define the orientation of intra- or inter-protein elements such as helices or domains.

RDCs were measured for ^{NCD}Complex and ^{NCD}GB116 proteins. For each of them isotropic TROSY, isotropic semi-TROSY, anisotropic TROSY and anisotropic semi-TROSY experiments were collected. In order to create an anisotropic media we chose Otting nonionic liquid crystalline media. RDCs were measured calculating the difference between the total couplings detected by the anisotropic media and the J coupling from the isotropic media. (Appendix 3.3)

Axial component of the alignment tensor (Da) and the Rhombicity (R) of the tensor were determined *de novo* plotting a histogram with all the RDCs. Calculated values were -18.9 for Da, -18.9 and 0.2218 for R.

Once we had the Da and R values we used Module program to represent the relative orientations of the different components of the protein complex. In order to do that, we overlayed the RDCs information on a modelled version of K_v7.4 helices AB structure (PDB: 4UMO) with the sequence of K_v7.2 channel. It should be taken into account that RDCs provide information about the relative orientation of the vectors, but do not provide information regarding the location in the space of each element. Therefore, the only thing that should be considered in Figure 3.22 is the orientation, nothing else. The location has been deduce without evidences in agreement with the crystal structure.

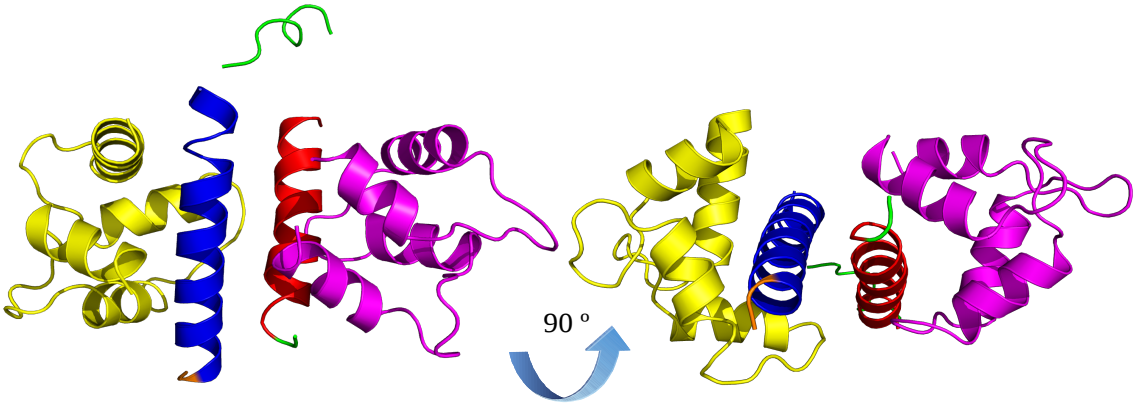


Figure 3.22. Relative orientation of protein elements forming the complex. In yellow, N-lobe and in purple, the C-lobe of CaM. In blue, helix B and in red helix A. In cyan, TW helix. Within the Helix A a green amino acid has been represented to represent the N-terminus and within Helix B a orange amino acid has been represented to indicate the C-terminus.

According to our RDC data helices A and B could be orientated in anti-parallel formation. CaM's Lobules also are could be orientated a closed conformation wrapping helices AB. But once again, this representation is based in crystal structure and the location in the space of each element is unknown.

3.3.6 GB116 Protein Structure Confirms an anti-parallel Coiled-Coil Conformation of Helices AB

After side-chain chemical shifts assignments, short- and media-range NOE assignments were done manually for ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC. Afterwards an automatic peak picking was done using Sparky program in order to create a list of ambiguous restrains.

ARIA2 program was used to create an atomic model using Torsion angles restrains and back-bone's chemical shifts. Moreover, NOE restrains were added in which an automatic NOE assignment using the ambiguous NOE list. Besides, RDCs were loaded in the calculation too.

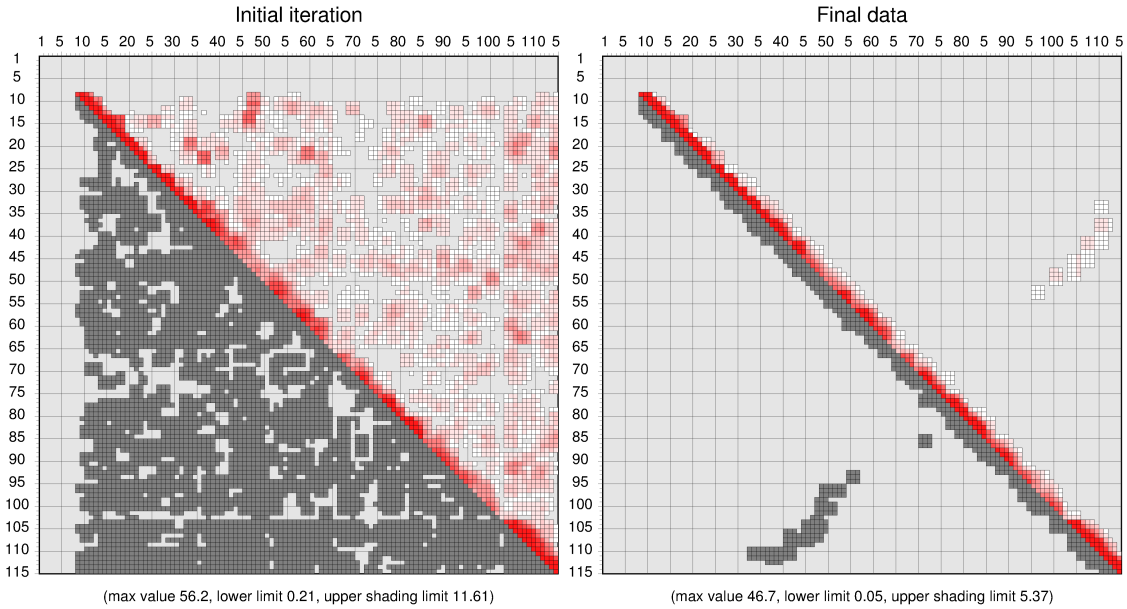


Figure 3.23. K_v7.2_AB contact map. All the possible contacts in at the beginning of the structure calculation (left) are refined to the final contact map (right).

From the first iteration in which a huge number of contacts were possible by ambiguous NOE list, only some of them survived the structural refinement. As expected, the short- and media-range NOE contacts are described in the diagonal of Figure 3.23 but the more interesting contacts are these out of the diagonal. By the shape of the graph we can observed that there are inter-helix contacts between helices A and B. Moreover, due to the contacts type we can conclude that the orientation of the helices is anti-parallel. The list of long-range restrains are listed is annexed in digital format (File 03_LongRange.txt).

When we take into consideration the final structure bundles we can observe that there is an agreement in all the structures. These protein models represents the best structures that fit experimental data as best as possible (Figure 3.24). Helices A and B are located in anti-parallel an canonical coiled-coil (Appendix 3.5) mode in which a and d positions are occupied with hydrophobic residues. The way in which both helices are bound is due to series of Van der Waals contacts as shown in Figure 3.31,

However, TW helix is moving around as it doesn't make any solid contact with other sites of the protein. Nevertheless, this does not mean that the TW site is moving in the final protein complex as it may have an interaction site with CaM. When we take the best structure, we see clearly how both helices are in anti-parallel coiled-coil conformation.

Apart from the visible His-tag tail in the structure, we can also observe that the GILGSGQ sequence, which is conserved among all the K_v7 family members. It appears to be flexible. This region has been proposed to be the connexion henge from the transmembrane domain of the channels and the C-terminal domain (Sihn et al., 2015). Therefore, all the intracellular signalling that take place in the C-terminus may be conducted through this henge to the channel pore, regulating its gating.

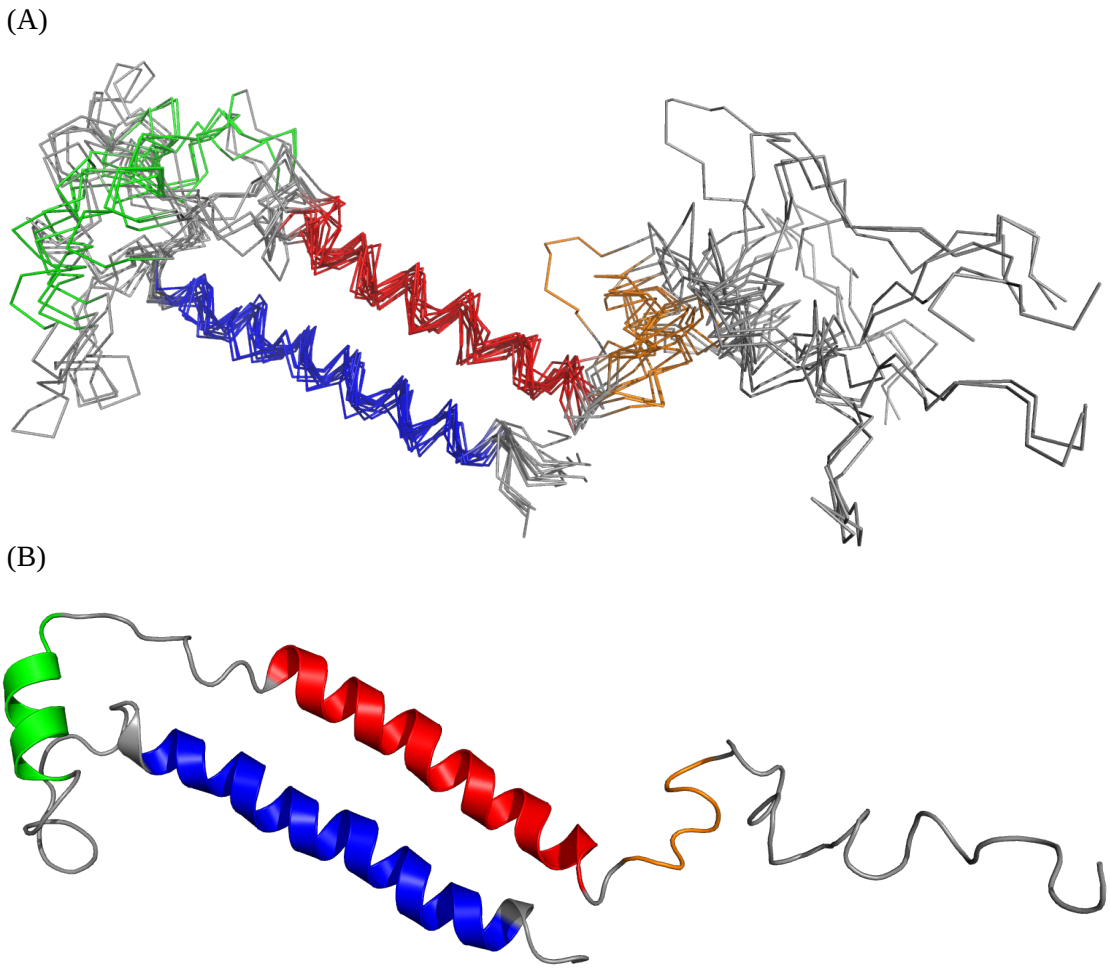
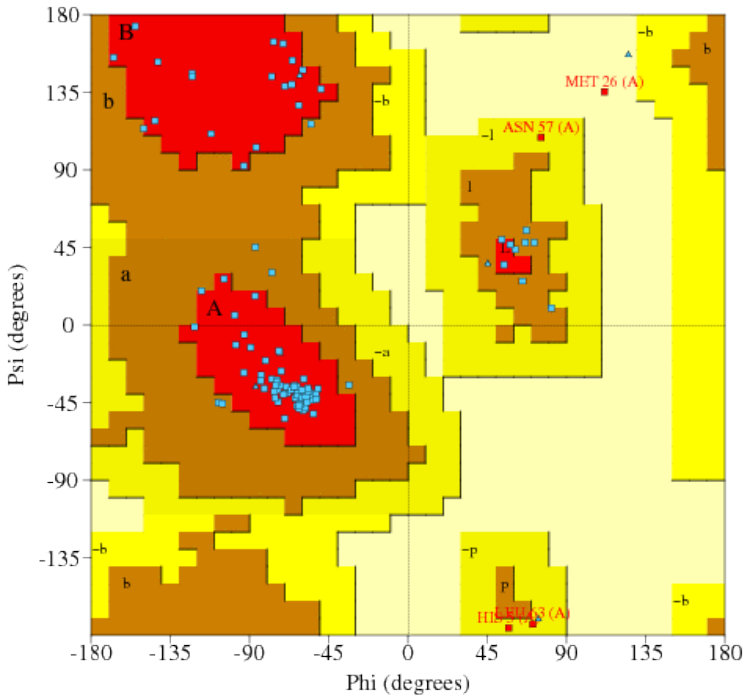


Figure 3.24. K$_v$7.2_AB structural model. (A) Protein bundles in which eight proteins are represented. (B) Best structural model that fits better all the experimental data. In blue, helix A; in green, TW helix; in red, helix B; and in orange, GILGSGQ region which is connected to the pore. (B) Overall structure of helices AB. Some torsions are found in the structure probable due to the absence of CaM in the calculation.

PROCHECK program was used to check the reliability of best protein structure (Figure 3.24). In general, Ramachandran plot reveals that the 96.1 % of the residues (98 residues) are in allowed regions. Moreover, there are 3 residues (2.9 %) that appear in generously allowed regions. However, there is one residues that appear in disallowed regions which are Met 26. When we compared the Ramachandran plots of the rest of the models we realized that the outsiders were varying for every model. In fact, these residues are located in the flexible parts that are not, in principle, trivial for the helices A and B formation. Therefore, we considered the best protein structure as a good quality atomic model.



		No. of residues	%-tage
Most favoured regions	[A,B,L]	86	84.3%*
Additional allowed regions	[a,b,l,p]	12	11.8%
Generously allowed regions	[~a,~b,~l,~p]	3	2.9%
Disallowed regions	[XX]	1	1.0%*
Non-glycine and non-proline residues		102	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		7	
Proline residues		4	
Total number of residues		115	

Figure 3.25. Ramachandran plot of best model for helices AB. 84.3 % of de residues (86 amino acids) are located in the most favoured regions. 11.8 % of the residues (12 amino acids) are located in additionally allowed regions. 2.9 percent of the residues (3 amino acids) are located in generously allowed regions. Finally, 1 % of the residues (1 amino acids) are in disallowed regions. Remaining residues are Gly (triangles) and Pro (not shown).

3.3.7 CALMODULIN STRUCTURE SHOWS A SEMI-OPEN CONFORMATION

Same procedure was repeated for CaM too, where NOE assignments were done manually for ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC. Afterwards an automatic peak picking was done using Sparky program in order to create a list of ambiguous restrains.

ARIA2 program was used to create an atomic model using torsion angles, NOE and RDC restrains. In this case, contact map differs from the previous structure as the intermolecular contacts follow a different pattern. Basically, contacts are visible within each lobe, specially between the loops forming each EF-hands (Figure 3.26)

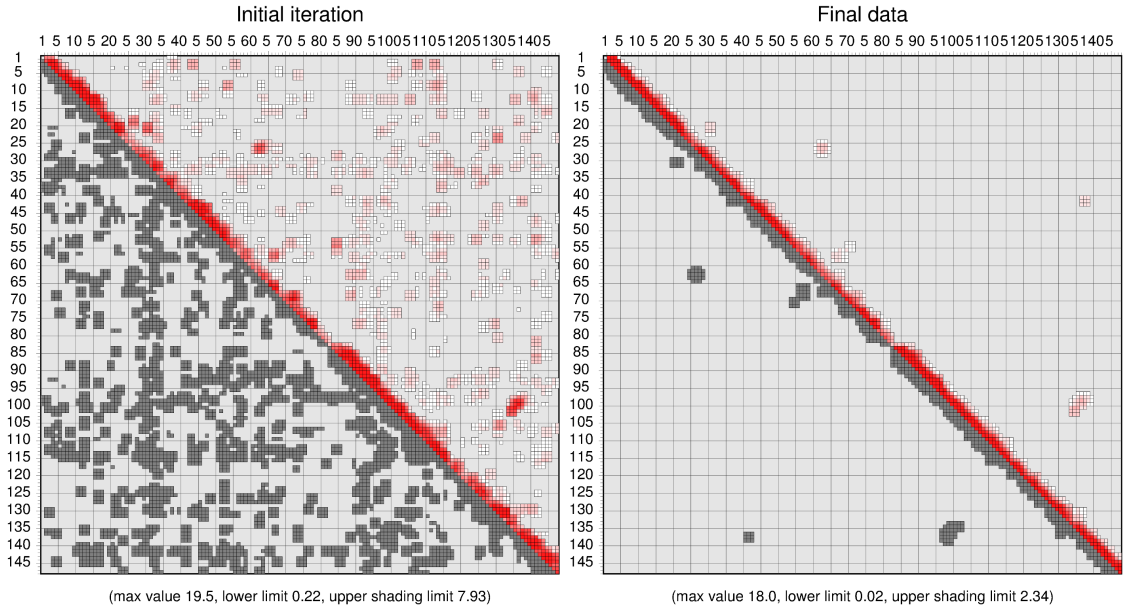


Figure 3.26. CaM contact map. All the possible contacts in at the beginning of the structure calculation (left) are refined to the final contact map (right).

When we take into consideration the final structure bundles we can observe that at the first sight is seems that CaM structure is not well-defined. When the bundles are aligned the image is not clear. However, it is easy to determine both lobes (Figure 3.27.A). Nevertheless, when aligning all the structures for one of the lobes the representation becomes much clearer as the structure differences are due to the flexible linker of calmodulin (Figure 3.27.B and C). The big flexibility of CaM is produced by the lack of restrains between helices A and B and CaM. Among all the models, the best one representing the maximal restrains has a closer conformation (Figure 3.27.D). In fact, this structure is a perfect candidate for protein docking as it has a semi-open conformation where a pocket formed in which helices AB could be wrapped.

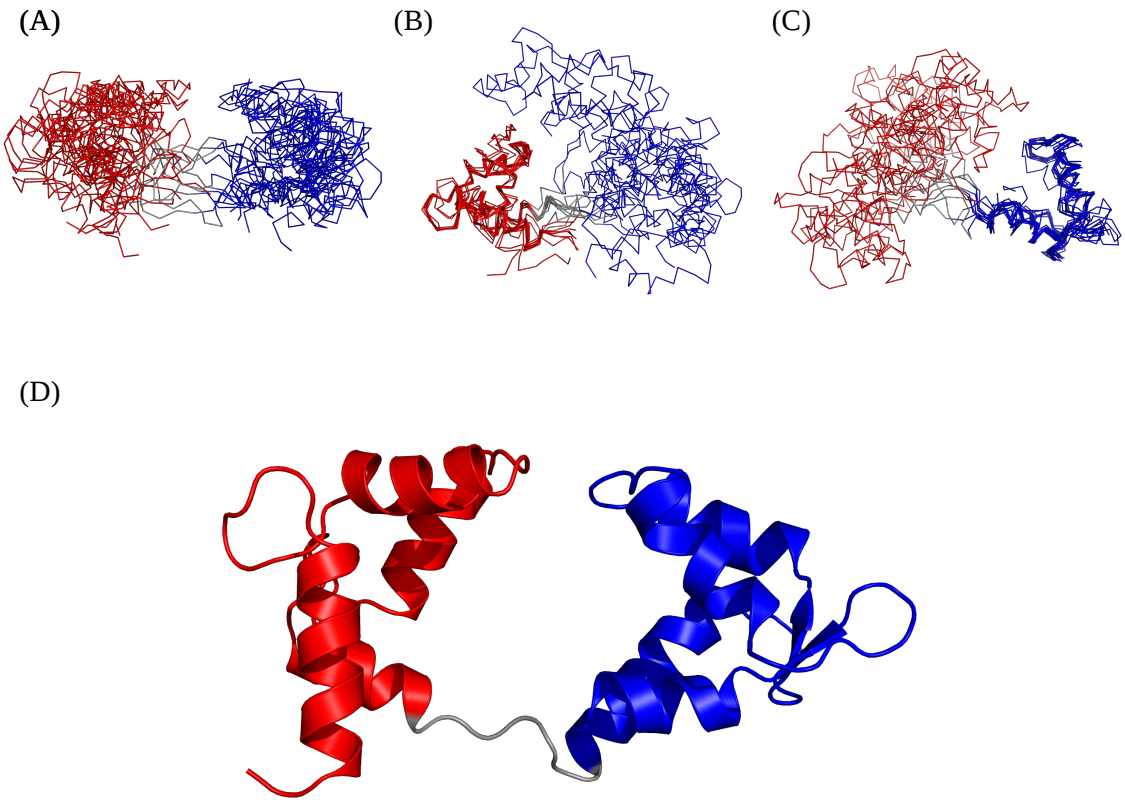
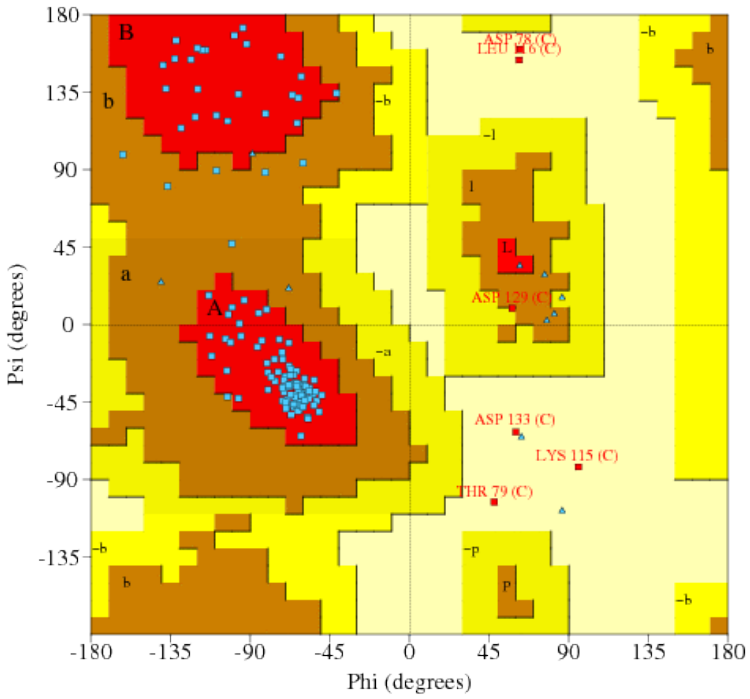


Figure 3.27. CaM structural model. (A) Alignment of the best models. (B) Alignment of the N-lobe and (C) C-lobe within the best structural models. Structural flexibility is due the linker between N- and C-lobe. (D) CaM structure of the model that fits more restrains. The possible interaction site could be placed in the pocked of the centre.

PROCHECK program was used to check the reliability of best CaM structure (Figure 3.28). In this case the result was really good, as the Ramachandran plot reveals that the 95.4 % of the residues (124 residues) are in allowed regions. Moreover, there in one residue that appears in generously allowed regions (0.8 %). However, there are 5 residues that appear in disallowed regions which are Asp 113, Lys 115, Thr 79, Asp 129 Leu 116 and Asp 78. Nevertheless, it was considered that this model was reliable as it represents the best the restrains.



		No. of residues	%-tage
Most favoured regions	[A,B,L]	117	90.0%*
Additional allowed regions	[a,b,l,p]	7	5.4%
Generously allowed regions	[~a,~b,~l,~p]	1	0.8%
Disallowed regions	[XX]	5	3.8%*
Non-glycine and non-proline residues		130	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		11	
Proline residues		2	
Total number of residues		145	

Figure 3.28. Ramachandran plot of best model for CaM. 90 % of de residues (117 amino acids) are located in the most favoured regions. 5.4 % of the residues (7 amino acids) are located in additionally allowed regions. 0.8 percent of the residues (1 amino acid) are located in generously allowed regions. Finally, 3.8 % of the residues (5 amino acids) are in disallowed regions. Remaining residues are Gly (triangles) and Pro (not shown).

3.3.8 Final Complex Formation: Protein Docking

After the structural definition of both proteins HADDOCK2.2 program was used to perform the protein docking. However, several complications arose when performing the docking.

1- CaM's conformation was semi-open which does not allow easily the docking as the interaction site is hidden.

2- CaM and GB116 have two interaction sites. Therefore, protein docking for both binding sites were done at the same time.

3- The N-terminus of GB116 was really big and besides it wasn't involved in the binding. In fact, it was disturbing the protein docking along the interaction sites.

For all these reasons protein docking parameters were changed so as to get more complexes were CaM was wrapping GB116. Basically, the force constant for centre of mass contact restraint was elevated to 2. Moreover, NOE long-range restrains and RDCs were loaded to the server to make the protein docking more robust.

Final docking represents a protein complex were CaM is hugging helices A and B (Figure 3.29). As predicted, Helix A in bound to N-lobe and Helix B is bound to C-lobe. Helices A and B are in anti-parallel conformation and the TW site is not making contact which CaM.

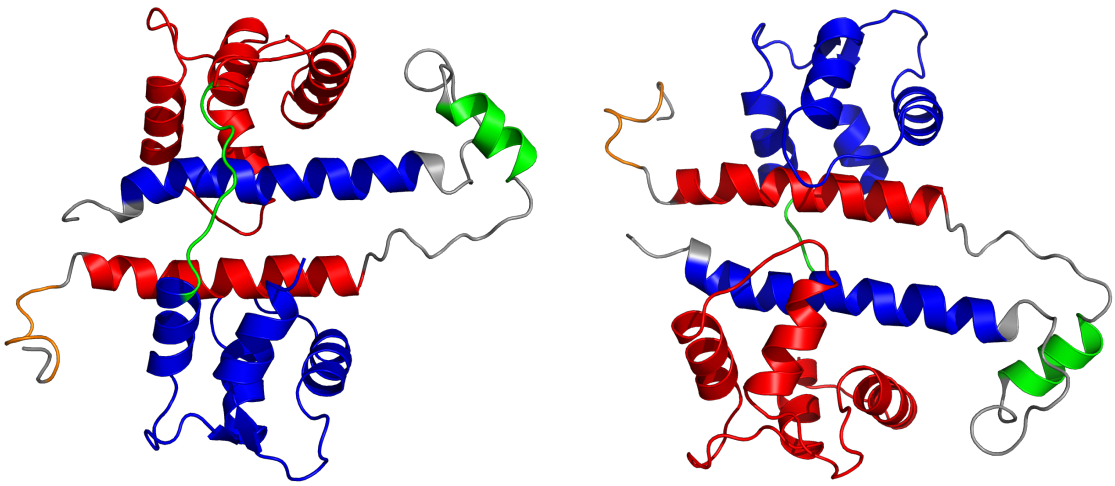


Figure 3.29. Structure of CaM/K_v7.2_AB complex. After protein docking using HADDOCK, CaM is wrapping helices A and B. N-lobe of CaM is in red, and the C-lobe in blue. Flexible linker is represented in dark green. For K_v7.2, helix A is represented in red, and helix B in blue. TW helix is represented in green. Besides GILGSGQ region has been coloured in orange.

3.4 DISCUSSION

In this chapter we have represented the structural insides of CaM binding to the C-terminus of K_v7.2 channels. This atomic model resolved by NMR provide us for a valuable information of of how this interaction occurs, which amino acids are involved and what kind of forces drive this protein complex formation.

To begin with, we have proven that the Δ6L (ΔR374_T501) made to our construct in order to be more soluble gives functional K_v7.2 channels. Wild type and mutated K_v7.2 channels have similar current intensities which suggest that the linker between helices A and B is not required for the channel function, assembly or CaM binding. Thus, we corroborate what previous studies in the laboratory and other authors have observed before when deleting this inter-helix loop (Aivar et al., 2012; Haitin et al., 2009). Nevertheless, the reason why this particular deletion presented a better thermostability profile compared to Δ2 (ΔT359_T501) and Δ6 (ΔR374_K493) remains unknown. Regarding Δ2 (ΔT359_T501), our group has recently reported that this deletion is disrupting an unconventional calmodulin anchoring site that stabilizes CaM binding to the C-terminus of the channel (Gomis-Perez et al., 2015). In the present thesis this region is represented by the helix TW. Therefore, it comes into an agreement that interrupting this small helix creates a complex destabilization in the thermostability assay. On the other hand, the only eight amino acid are different between Δ6 (ΔR374_K493) and Δ6L (ΔR374_T501): Ser 494, Cys 495, Pro 496, Cys 497, Glu 498, Phe 499, Val 500 and Thr 501. After aligning human K_v7 channels' sequences we observed that Cys 495 and 497 are only conserved in K_v7.5 channel, indicating that this region has had a big mutational propensity. Thereby, the thermostability improvement after the removal of these amino-acids could be more related to unspecific interactions rather than CaM binding effects.

In regard to the complex structure, here we have shown for the first time a monomeric protein complex in which K_v7.2 helices A and B are in an antiparallel conformation and are wrapped by CaM. Moreover, we also have resolved the pre-helixA region in which the mechanical transduction may occur when for the protein signalling that takes places in the C-terminus.

When we search for other related structures in the protein data bank, we rapidly find that the presented complex is very similar to other K_v7.1 protein complex structures (Figure 3.30) published (Sachyani et al., 2014) (PDB: 4UMO and 4V0C). In these cases, two CaM proteins are binding helices A and B from different subunits creating a dimer with a stoichiometry 2:2. In other words, helix A of one molecule is making contacts with the helix B of another molecule and CaM is binding both at the same time. Nevertheless, authors confirm that this dimer formation occurred during the crystallization and that the relevant physiological protein complex should be the one composed by a monomeric complex in which helices A and B from

the same molecule are been wrapped by CaM. In this regard, our CaM/K_v7.2 _AB complex confirms this hypothesis.

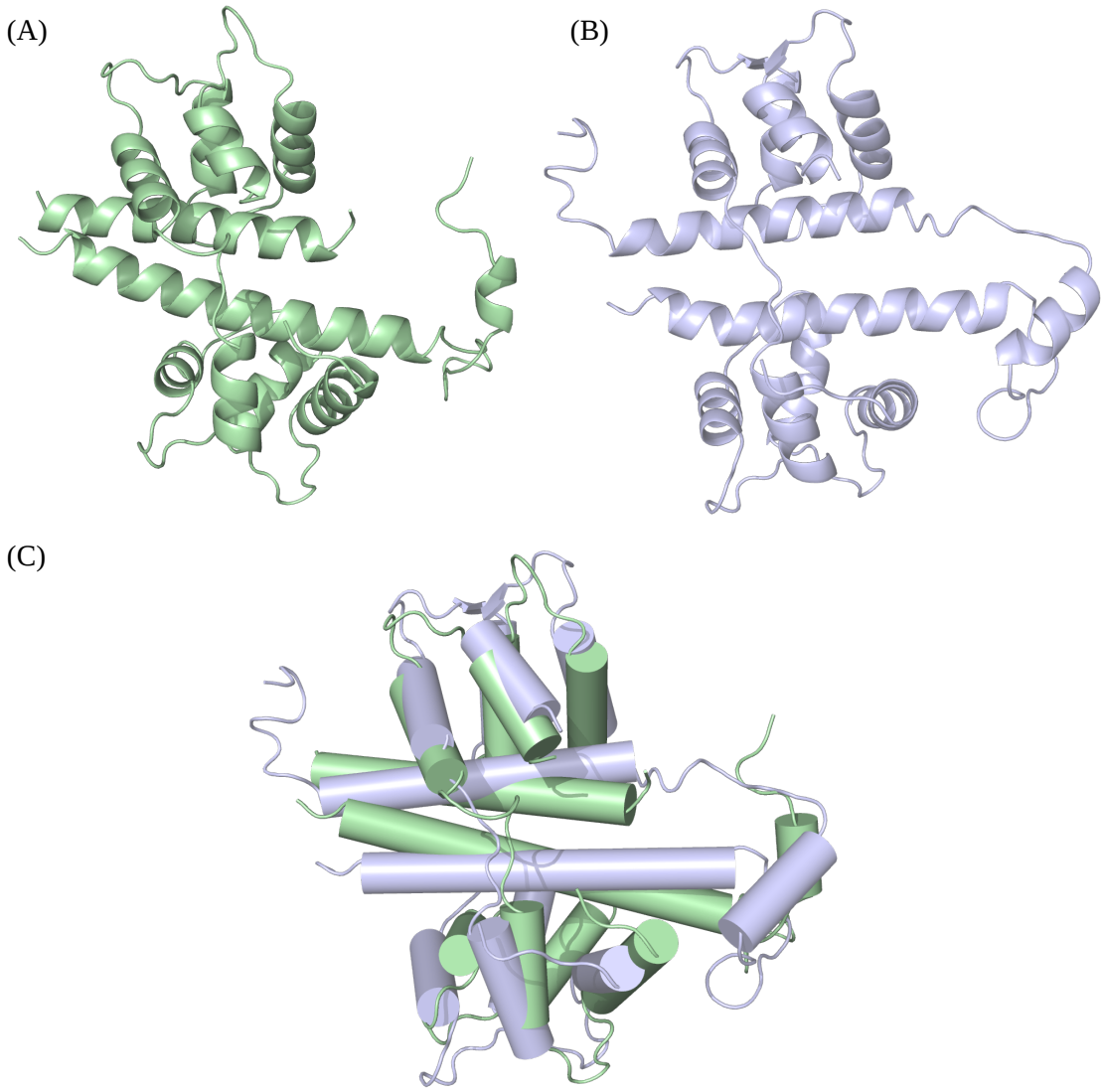


Figure 3.30. Overlay of CaM/K_v7.2_AB complex and CaM/K_v7.1_AB complex. (A) Cartoon representation of CaM/K_v7.1_AB complex (PDB 4UMO). (B) Cartoon representation of CaM/K_v7.2_AB complex characterize by NMR. (C) Cylindrical representation of helices where a big a small rotation is observable, but the overall structures are very similar.

CaM similarities between the NMR and the crystal structure are really high. The respective interactions of each CaM's lobules and their corresponding channel's helices are almost the same. For instance, N-lobe is binding helix B through the hydrophobic pocket created by residues 88 to 92 and by 109 to 115. The C-lobe also interacts with helix A by the hydrophobic cluster created by residues 18 and 19.

As mentioned before, either $K_v7.2$ or $K_v7.1$ structures shows an anti-parallel coiled-coil conformation for helices A and B. When we overlap both structures we can see that they perfectly match. These inter-helix interactions are driven by Van der Walls forces between the residues listed in Figure 3.31. Nevertheless it has been also found a polar contact between Gln 48 (341 in $K_v7.2$ sequence) and Arg 100 (519 in $K_v7.2$ sequence).

$K_v7.2$_AB

#	A.A.	Position
330	GLU	f
331	LYS	g
332	ARG	a
333	ARG	b
334	ASN	c
335	PRO	d
336	ALA	e
337	ALA	f
338	GLY	g
339	LEU	a
340	ILE	b
341	GLN	c
342	SER	d
343	ALA	e
344	TRP	f
345	ARG	g
346	PHE	a
347	TYR	b
348	ALA	c
349	THR	d
350	ASN	e
351	LEU	f
352	SER	g

Helix A

#	A.A.	Position
532	ARG	f
531	LEU	e
530	SER	d
529	GLU	c
528	LYS	b
527	PHE	a
526	LYS	g
525	ARG	f
524	LYS	e
523	SER	d
522	VAL	c
521	LEU	b
520	PHE	a
519	ARG	g
518	MET	f
517	VAL	e
516	CYS	d
515	VAL	c
514	ALA	b
513	ARG	a
512	ILE	g
511	SER	f

Helix B

$K_v7.1$_AB

#	A.A.	Position
369	PRO	c
370	ALA	d
371	ALA	e
372	ALA	f
373	SER	g
374	LEU	a
375	ILE	b
376	GLN	c
377	THR	d
378	ALA	e
379	TRP	f
380	ARG	g
381	CYS	a
382	TYR	b
383	ALA	c
384	ALA	d
385	GLU	e

Helix A

#	A.A.	Position
531	GLN	c
530	GLN	b
529	PHE	a
528	LYS	g
527	LYS	f
526	LYS	e
525	ALA	d
524	VAL	c
523	PHE	b
522	TYR	a
521	GLN	g
520	MET	f
519	ARG	e
518	ARG	d
517	ILE	c
516	VAL	b
515	LYS	a
514	ILE	g

Helix B

Figure 3.31. Diagram of inter-helical interaction in the anti-parallel coiled coil helices. The hydrophobic contacts of $K_v7.2$_AB are very similar to $K_v7.1$_AB structure.

Due to the NMR limitations, *a priori* we are not able to concrete whether or not the calmodulin in the complex has calcium or not. Due to the presence of EGTA in the sample and the phosphate buffer one would expect not to have calcium in any of the lobes. Nevertheless, more experiments were needed to specify the the calcification level of the CaM as we will see in the next chapter.

4 THE CALCIUM EFFECT

4.1 INTRODUCTION

4.1.1 PROTEIN COMPLEXES BEYOND STATIC REPRESENTATIONS

Atomic structural studies have classically provided a static representation of proteins which facilitates the resolution. Nevertheless, there are not static proteins in life as they are continuously changing their conformation into an equilibrating set of time-dependent structures. In facts, proteins in the intracellular environment are constantly making physical contacts with other ligands, proteins, or macromolecular complexes. Basically, all those recognitions are mediated by proteins' atomic thermal fluctuations or segmental motions. Thus, proteins dynamics are involved in all the biological functions such as protein folding, ligand binding, protein complex formation, enzymatic catalysis, signalling, cell regulation and so on.

According to the free energy landscape model, proteins have different thermodynamically accessible conformations within a hierarchy of time-scales (Figure 4.1). The population of each ensemble follows a Maxwell-Boltzmann distribution which depends on the Gibbs free energy. Local minima corresponds to well-established states. Rates of interconversion between the states are determined by the energy barriers that separate them, which are proportional to the time-scale that is needed to stochastically convert between states. Thus structure of a certain protein is averaged to the time-window we are looking at. Moreover, energy-landscapes are unique for each protein, but changes with other conditions such as temperature, pressure, ionic strength, ligand binding, protein binding, pH, etc.

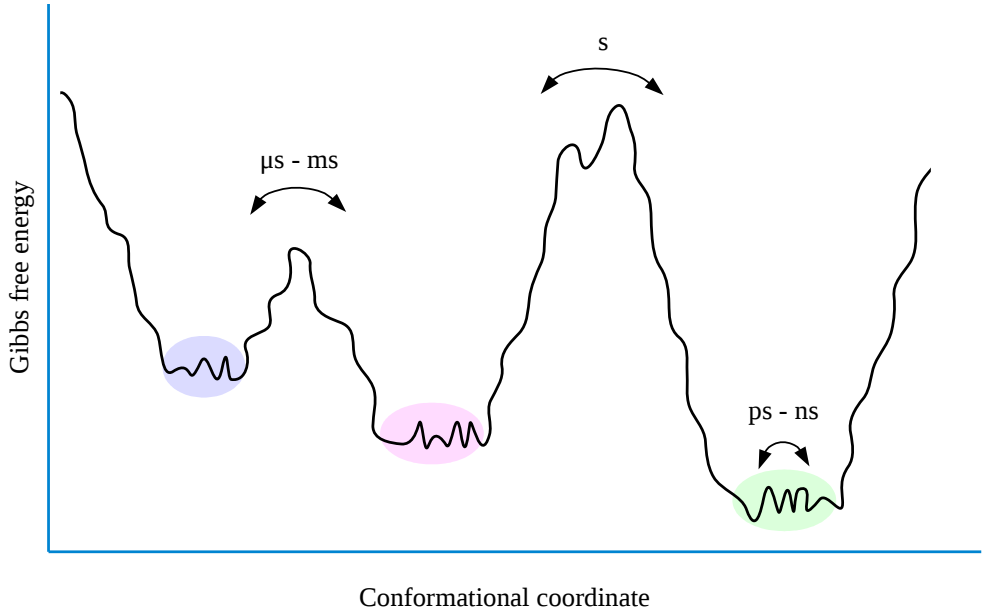


Figure 4.1. Free-energy landscape of proteins. Dynamic protein models are represented as a function of conformational coordinates. Minima correspond to well-defined structures. Population of each state is defined by its Gibbs free energy. Fast bond vibrations occur in picosecond (ps) to nanosecond (ns) time-scale, segmental motions occur in microseconds (μs) to milliseconds (ms), and finally large domain rearrangements occur in slow time-scales (s).

4.1.2 TECHNIQUES TO STUDY PROTEIN DYNAMICS

As for the high-resolution structure determination, X-ray and NMR are the best techniques.

4.1.2.1 X-Ray Crystallography

For protein crystallisation dynamic regions are better to be avoided. In fact, crystallographic structures are considered to be frozen protein structures as proteins are packed to their local minima energy. Nevertheless, X-ray structures are used to obtain indirect information about protein flexibility at fast bond vibrations or segmental motions level:

- Flexible regions in proteins often do not fit properly electron density maps.

- Residues that adopt different conformations usually appears with partial occupancy in crystallographic structures.

- B-factor represent the fluctuation of a certain residue around its average position under certain temperature.

However, X-ray crystallography has a big handicap regarding big conformational changes. It can only detect a single energy minima but not the transition between them. Therefore, scientists often obtain the starting and ending pictures of a conformational change but they are not able to know how or when these transitions occur. Thus, molecular dynamics are used to fill the gaps from the different energetic states of a certain protein.

4.1.2.2 NMR

An indirect source of protein dynamics information emerges from analysing the NMR structures deposited in the protein data bank. Often, NMR entries have several structures that fit the structural restrains. Even if these diverse models only reflects the technical noise, in practice these dynamics patterns suit well with the flexibility patterns described by other methods.

Nevertheless, nuclear magnetic resonance is a very rich source of direct information about protein dynamics. Keeping in mind the NMR limitations discussed in Chapter 3, this versatile technique provides structural and dynamical information at residue level in many different time-windows. Moreover, each time-window can be studied by different NMR experiments (Figure 4.2).

- **Real-time (RT) NMR** (slow exchange) is a powerful tool to study dynamic processes at second (s) time-scale where conformational changes and solvent hydrogen exchanges occur. This time-scale is specially interesting as movement of large protein domains are registered. These movements are highly dependent on higher energy. Dynamics are quantified the signal intensities in time.

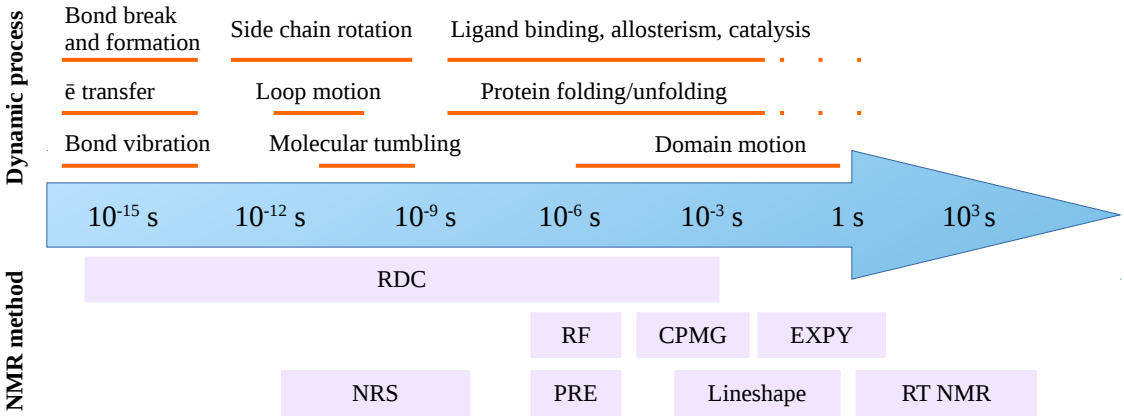


Figure 4.2. Protein dynamics over different time ranges. These protein motions can be studied by different NMR experiments as discussed in the text.

- **EXchange Spectroscopy (EXSY)** (slow exchange) is very useful to quantify dynamic processes at 10-5000 ms time-scales. Segmental motions, ligand bindings or releases, topological interconversion and other dynamic regulators of functional processes occur at this time-window. This method could be really useful for studying dynamics between two structural states.

- **Lineshape analysis** (slow-intermediate exchange) is used for analysing some physiological processes such as protein-ligand or protein-protein as well as slow-intermediate conformational changes in small domains. The time scale for this experiment is from 10 to 100 ms. Often, a series of spectral acquisitions are done doing a titration for ligand, temperature, pH and so on.

- **Carr-Purcell Meiboom-Gill Relaxation Dispersion (CPMG RD)** (intermediate-fast exchange) is a convenient tool for measuring dynamics at 0.3 to 10 ns to obtain thermodynamics, kinetics and structural information for exchange processes.

- **Rotating Frame Relaxation Dispersion (RF RD)** (intermediate-fast exchange) is employed to study dynamics at 20 – 100 μs. At this time-window some physiological processes such as side chains', secondary structure elements' and loops' motions occur. Thus, those processes are implicated in ligand binding or release, allostery and enzymatic activity.

- **Residual Dipolar Coupling (RDC)** (intermediate-fast exchange). Previously, in chapter 3, we have talked about this method. RDCs are collected using at least five different anisotropic media where the molecular motions can be interpreted within the frame of the aliment tensor. Thus motions at ns-μs time-scale can be detected.

- **Paramagnetic Relaxation enhancement (PRE)** (intermediate-fast exchange) shows protein dynamics at μs time-scale. This methods detects the line-broadening due to two additional relaxation mechanisms: electron-nucleus dipolar or Curie-spin relaxation. Besides

this utility, paramagnetic probes could also be used to collect long-range structural restrains by collecting PseudoContact Shifts (PCS) that alter the resonance frequency.

- **Nuclear Spin-Relaxation (NSR)** (fast exchange) is used to study processes at ns-ps. Fast bond vibrations and librations are produced by thermal fluctuations at this time-scale and could influence the entropic component of a protein, regulating the stability of some protein conformations. Three mechanisms are related to these motions and the MHz-frequency oscillation fields: Chemical Shift Anisotropy (CSA), Dipolar Couplings (DC) and Quadrupolar interactions.

For further reading: (Göbl & Tjandra, 2012; Kleckner & Foster, 2011; Sivanandam & Millet, 2015).

4.1.3 TECHNIQUES TO STUDY PROTEIN INTERACTIONS

In the crowded cytosol, proteins interact with each other forming protein-protein or protein-ligand complexes. Protein quaternary structure includes all proteins (homomers or heteromers) and elements that assemble together for a determined time period and for an specific duty. Complex formation is driven by the combination of the protein dynamics of each member, which suffers conformational changes between the free and bound states.

Protein-protein or protein-ligand binding can be determined by following techniques (from high to low resolution):

4.1.3.1 X-Ray Crystallography

Until 2015 most proteins have been determined using X-ray crystallography. If we have a look at structures stoichiometry we find that around 44000 monomers, 41000 homomers and 12500 heteromers have been solved. Thus, the 87 % of the structures involved only a single polypeptide chain which self-assembles into homomers. However, the number of heteromers is still low but will increase in the following years as the technique is evolving.

Depending on the structure resolution protein-protein or protein-ligand interactions can be analysed in a certain crystal, but we should take into the account that those structures do not show dynamics, just a frozen picture. Nevertheless, critical residues or regions involved in protein interaction can be detected depending on the structure resolution. Afterwards, more experimental data are required to corroborate the importance of selected residues.

4.1.3.2 Electron microscopy

Electron microscopy is perhaps the best technique for studying big protein complexes. Even if this technique is very young, interestingly only 60 structures correspond to monomers, around 250 to homomers and 450 to heteromers. Unlike NMR or crystallography, this method is very

suitable for large structures or complexes, particularly heteromers with different subunits. Nevertheless, atomic resolution if often lost when dealing with EM structures acquisition, which means that some small ligands such as ions or chemical compounds are not detected.

4.1.3.3 NMR

In chapter 2 we have already explained how a protein structure can be elucidated from NMR data. Due to the technical limitations most of the protein structures in PDB correspond to monomers (~ 8400), 950 to homomers and 45 to heteromers. Nevertheless, in the contrary of previous techniques, NMR is a very powerful technique to characterize protein-protein or protein-ligand interactions at atomic level without the necessity of resolving any protein structure. There are several techniques to determine protein interaction sites and constants. Here we summarize some of the most popular methods (Figure 4.3).

4.1.3.3.1 Paramagnetic Relaxation Enhancement (PRE)

Spin-spin relaxation rates are proportional to the product of the squares of the giromagnetic ratios of each spin. Thus, considering that the giromagnetic relation of an electron is 658 times bigger than a proton, the presence of a single unpaired electron in a close proton will have more relaxation effect than another nearby proton. In other words, all those proton signals close to a paramagnetic element will “disappear” from the spectra without affecting the rest of signals (Figure 4.3.B). This technique provides structural restraints up to 30 Å.

For example, for proteins that bind metals or cationic ions (like CaM), paramagnetic relaxation will occur when a paramagnetic metal (Mn^{2+}) binds to a binding region. In this case, all the signals around the binding site would disappear indicating the residues implicated in the metal binding.

On the other hand, paramagnetic tags such as nitroxide radicals, Mn^{2+} chelates or lanthadines can be introduced selectively into one of the protein-protein or protein-ligand binding partners. Thus protein interaction sites can be easily monitored.

4.1.3.3.2 Intermolecular NOE

Short distance atomic contacts (5 – 6 Å) are detected by NOE effect (see Chapter 3). These give very valuable information not only about which residues are implicated in the contact, but also the distances between interacting faces.

In case of protein-protein interactions when interaction partners can be expressed and purified separately, different isotopic labelling can be use for each protein in order to avoid discriminate easily inter-molecular contacts from intra-molecular contacts. Nevertheless, in case of same isotopic labelling in both proteins is more complicated, but not impossible, to distinguish between inter and intra-molecular NOE signals (Figure 4.3.C).

In case of small ligands, protein-ligand interactions can be easily quantified either looking at the NOE differences in the proton signals coming from the ligand or from the protein. However, this method is indicated specially for high affinity complexes (K_d > ~ 50 nM) where two proteins or protein-ligand stay in contact for longer time.

4.1.3.3.3 Chemical Shift Perturbation (CSP)

This method is widely used either for high or low affinity interactions. In the most common scenario, only one of the interacting partners is isotopically labelled (usually by ^{15}N). Then ^{15}N-HSQC spectrum are collected with increasing concentration of the other interacting partner. As chemical shifts are really sensitive to small changes in the electronic environment, the interaction interface is easily detected (Figure 4.3.D). However, if the protein suffers a conformational change in a region far from the interaction site, this approach could lead to a misinterpretation (for example protein allosterism).

Once the interaction is taking place, the two molecules are in equilibrium between their free and bound states which is described as a dissociation constant K_d (Figure 4.3.A). Taking into the account the exchange rate of the complex formation and the chemical shift differences between the bound and free protein states, three types of HSQC correlation can be observed:

- Slow exchange regime: two set of signals are detected, one of the reporter free state and another of the reporter in bound state. In this case signal intensities can be measured to monitor the populations or protein concentrations of the free or bound protein. Nonetheless, this situation is not very convenient as signals have to be reassigned *de novo*.

- Fast exchange regime: interconversion from the free and bound forms is faster than the chemical shift differences. Only a single set of signal is detected and often peaks move in a continuous manner. Therefore there is not need of reassigning.

- Intermediate exchange regime: could be defined as a situation in between slow and fast exchange where for some residues two set of signals are detected and for other only a single set. Often this situation lead to signal line broadening that may prevent signal from observation.

4.1.3.3.4 RDC

Either one or both interacting partners can be labelled. After collecting the free state in an anisotropic media, ligand is added to the sample. On one hand molecular tumbling of the free and bound state is different and on the other hand, the residues exposed to the anisotropic media can change after the ligand addition. Thereby, measuring the RDC differences before and after protein complex formation gives valuable information of the residues in the interacting interface (Figure 4.E).

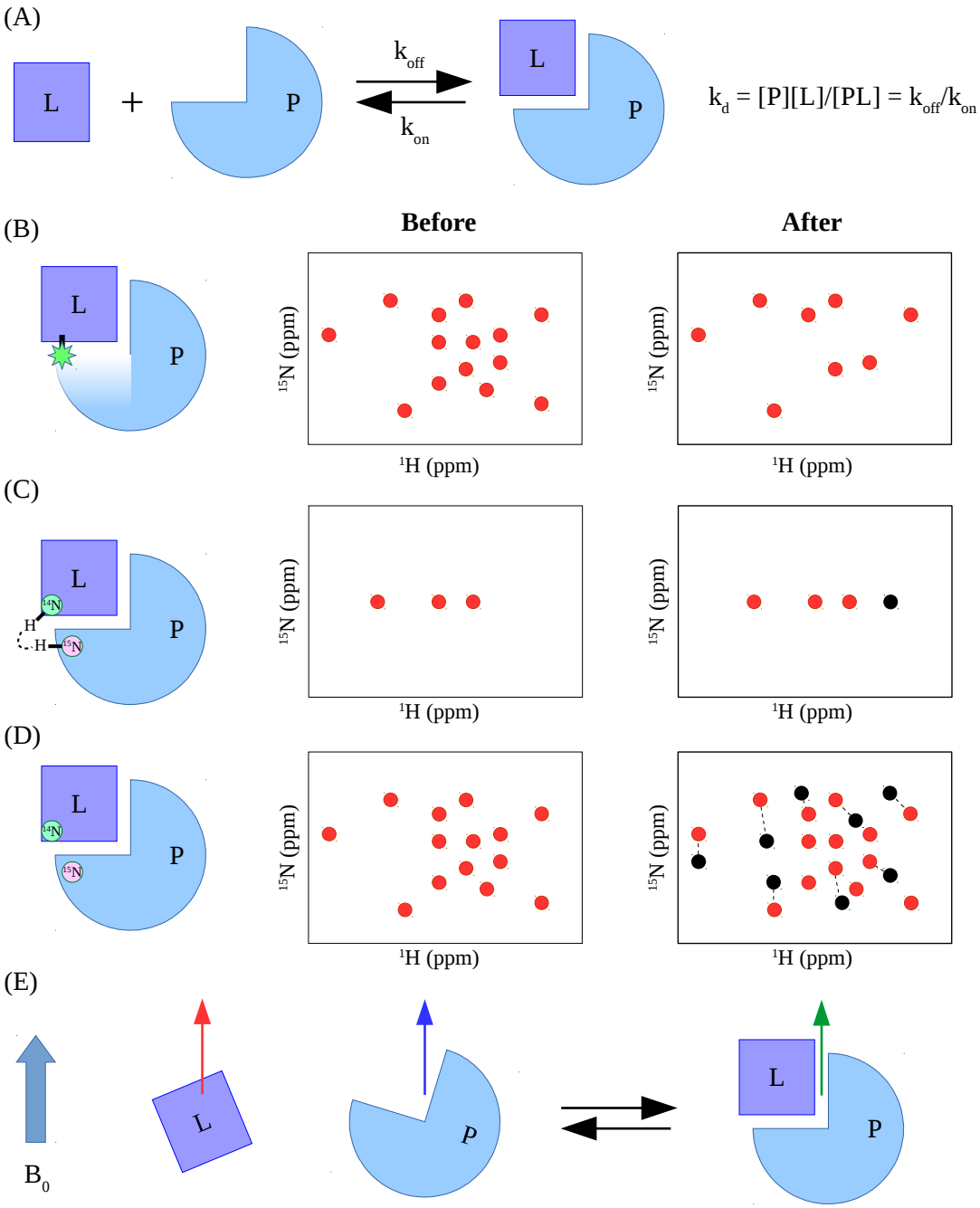


Figure 4.3. NMR detection of protein interactions. (A) A protein (P) interacts with a ligand (L) which could be another protein, small molecule, ion, DNA, etc. Dissociation constant is defined as represented. (B) A paramagnetic tag is attached to a ligand. Resonances close to tag in protein are broadened while other remains unaffected. (C) Internuclear NOE contacts can be discriminated from a intranuclear NOE by labelling a single partner. (D) Unlabelled binding partner is added to a 15N-protein. This causes changes in the chemical environment of nuclei. (E) RDCs are based on the orientation of molecules towards the magnetic fields (B_0). When free protein and ligand are aligned differently than when they are bound together useful relative orientation restrains can be obtained.

4.1.3.4 Small Angle X-Ray Scattering (SAXS)

Global spatial changes and protein-protein binding can be detected also by Small Angle X-ray Scattering technique. One of its best features is that proteins are in solution and the concentration needed for SAXS determination is very low. In this case protein shapes are determined where usually protein structures are fitted inside. The sensitivity of ligand binding is represented a big structural changes. Nevertheless, in case of small ligands interaction may not be visible.

4.1.3.5 Föster Resonance Energy Transfer (FRET)

Foster's Resonance Energy Transfer is a non-radiative dipole-dipole coupling transfer of energy that requires some interaction between a donor and an acceptor molecule. It can occur if the emission spectrum of the donor overlaps with the absorption spectrum of the acceptor. This method could be use to detect protein interactions with ligands and the relative orientation and distance between the fluorophores. This technique is used either *in vitro* or *in vivo*. Moreover, in vitro assays do not require purified samples even if it is advisable.

This phenomena is highly distance-dependent where the distance between donor and acceptor is typically in the range of 1-10 nm. Moreover, the orientation of transition dipoles also affect energy transmission. In fact, maximum transition occurs when chromophores are located in parallel to each other. When chromophores are located perpendicularly there is not FRET effect. Usually there is a big controversy when evaluating FRET results as the relative orientation or distance between fluorophores are not always known.

4.1.3.6 Surface Plasmon Resonance (SPR)

Proteins a fixed to a metal surface where ligands flow past. When ligand binds immobilised proteins it causes changes in the refractivity and reflectivity properties of metal. These transitions is detected by an optical device and are related to the mass of both ligand and protein. The protein needed for this experiment is low compared to ITC and a hundreds of compounds can be screened in one day.

4.1.3.7 Isothermal titration calorimetry (ITC)

This tool measures heat release after a ligand binding. It is really accurate in determination of binding constants. Besides, this technique is able to deconvolute the different contributions of entropy and enthalpy in ligand binding. These relative contributions can provide information about polar and hydrophobic interactions respectively. As a drawback, this method requires large amount of proteins.

4.1.3.8 Other methods

In order to detect protein-protein interactions there are a huge variety of techniques that, even if they are extremely accurate, they do provide valuable information. Fluorescent methods, sedimentation equilibrium, gel filtration, native-gel electrophoresis, cross-linking, protein affinity chromatography, affinity blotting and a big etcetera are used for this purpose. For further reading (Phizicky & Fields, 1995).

4.1.4 K_v7 CHANNELS AND DYNAMICS

Why are we drawing our attention to protein dynamics? As we have seen in the introduction CaM is magnificent Ca^{2+} sensor which suffers conformational rearrangements upon Ca^{2+} or target binding. Besides, this calcium signalling is often sensed by CaM and transmitted to its targets by mechanical energies. Every single CaM state (Apo-CaM, Holo-CaM, Apo-CaM/Target, Holo-CaM/Target) could be considered as local minima in the free-energy landscape. Thus, previously resolved CaM/K_v7.2_AB structure is the average structure of an energy minima in a certain time-window. In other words, the studied protein complex was undergoing structural changes during NMR data acquisition but the resolved structure model corresponds to the most prevalent conformation.

But, in which step of the CaM regulation of K_v7.2 channels have we captured the complex structure? Is it in the “active” conformation or in an “inactive” conformation? Is the protein complex calcified? What is the effect of calcium in the complex?

In this chapter, we will try to unravel the steps for CaM to bind the C-terminus of the K_v7.2 channel. Moreover, the influence and effect of the Ca^{2+} will be discussed. In order to do that, first we will study CaM lobes' preferences to each binding site (helices A or B) in the presence or absence of Ca^{2+} by *in vitro* fluorimetry based on Dansyl-CaM. A calcium^{+} titration will be done using ^{15}N-Complex and spectra will be collected so as to explore CaM/K_v7.2_AB complex rearrangements upon Ca^{2+} addition. Thus, the calcification state of CaM in the resolved complex will be defined. Finally, a model for CaM binding and regulation of K_v7.2 channels will be discussed.

4.2 MATERIALS AND METHODS

Most of the materials and methods in this chapter were the same as Chapter 2 or Chapter 3. However some differences are described:

4.2.1 MOLECULAR BIOLOGY TECHNIQUES

Following **plasmids** were used:

Name	Provider	Size (bp)	Resistance	Tags	Cleavage
Easy-T	Promega (A1360)	3015	Amp		
pProEX-HTc	Invitrogen (10711018)	4780	Amp	His	TEV
pGEX-6P-1	GE Healthcare (28-9546-48)	4984	Amp	His-GST	PreScission
pOKD4	(Dzivenu et al., 2004)	6473	Kan		

Following ***E. coli*** strains were used:

Strain	Provider	Humanized?	Extra plasmid resistance?
BSJ	SUNY Stony Brook	No	No
BL21(DE3)	Novagen (69450-3)	No	No
CodonPlus(DE3)(RIL)	Agilent (#230240)	Yes	Cam

4.2.2 PROTEIN BINDING CHARACTERIZATION FLUORESCENCE SPECTROSCOPY

The K_v7.2 CaM binding domain (residues 310-548 from human K_v7.2, named here Q2ABc WT) helix A (hA residues G310-T359) and helix B (hB residues S450-S590), fused to GST, were expressed in *E. coli* BL21-DE3 cells and purified using glutathione-sepharose as described previously (Chapter 2, Section 2.2.4.5). Recombinant CaM was produced in BL21–DE3 bacteria and purified as described (Chapter 2, Section 2.2.4.1). The cDNAs encoding the human CaM N-lobe (residues 1–78) and C-lobe (residues 79–148) were provided by the group of Daniel L. Minor Jr. (Cardiovascular Research Institute, University of California). Both lobules were purified by Ni-Sepharose resin as described before (Chapter 2, Section 2.2.4.4) and afterwards MBP was removed by TEV protease. There are no differences in the primary amino acid sequence of rat and human CaM. The oligomerization state of the purified proteins was then examined by dynamic light scattering (DLS).

In the fluorometric experiments using Dansyl-CaM (D-CaM), the CaM dansylation, sample preparation and fluorescent measurements were performed as described (Chapter 2, section 2.2.5.2). The D-CaM binding buffer used was: 25 mM Tris-HCl, 120 mM KCl, 5 mM NaCl, 2 mM $MgCl_2$, 10 mM EGTA, pH 7.4). The level of contaminant Ca^{2+} in the protein preparations was determined by inductively coupled plasma mass spectroscopy carried out at the Department of Analytical Chemistry (University of the Basque Country), and was found to be less than 40 nM. Experiments were also performed in the presence of an excess of free Ca^{2+} (3.9 or 100 μM), adding 9.63 or 9.985 mM Ca^{2+} to the D-CaM binding buffer. The data obtained at lower (3.9 μM) and higher (100 μM) free Ca^{2+} concentration were indistinguishable. Fluorescence enhancement was plotted against that in [Ca^{2+} free] to generate the concentration-response curves. The parameters of the Hill equation ($E = Bmax*([Ligand]^h/([Ligand]^h+IC_{50}^h)$), or the parameters of a two sites Hill equation ($E = B_{max1}*([Ligand]^{h1}/([Ligand]^{h1}+IC_{50}1^{h1}))+B_{max2}*([Ligand]^{h2}/([Ligand]^{h2}+IC_{50}2^{h2})$) were fitted to the data by curvilinear regression, enabling the apparent affinity (EC_{50} or concentration that gives half-maximal change in fluorescence emission intensity) and Hill coefficient to be determined.

Protein concentrations were measured by UV absortption. In the competition assays, the purified CaM, N-, and C-lobe, and an equimolar mixture of both lobes or intact CaM were added in increasing amounts to the starting mixture of D-CaM (12.5 nM) and a ligand (Q2ABc, hA or hB). Steady state fluorescence was reached within 30 seconds, and measurements were taken more than 60 seconds after sample addition. No significant effect of the isolated lobes or GST was found. The reduction in fluorescence was plotted against the concentration of the lobes.

To make the assay more sensitive we used the concentration of the ligand that corresponded to its calculated EC_{50} for the increase in D-CaM fluorescence emission. The data was plotted as the reduction in fluorescence (taking the initial complex as 100%) relative to the [competing lobe/s]. The following equation, which assumes as 1:1 stoichiometry, was used to estimate the K_d, where F is the increase in fluorescence, F_{max} is a free parameter that represents the maximal fluorescence, [peptide] is the known total peptide concentration [CaM] is the known concentration of total D-CaM, and K_d is a free parameter for the affinity constant:

$$F = F_{max} * \left(1 - \frac{2 * K_d}{[peptide] - [CaM] + K_d + \sqrt{([peptide] - [CaM] - K_d)^2 + 4 * K_d * [peptide]}} \right)$$

The results are expressed as the means ± S.E.M from three or more experiments. For statistical comparison, significance was evaluated using the unpaired Student t test, with values of $P < 0.05$ (*), $P < 0.01$ (**) and $P < 0.001$ (***) considered to be statistically significant.

4.2.3 CaM-K_V7.2 Complex Chemical Shifts Library Comparison

Biological Magnetic Resonance Data Bank (Doreleijers et al., 2003) was used to find the chemical shifts for 34 different CaMs:

#	BMRB	Ligand	N-lobe	C-lobe
1	15470	Munc13-1 peptide	Ca^{2+}	Ca^{2+}
2	15624	Calcineurin peptide	Ca^{2+}	Ca^{2+}
3	15650	DRP1p peptide	Ca^{2+}	Ca^{2+}
4	15852	DAPk peptide	Ca^{2+}	Ca^{2+}
5	1634	Skeletal Muscle Myosin Light-Chain Kinase	Ca^{2+}	Ca^{2+}
6	16465	Ca^{2+} Pump 18-1 Motif	Ca^{2+}	Ca^{2+}
7	17264	Na_V1.5 peptide	Apo	Apo
8	17771	Olfactory CNGC channel	Ca^{2+}	Ca^{2+}
9	17807	CaM BD of L-selectin	Ca^{2+}	Ca^{2+}
10	18027	CaM BD of iNOS	Ca^{2+}	Ca^{2+}
11	18028	CaM BD of eNOS	Ca^{2+}	Ca^{2+}
12	18082	ER alpha peptide	Ca^{2+}	–
13	18084	ER alpha peptide	–	Ca^{2+}
14	18302	CaM BD of L-type Ca_v	Ca^{2+}	–
15	18556	HIV-1 matrix protein peptide	Ca^{2+}	Ca^{2+}
16	19036	N-Terminally Acetylated -Synuclein oeptide	Ca^{2+}	Ca^{2+}
17	19050	CaM BD Na_V1.2	Ca^{2+}	–
18	19586	CaM BD of eNOS	Ca^{2+}	Ca^{2+}
19	19604	HIV-1 matrix protein	Ca^{2+}	Ca^{2+}
20	4270	Ca^{2+}-CaM Dependent Protein Kinase	Ca^{2+}	Ca^{2+}
21	4284	CaM Binding Peptide of the Ca^{2+}-Pump	Ca^{2+}	Ca^{2+}
22	4310	SEF2-1 peptide	Ca^{2+}	Ca^{2+}
23	5286	CaMKIp peptide	Ca^{2}	Ca^{2}
24	5480	Olfactory CNGC channel peptide	Ca^{2+}	Ca^{2+}
25	5770	Plant Glutamate Descarboxylase peptide	Ca^{2+}	Ca^{2+}
26	5893	Microtubule associated protein F-STOP	Ca^{2+}	Ca^{2+}
27	6023	CaM BD of Protein Phosphatase 2B	Ca^{2+}	Ca^{2+}
28	6541	CaM with Ca^{2+}	Ca^{2+}	Ca^{2+}
29	7416	CaM with Ca^{2+}	Ca^{2+}	Ca^{2+}
30	7417	CaM with Dysprosium(III)	Dy^{3+}	Dy^{3+}
31	7418	CaM with Thulium(III)	Tm^{3+}	Tm^{3+}
32	7423	CaM/DAPk peptide with Ca^{2+}	Ca^{2+}	Ca^{2+}
33	7424	CaM/DAPk peptide with Terbium(III)	Tb^{3+}	Ca^{2+}
34	7425	CaM/DAPk peptide with Thulium(III)	Tm^{3+}	Ca^{2+}

CaM's chemical shift values of the assigned K_v7.2/CaM complex were used as a reference (REF) and they were compared with the library (LIB) proteins. For each residue chemical shift differences were defined as follows.

$$\Delta\delta_{NH} = \sqrt{(\delta_H^{LIB} - \delta_H^{REF})^2 + [0.17 * (\delta_N^{LIB} - \delta_N^{REF})]^2}$$

* Nitrogen normalizing value 0.17 was taken from (Piserchio et al., 2002)

N-lobe was defined from 1 to 76 and C-lobe from 83 to 148 amino acids. EF1 (20-31), EF2 (56-67), EF3 (93-104) and EF4 (129-140) were also defined for data analysis.

When comparing these regions with the library the following equation was used.

$$\overline{\Delta\delta}_{REGION} = \frac{\sum^{n} \Delta\delta_{NH}}{n}$$

All the calculations and graphical representations were done with MatLab scripts.

4.2.4 Calcium Titration on ^{15}N-CaM/K_v7.2_AB Complex

CaM and His-Q2AB with deletions ΔF316_R325, Δ6L and ΔP533_H546 (GB116) were both labelled with ^{15}N and purified as explained before (Chapter 3, Section 3.2.4). Nevertheless, some changes were done in order to complete the calcium titration in the magnet.

In the affinity chromatography following buffers were changed: *Lysis Buffer* (120 mM KCl, 20 mM Tris-HCl pH 7, 1 mM PMSF, 2 mM Imidazole and 1 tablet of Protease inhibitor without EDTA), *Wash Buffer* (120 mM KCl, 20 mM Tris-HCl pH 7) and, *Elution Buffer* (120 mM KCl, Tris-HCl 20 mM pH 7, 300 mM Imidazole). After protein elution sample was dialysed two times against *Dialysis Buffer* (120 mM KCl, 20 mM Tris-HCl pH 7, 5 mM EGTA) to get rid of free calcium. Afterwards, sample was separated in two and two different dialysis were done. 1) *Apo Running Buffer* (120 mM KCl, mM MES pH 6, 0.5 mM EGTA) and 2) *Calcium Running Buffer* (120 mM KCl, 100 mM MES pH 6, 0.5 mM EGTA, 5 mM $CaCl_2$).

Gel filtration was done using Superdex 75 pg 26/60 equilibrating the column for each sample. In both cases the elution profiles were almost identical. Positive fractions were concentrated till 500 µl using a 3 kDa cut-off Amicon ultra-15 concentrator in a fixed angle rotor at 3500 g. Protein concentrations were 125 µM and 134 µM for Apo and Ca^{2+} sample.

^{15}N-HSQC spectra were collected using 800 MHz Bruker Avance III spectrometer at 303 K. First, both separated samples were measured. In these cases different protein fingerprint was detected easily as a calcium consequence. Afterwards, a Ca^{2+} titration was performed adding specific volumes of the calcified sample to the apo-sample. Free calcium was calculated using CHELATOR software (Schoenmakers et al., 1992), which does not take into account the Ca^{2+}

buffer capacity of CaM. For this reason this “free”-calcium is name as “available”-calcium for the complex. In other words, available calcium represent the amount of calcium that is not chelated by EGTA. This process was repeated eight times reaching following set of data:

1. Apo-Complex
2. Complex with Total 0.6 mM of Calcium → 210.13 μM of available-calcium
3. Complex with Total 0.8 mM of Calcium → 369.25 μM of available-calcium
4. Complex with Total 1 mM of Calcium → 548.77 μM of available-calcium
5. Complex with Total 2 mM of Calcium → 1.5 mM of available-calcium
6. Complex with Total 3 mM of Calcium → 2.5 mM of available-calcium
7. Complex with Total 4 mM of Calcium → 3.5 mM of available-calcium
8. Complex with Total 4 mM of Calcium → 4.5 mM of available-calcium

^{15}N-HSQC spectra were processed using TopSpin software and analysed using Sparky program. ^{1}H-^{15}N chemical shift correlations were assigned using the chemical shifts employed for the protein structure characterization.

However, due to the slow exchange rate of the protein populations more than one set of chemical shifts were visible. Thus, for these chemical shifts from CaM that were difficult to assign, ^{1}H-^{15}N correlations were compared to BMRB 17264, 19604 and 18027 files.

Chemical Shift Perturbation values were calculated using the peaks list of each spectra and using MatLab software. For each residue chemical shift differences were defined as follows, where “Titration” refers to the data used in this experiment and “structure” refers to the data from the chemical shifts used to get the structure of the complex.

$$\Delta\delta_{NH} = \sqrt{\left(\delta_H^{Titration} - \delta_H^{Structure}\right)^2 + \left[0.17 * \left(\delta_N^{Titration} - \delta_N^{Structure}\right)\right]^2}$$

* Nitrogen normalizing value 0.17 was taken from (Piserchio et al., 2002)

4.3 RESULTS

4.3.1 D-CAM BINDING CHARACTERIZATION TOWARDS HELICES A AND B

Fluorescence spectroscopy is a very sensitive technique which requires small amounts of protein. In this assay, dansyl group was covalently attached to CaM in order to monitor ligand binding. When the chromophore group is located in more hydrophobic environment a fluorescence enhancement is visible. Also, it has been demonstrated that dansyl group does not alter CaM ligand binding (Kincaid et al., 1982). These results are based on Dr. Alessandro Alaimo's work (Alaimo, Alberdi, Gomis-Perez, Fernández-Orth, et al., 2014).

4.3.1.1 CaM complex formation with Q2ABc weakens its Ca^{2+} binding affinity

First, CaM's affinity for Ca^{2+} was measured in the absence and in the presence of GST-Q2ABc. For the first part of the assay calcium titration was done using 12.5 nM of D-CaM. For the second part, same procedure was done in the presence of 200 nM of GST-Q2ABc. The experiment showed that Ca^{2+} binding affinity was decreased in the presence of Q2ABc, which was shifted from 0.7 µM to 3.6 µM of free Ca^{2+} respectively (Figure 4.4).

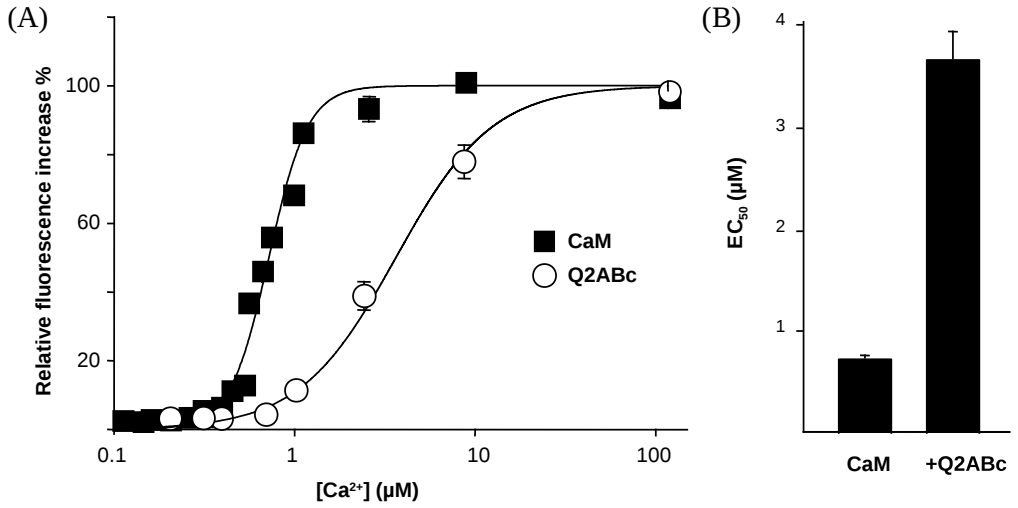


Figure 4.4. Q2ABc binding weakens the Ca^{2+}-CaM interaction. (A) Relative increase in fluorescence emission of D-CaM (12.5 nM) in response to increased Ca^{2+} of concentrations in the presence (open circles) and absence (filled circles) of a molar excess of GST-Q2ABc (200 nM). The lines are the result of fitting a Hill equation to the data. The data represent the means ± standard error from three or more independent experiments. The error bars are smaller than the symbols. The EC_{50} values obtained are (in µM): CaM = 0.72 ± 0.02 and Q2ABc = 3.64 ± 0.26. (B) Plot of the apparent binding affinity derived from the data.

4.3.1.2 CaM interaction with Q2ABc is dominated by N-lobe in the absence and by C-lobe in the presence of Ca^{2+}

The ability of the isolated N- and C-lobes to interact with Q2ABc in the absence and in the presence of Ca^{2+} was examined (Figure 4.5).

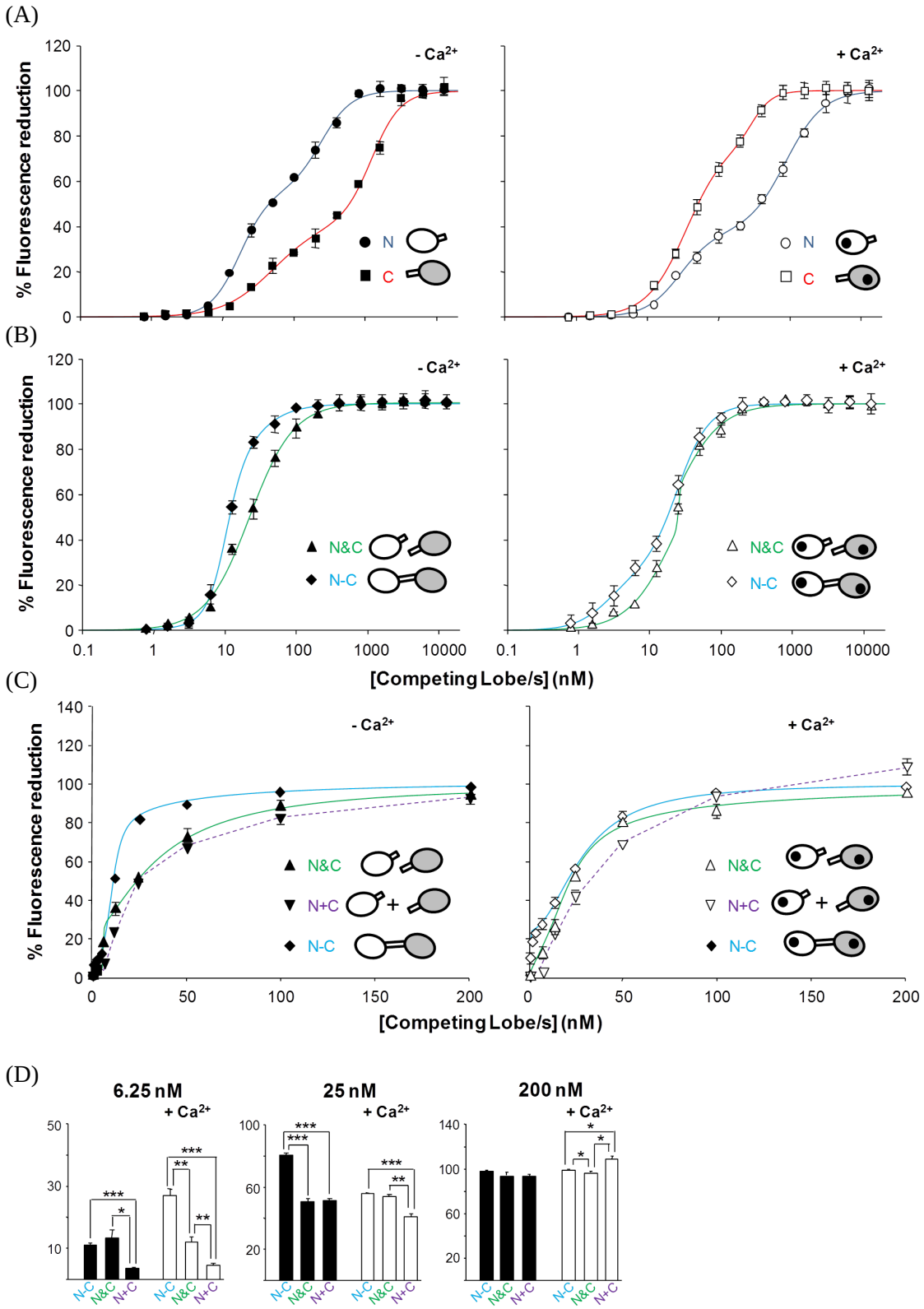


Figure 4.5. Q2ABc binds preferentially to the N-lobe in the absence and to the C-lobe in the presence of Ca^{2+}. (A) Relative concentration-dependent reduction in 12.5 nM D-CaM fluorescence emission when complexed with Q2ABc. To achieve maximal sensitivity in the assay, concentrations of Q2ABc that caused 50% of the maximal increase in D-CaM fluorescence emission were used (11 nM and 27 nM in the absence and presence of Ca^{2+}, respectively). D-CaM was mixed with Q2ABc and then each lobe (N or C), both lobes (N&C) or intact CaM (N-C) was added incrementally at the concentrations indicated. The data represent the means 6 standard error from three or more independent experiments. Some error bars were smaller than the symbols. The result of fitting a two sites Hill equation to the competition curves is compiled in Table 4.1. (A) The effect of incremental addition of the indicated lobes obtained in the absence of Ca^{2+} (Left, 10 mM EGTA added) and in the presence of 100 μM free Ca^{2+} (right). (B) The effect of incremental addition of CaM (N-C) and of an equimolar mixture of the lobes (N&C) obtained in the absence of Ca^{2+} (left) and in the presence of 100 μM free Ca^{2+} (right). (C) Comparison of the arithmetic addition of the curves obtained for each individual lobe (N+C) with the effect of an equimolar mixture (N&C) and of CaM (N-C) at concentrations under 200 nM in absence (left) or in the presence of 3.9 μM Ca^{2+} (that were indistinguishable from the results obtained in the presence of 100 μM Ca^{2+}). (D) Plot of the reduction on fluorescence at the indicated concentration of competing lobe(s). ***, significance at $P \leq 0.001$, **$P \leq 0.01$, *$P \leq 0.05$, unpaired Student's t test.

In the absence of free Ca^{2+}, both lobes interacted with $K_v7.2$ (Figure 4.5.A, red and blue lines), but the N-lobe displayed higher affinity than the C-lobe. In addition, the difference between an equimolar mixture (N&C) and the arithmetic addition of the effect of the lobes (N+C) at low concentrations was small. By contrast, the affinity for CaM was higher than that of the equimolar mixture of the lobes (Figure 4.5.B and C). Thus, linkage of the lobes favours binding to Q2ABc.

Calcium changed the interaction mode of both lobes, weakening binding to the N-lobe (Figure 4.5.A, blue line) and strengthening the interaction with the C-lobe (Figure 4.5.A, red line; see Table 4.1). The effect of holo-CaM (N-C) was again larger than that of the equimolar mixture of lobes (N&C), although such differences were no longer relevant at concentrations ≥25 nM. In addition, the effect of an equimolar mixture of lobes (N&C) at low concentration differed from that of the arithmetic addition of lobes (N+C) (Figure 4.5.C and D), suggesting the existence of cooperativity.

		$Bmax_1$	$IC50_1$ (nM)	h1	$Bmax_2$	$IC50_2$ (nM)	h2
- Ca^{2+}	N	55.3 ± 4.1	17.3 ± 1.6	2.1 ± 0.2	44.7 ± 4.1	240 ± 28	2.0 ± 0.2
	C	41.6 ± 6.3	49.7 ± 11.1	1.2 ± 0.2	58.4 ± 6.3	1224 ± 133	2.1 ± 0.3
	N&C	10.3 ± 4.6	6.1 ± 2.8	≥5	89.7 ± 4.6	25.1 ± 1.4	1.4 ± 0.1
	N-C	62.5 ± 7.1	10.3 ± 0.5	3.4 ± 0.6	37.5 ± 7.1	18.6 ± 5.2	1.5 ± 0.2
+ Ca^{2+}	N	37.2 ± 2.9	28.9 ± 3.4	1.8 ± 0.2	62.8 ± 2.9	882 ± 69	1.6 ± 0.1
	C	77.6 ± 4.5	35.7 ± 3.2	1.5 ± 0.1	22.4 ± 4.5	303 ± 33.1	3.1 ± 0.7
	N&C	82.8 ± 2.1	21.5 ± 2.1	1.3 ± 0.1	17.2 ± 2.1	22.1 ± 2.6	≥5
	N-C	25.9 ± 8.1	2.6 ± 0.9	1.7 ± 0.5	74.1 ± 8.1	24.4 ± 2.5	1.9 ± 0.2

Table 4.1. Summary of binding parameters for $K_v7.2AB$ obtaining after fitting a two sites Hill equation to the Q2ABc data in figure 4.5.

4.3.1.3 CaM interaction with Q2ABc is dominated by Helix B in the absence and by Helix A in the presence of Ca^{2+}

By competition assay the contribution of each individual segment of the $K_v7.2$ binding site was examined. Although the binding of CaM with $K_v7.2$ is not fully recapitulated by its individual segments, this analysis helps in understanding how CaM interacts with the channel.

In the absence of Ca^{2+} the interaction with the N-lobe was dominant (Figure 4.5.A), and the affinity of this lobe was higher for helix B (Figure 4.6 and Table 4.2). The affinities of the C-lobe for either helix were essentially the same, albeit lower than the affinity of the N- lobe for helix B. By contrast, in the presence of Ca^{2+}, when the interaction with the C-lobe dominates (Figure 4.5.A), the interaction of highest affinity was that between the C-lobe and helix A. Overall, these data revealed that Ca^{2+} strengthened the interaction of helix A with both lobes, whereas for helix B a reduction in the affinity of the interaction with the N lobe was observed.

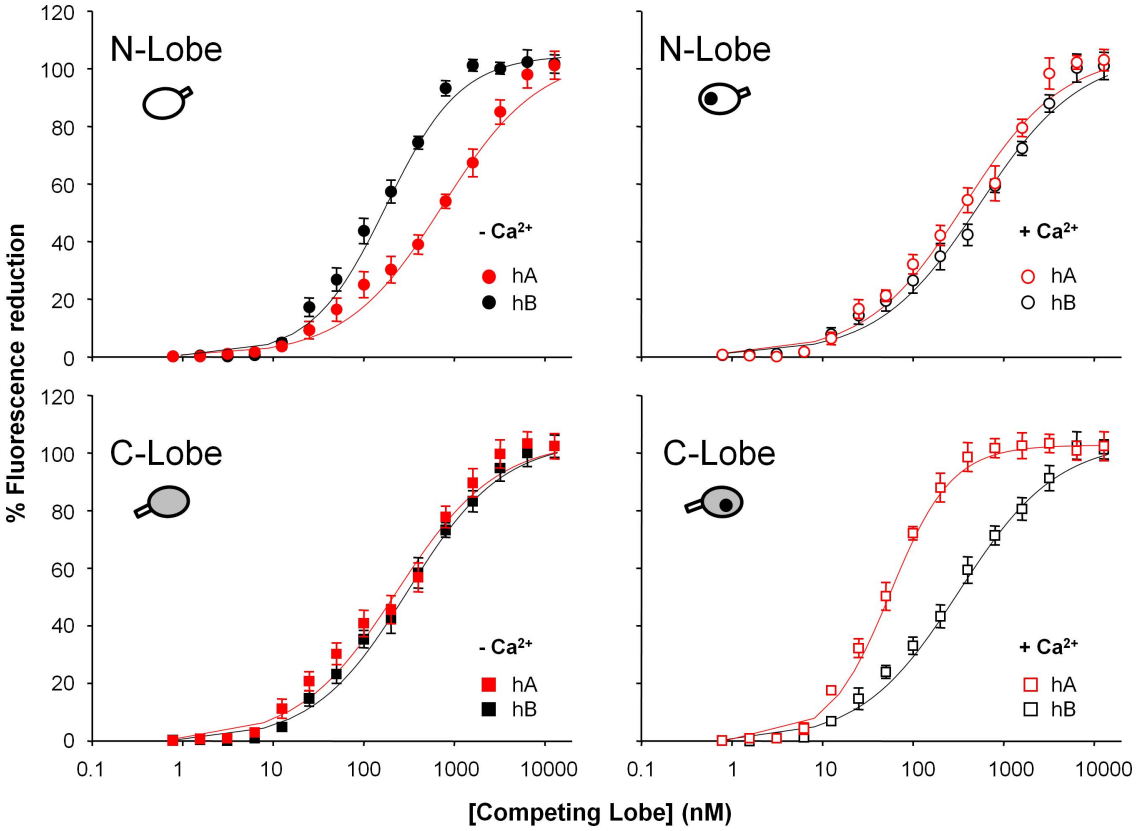


Figure 4.6. CaM's lobes preference for each helix (A and B). The N-lobe binds preferentially to helix B in the absence of Ca^{2+}, whereas the C-lobe binds preferentially to helix A in the presence of Ca^{2+}. Competition curves with isolated CaM lobes (N or C) obtained using the individual CaM lobes and performed in the absence (filled symbols) or in the presence of Ca^{2+} (open symbols). D-CaM (12.5 nM) was mixed with helix A (hA, red symbols) or helix B (hB, black symbols) at a concentration corresponding to its calculated EC_{50} for the increase in D-CaM fluorescence emission (46.4 and 65.6 nM in absence or presence of Ca^{2+} for helix A respectively, and 20.1 and 42.6 nM in absence or presence of Ca^{2+} for helix B respectively) and then each lobe was added incrementally at the concentrations indicated. The data represent the means 6 standard error from three or more independent experiments. Some error bars were smaller than the symbols. The result of fitting a Hill equation to the competition curves is compiled in Table 4.2.

	Helix	Lobe	IC_{50} (nM)	h	K_d (nM)
- Ca^{2+}	hA	N	533 ± 51	0.9 ± 0.1	530 ± 55
		C	181 ± 21	0.8 ± 0.1	169 ± 23
	hB	N	130 ± 8	1.1 ± 0.1	119 ± 9
		C	238 ± 16	0.9 ± 0.1	231 ± 18
+ Ca^{2+}	hA	N	284 ± 33	0.9 ± 0.1	267 ± 37
		C	46 ± 2	1.4 ± 0.1	17 ± 2
	hB	N	415 ± 42	0.8 ± 0.1	414 ± 48
		C	244 ± 17	0.9 ± 0.1	229 ± 21

Table 4.2. Summary of binding parameters for each CaM lobe obtaining after fitting a one site Hill equation to the helices data in figure 4.6.

By these short of experiments CaM lobes' preferences for helix A or B in the absence or presence of Ca^{2+} have been unravelled. Now it is clear that in the absence of calcium, N-lobe is the first making contact with helix B of K_v7.2 channels whereas, in the presence of calcium, C-lobe is the first making contact with helix A of K_v7.2 channels.

But, how can be known in which calcification stated is the structural model of CaM/K_v7.2-AB complex? Is CaM in its Apo, Holo or in a intermediate state? Is it possible to observe structural rearrangements in the complex by the effect of calcium? If so, could be possible to trace a model for Ca^{2+}-dependent CaM/K_v7.2-AB interaction?

4.3.2 Chemical Shifts Comparison Shows a Likely calcification-state of the N-lobe and an Apo-state of the C-lobe

In order to determine the calcification state of CaM a small database was created with the back bone's chemical shifts of 34 different assigned calmodulin. H-N values of the characterized CaM complexed with $K_v7.2$ were used as a reference in the comparison. Thus, values closer to zero reveals more similarity and higher values higher diversity. To make the comparison easier to understand, two groups were done: 1) The N-lobe was defined as amino acids from 1 to 76 and 2) C-lobe from 83 to 148 amino acids. Chemical shift comparison was done as explained in Materials and Methods. Afterwards values were plotted (Figure 4.7)

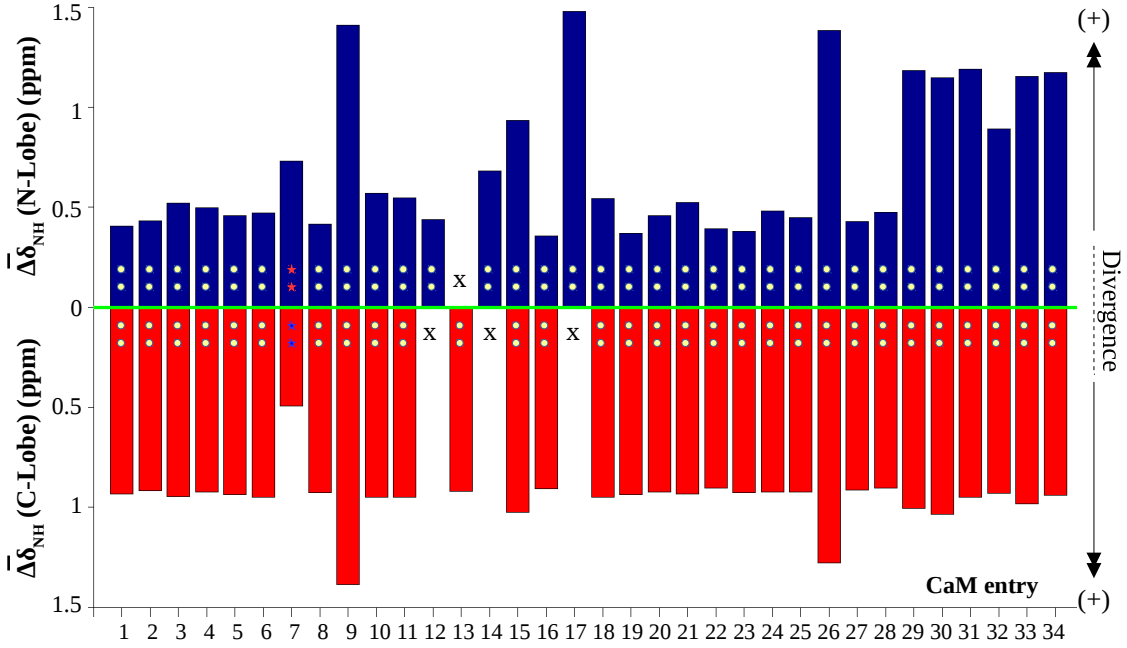


Figure 4.7. Similarities between calmodulin lobes complexed with $K_v7.2$ and database forming 34 different calmodulins. N-lobe values are represented in blue bars and C-lobe values in red. Inside bars the calcification state of this region can be distinguish, where yellow circles mean "calcified"- or "holo"-state and stars "apo"-state. For regions without any information a cross is shown. CaM-$K_v7.2$ chemical shifts were used as a reference (green line). There are not negative values. The closer to zero the more similar. Each bar represents a single CaM values from a concrete BMRB entry: 1, 15470; 2, 15624; 3, 15650; 4, 15852; 5, 1634; 6, 16465; 7, 17264; 8, 17771; 9, 17807; 10, 18027; 11, 18028; 12, 18082; 13, 18084; 14, 18302; 15, 18556; 16, 19036; 17, 19050; 18, 19586; 19, 19604; 20, 4270; 21, 4284; 22, 4310; 23, 5286; 24, 5480; 25, 5770; 26, 5893; 27, 6023; 28, 6541; 29, 7416; 30, 7417; 31, 7418; 32, 7423; 33, 7424; and 34, 7425.

Taking into consideration the chemical shifts differences from the N-lobe, it is noticeable that the similarities between different BMRB entries and CaM values when complexed to $K_v7.2$ are more uniform. In fact, the closer entry corresponds to number 16 (BMRB: 19036), a NMR structure of the complex of an N-terminally acetylated α-synuclein peptide with calmodulin (PDB: 2M55). Nevertheless, there is not clear differences with other calcified entries. Considering the unique entry without calcium in the N-lobe, it is worth to highlight that this entry shows more divergence than many of the calcified lobes.

On the other hand, C-lobe has just one entry which shows bigger similarities to the reference. This entry corresponds to number 7 (BMRB: 17264), a solution NMR Structure of apo-calmodulin in complex with the IQ motif of Human Cardiac Sodium Channel $Na_V1.5$ (PDB 2I53). For further analysis of which regions were altered, chemical shifts of closest entries were plotted (Figure 4.8). Once again it was corroborated that the N-lobe was more related to the calcified-version and the C-lobe was closer to the apo-version.

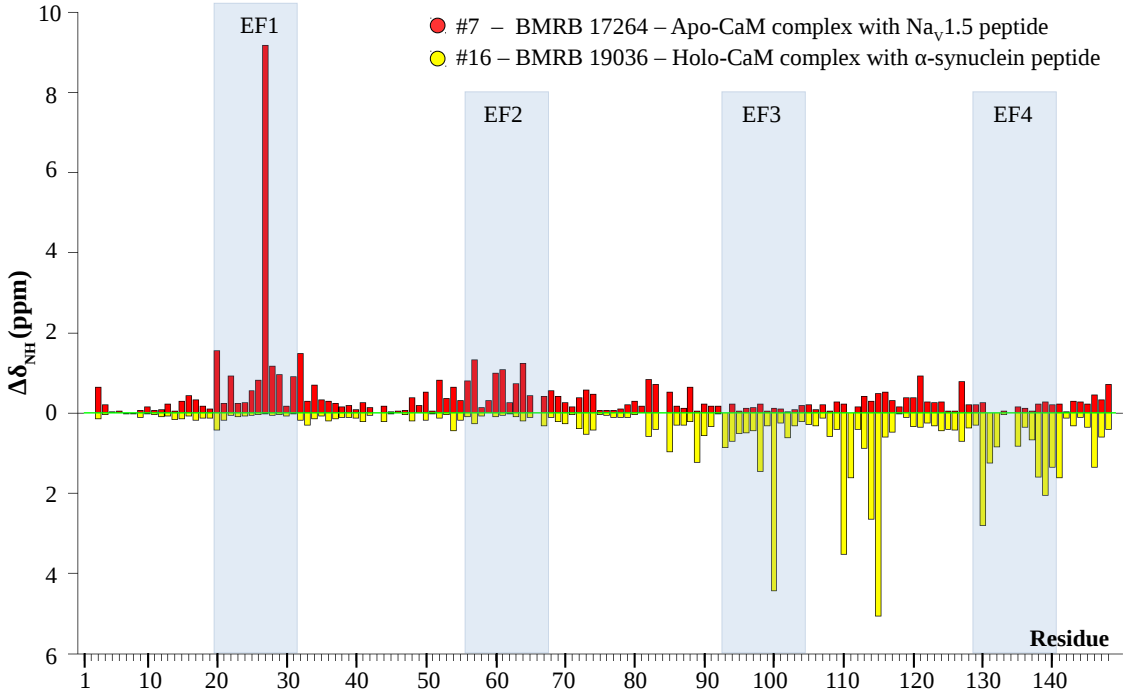


Figure 4.8. H-N chemical shift comparison of #7 and #16 entries towards reference. More differences were shown for the Apo-CaM N-lobe suggesting that CaM-K_v7.2 has Ca^{2+} in the N-lobe. In case of the C-lobe bigger differences were shown for Holo-CaM suggesting calcium absence in this lobe. Besides, F1 (20-31), EF2 (56-67), EF3 (93-104) and EF4 (129-140) were also defined as blue boxes. There are not negative values.

Summarizing, the result suggest that the CaM-K_v7.2 complex has Ca^{2+} in the N-lobe whereas the C-lobe remains in its apo-state. Nevertheless, following experiment conclusively determines the calcification state of the structure and the possible calcium effect on it.

4.3.3 Calcium Titration Analysed by NMR

Both CaM and His-Q2AB with deletions ΔF316_R325, Δ6L and ΔP533_H546 (GB116) were labelled with ^{15}N. During the purification process, protein eluted from the affinity column was divided in two and were dialysed using two different buffers: 1) *Apo Running Buffer* (120 mM KCl, mM MES pH 6, 0.5 mM EGTA), and 2) *Calcium Running Buffer* (120 mM KCl, 100 mM MES pH 6, 0.5 mM EGTA, 5 mM $CaCl_2$). Gel filtrations were done separately and showed an identical peak for both samples (Figure 4.9). Thus, calcium effect is not provoking a change in the oligomerization state and moreover, the hydrodynamic radius of the complex is not changing. After further protein purification and concentration, ^{15}N-HSQCs were collected for both samples (with and without calcium). Then samples mixtures were done to perform the titration as described in Materials and Methods (Section 4.2.4).

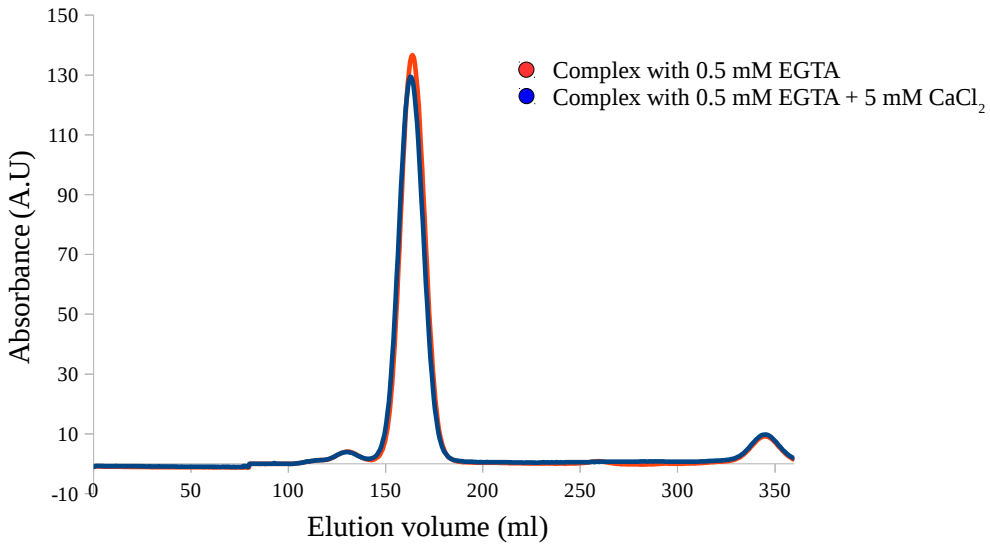


Figure 4.9. Size-Exclusion Chromatography. Superdex 26/60 column profile after injecting samples with and without calcium. Elution peaks are almost identical. Maximum peaks values were at the elution volume 162.66 ml and 163.896 ml respectively. According to a calibration curve, these elution volumes correspond to protein of 31.13 and 30.3 kDa respectively. CaM/K_v7.2_AB complex has a theoretical weight of 30.27 kDa.

4.3.3.1 Slow-Exchange dynamics uncovers three sets of signals in ^{15}N-HSQCs

Calcium titration of CaM/K_v7.2_AB complex was recorded by NMR. Protein complex was labelled with ^{15}N. Calcium was added to a final available concentrations of 0, 0.21, 0.37, 0.55, 1.52, 2.5, 3.5 and 4.5 mM. ^{15}N-HSQCs were acquired and overlapped as shown in Figure 4.10.

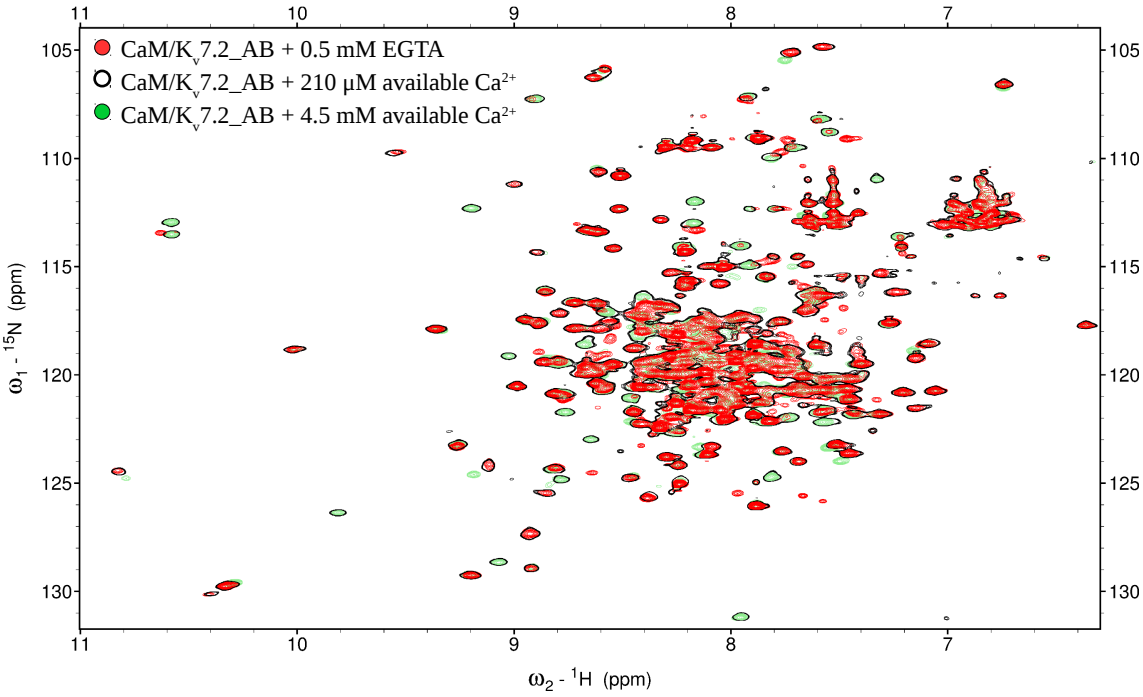


Figure 4.10. ^{15}N-HSQC overlay of three complex states. In red, "Apo"-complex spectra is shown, in pale green "Holo"-complex spectra and in black circles the intermediate spectra at 0.21 mM of available calcium. Some of the signals from the intermediate spectra overlaps with the "Apo", the "holo" or with both of them.

Calcium addition to the complex altered the ^{15}N-HSQC spectra, proving that calcium was interacting with the sample. Furthermore, due to the slow exchange regime between different complex states, three set of signals were detected for the titration:

1. Apo-CaM/K_v7.2_AB complex (without calcium): It was only visible for the first spectra.
2. Holo-CaM/K_v7.2_AB complex (with calcium): It was noticeable for concentrations of free calcium above 548 µM.
3. An intermediate state between saturated and unsaturated complexes, which only appeared at 210 and 369.25 µM.

These three sets of signals were overlapping sequentially. In other words, they were not direct changes from the "apo" to the "holo" state without going through the "intermediate" state. Moreover, the existence of these three sets of signals also provides the idea that, going back to the free-energy protein landscape, three different minima exist during Ca^{2+} signalling process. But, which CaM residues are making contact with the Ca^{2+}? Can the intermediate be defined? In which calcification state did we characterized the protein structure of the complex?

4.3.3.2 Calcium is present in the N-lobe and is absent in C-lobe of the structurally characterized CaM/K_v7.2_AB complex

In order to determine in which calmodulin areas were the interaction with Ca^{2+} taking place, the signal assignment was performed. Most of the signals for the "intermediate"-state were determined using the H-N values from the structurally characterized protein complex. However, in the assigning process, it was shown that some signals were shifted, specially in the CaM's N-lobe in the "apo"-state spectra and in the CaM's C-lobe in the "holo"-state spectra. For these signals that were dramatically moved, CaM chemical shifts from the BMRB entry 17264 and 19036 were used for "*de novo*" assigning. Nevertheless, few signals were not assigned in all the HSQCs.

When signals were assigned, Chemical Shifts Perturbation analysis was done, taking as a reference the chemical shift values from the structurally characterized CaM/K_v7.2_AB complex (Figure 4.11).

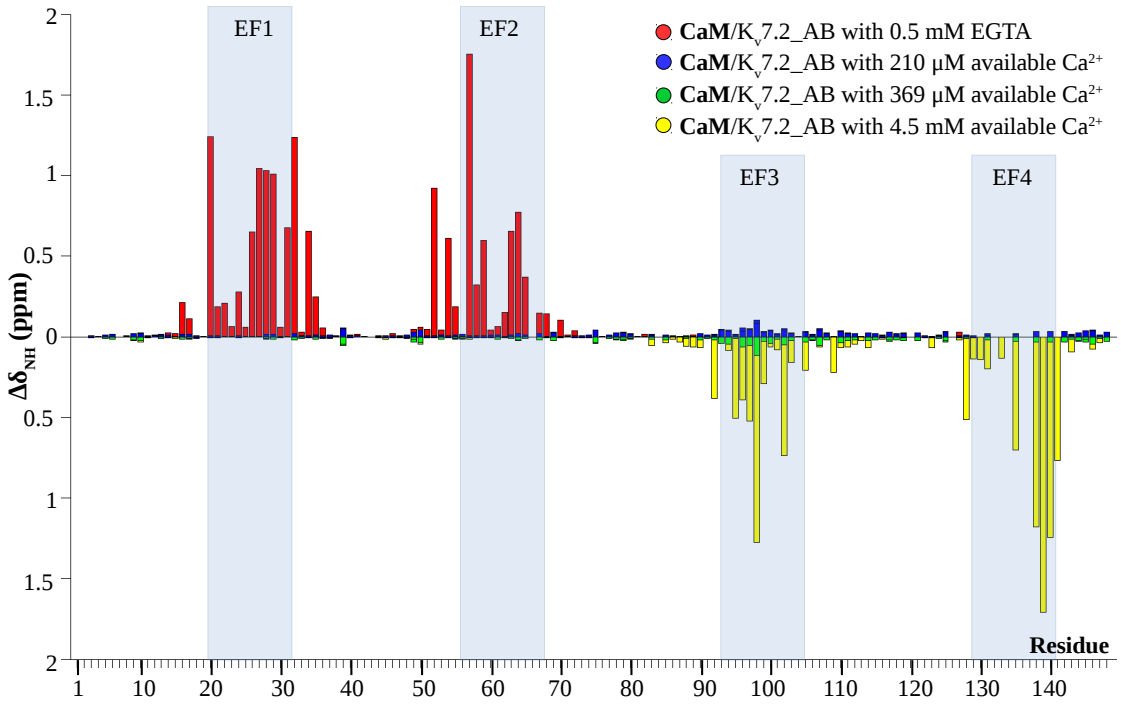


Figure 4.11. CSPs induced by calcium in CaM. CSPs from four spectra are only shown to make the interpretation easier. Apo-complex shows big chemical shift perturbations in the EF1 and EF2 hands, whereas holo-complex shows big chemical shift perturbations in the EF3 and EF4 hands. "Intermediate" states are more related to the CaM/Kv7.2_AB complex state. Divergent areas demonstrated the opposite calcification state of the lobe (EF1 and EF2 have calcium and EF3 and EF4 are without calcium.

The CSP outcome was very lightening. The closer data sets to the references were by far the "intermediate"-states reported at 210.13 and 369.25 μM of free-calcium. Regarding "apo"-state, big differences were visible at the N-lobe, specially in the EF1 and EF2 regions. On the other hand, the same effect was visible for the "holo"-state, where big differences were also appreciable in the C-lobe, specially in the EF3 and EF4 regions.

Therefore, this experiments finally concluded that the **structurally characterized CaM/K_v7.2_AB complex has a calcium ion in each EF-hand from the N-lobe**, as it demonstrates the convergence with the calcified N-lobe, **and has both EF-hands in the C-lobe empty**.

Nonetheless, these CSPs are very enclosed almost exclusively to the EF-hands. There are not big changes in the rest of the residues, concluding that the chemical environment for each residue does not change. This suggests that Ca^{2+} binding does not generate observable conformational changes a least for the calmodulin part of the CaM/K_v7.2_AB complex. But, are helices A and B changing their chemical shifts as a result of Ca^{2+} addition?

4.3.3.3 Calcium does not provoke any CSPs in AB helices

Chemical Shift Perturbations were analysed in the $K_v7.2$ helices region. As previously described, chemical shift values from the CaM/$K_v7.2$_AB complex were taken as a reference. Afterwards, values were plotted using MatLab (Figure 4.12).

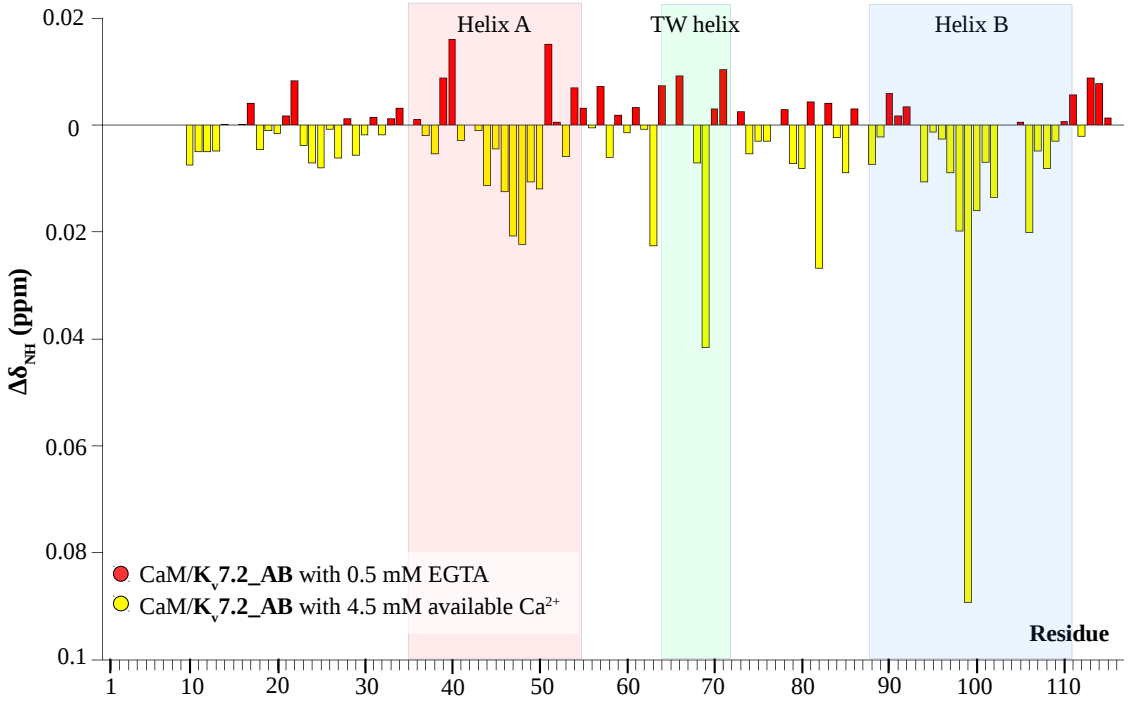


Figure 4.12. CSPs induced by calcium in $K_v7.2$ helices AB region. CSPs from two spectra are only shown to make the interpretation easier. There are not significant perturbation in the chemical shift values in any of the cases (note the scale). Helices A, TW and B are highlighted by colour boxes.

CSP map reveals that there are not significant changes in any of the regions of the $K_v7.2$. Values are small compared to CaM's EF-hands region. On one hand, the result confirms that Ca^{2+} only interacts with the CaM region of the CaM/$K_v7.2$_AB complex, as expected. On the other hand, the assay also shows that there are not big changes in the chemical shifts' environment in any of the residues from the $K_v7.2$ region, which suggest that big structural rearrangements are not taking place due to Ca^{2+} effect.

4.4 DISCUSSION

Since the beginning of this chapter, it has been empathized the importance of dynamic processes in all living cells. Models representing proteins' atomic structures are only snapshots of dynamic processes, related to a single conformation or a local minima in a protein energy-landscape (Figure 4.1). These structural rearrangements and protein dynamics upon a ligand binding, temperature change, pressure change, etc, are not easy to follow as they occur in a very different time-scales. Nevertheless, by different complementary techniques can be described in an accurate way, specially thanks to the NMR versatility.

To the best of our knowledge, **by this thesis it has been for the first time reported the calcification states of CaM complexed to K_v7.2_AB at residue level.** As we have seen, the calcification state of complexed calmodulin vary within intracellular Ca^{2+}-range (10-100 nM at resting concentrations to 1-2 μM after chemical or electrical signalling), and depending on the calcium concentrations, three states are possible:

1. Apo-state, in which calcium is not bound to any of the EF-hands but calmodulin remains wrapping helices A and B.

2. Intermediate state, where calcium is bound to the N-lobe whereas the C-lobe remains free of calcium in the presence of helices A and B. This state has been structurally characterized as it was the most populated conformation at the experimental conditions (Figure 4.13).

3. Holo-state, in which calcium is saturating CaM and this one is hugging helices A and B too.

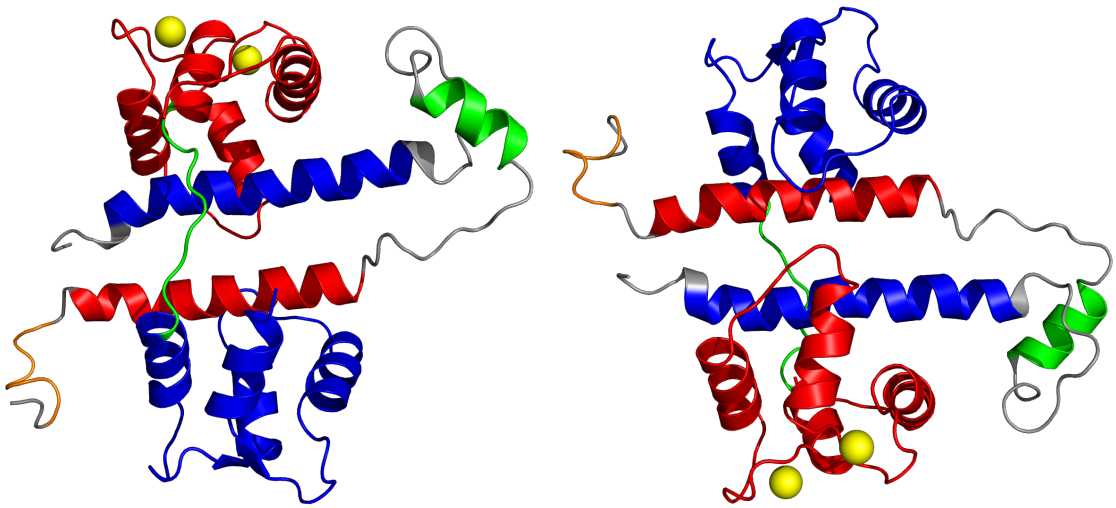


Figure 4.13. Structure characterized CaM/K_v7.2_AB complex with corresponding Ca^{2+} ions. After protein docking using HADDOCK, CaM is wrapping helices A and B. N-lobe of CaM is in red, and the C-lobe in blue. Flexible linker is represented in dark green. For K_v7.2, helix A is represented in red, and helix B in blue. TW helix is represented in green. Besides GILGSGQ region has been coloured in orange. In yellow spheres Ca^{2+} ions are shown.

Are these three states the key for calcium's regulation of M-current? For many years it has been hypothesized that the mechanism by which calcium was regulating K_v7 channels was by conformational changes derived from Ca^{2+} and CaM interaction. This idea is very appealing as it has been described for other calcium/calmodulin regulated channels. In the case of SK2 channels (Zhang et al., 2012), Ca^{2+} binding to CaM when complexed to SK2 channels creates a big conformational change. Thus, it could be easily interpretable that CaM's conformational change could be a big mechanical switch that can be turn ON or OFF closing or opening the pore depending on the calcium levels. However, drastic conformational changes are not the switch for calcium regulation of K_v7 channels as we have not been able to observed big conformational changes by NMR in helices A and B region.

Is calcium necessary for channels' regulation? Even if Apo-CaM/K_v7 complex state could be difficult to find within the cell at resting conditions, previous studies in our laboratory suggest that apo-Calmodulin is able to provoke a current potentiation in $K_v7.2$ and $K_v7.3$ channels (Gomis-Perez et al., in preparation). These experiments were done co-expressing K_v7 channels with wild-type calmodulin and CaM_{1234}, which harbours a Asp to Ala substitution in each of the EF-hands, preventing Ca^{2+} binding. Results reveal that current potentiation is independent from Ca^{2+} binding to CaM, in agreement with previous observations (Soldovieri et al., in preparation). In fact, calcium signalling may not be occurring *via* CaM pathway as reported by some groups who have observed Ca^{2+} modulation of K_v7 channels activity when using CaM_{1234} (Ghosh et al., 2006; Sachyani et al., 2014; Shamgar et al., 2006). Therefore, these outcomes are in line with our NMR data where we observe that there is not a mechanical switch for calcium regulation of M-current.

Nevertheless, for all the members of the K_v7 family apo-CaM is not sufficient for a correct channel modulation. In fact, for $K_v7.1$ channels, it has been exposed that using same conditions, the co-expression of CaM_{1234} produces currents inhibitions (Sachyani et al., 2014). Under their point of view, at least CaM's N-lobe needs to be ligated to calcium so as to stabilize CaM association to the $K_v7.1$ proximal C-terminus which is necessary for channel gating, but they also highlight that calcium is not provoking a dramatic conformational changes in $K_v7.1$ and CaM complex. Even though some of the observations are in agreement with ours, the differences regarding Apo-CaM functionally may rely on the fundamental structural and mechanistic differences between $K_v7.1$ and the rest of K_v7 family members ($K_v7.2$-$K_v7.5$) (Gamper et al., 2005).

Summing up, we know that calmodulin is a essential for channel's trafficking, gating and regulation. We also know that even if CaM's lobes have preferences, CaM can bind the proximal C-terminus of $K_v7.2$ channels in the absence or presence of calcium. Moreover, calcium binding

to CaM/K_v7.2 complex does not produce conformational changes. Channel's current potentiation is independent from Ca^{2+}. So, **what is the role of calcium?**

It is very probable that in neurons CaM/K_v7.2 complex are exploring an extensive landscape of conformations. In fact, these range of possibilities include a transient interactions that do not reach productive conformations. In this regard, calcium could be considered as a catalyst towards a final "functional" CaM/K_v7.2 configuration that discards transient interactions and selects and ensures the long-life productive interactions.

In this regard, fluorescence spectroscopy has been specially useful to determine which lobe of CaM interacts with which region of K_v7.2 channels under a certain Ca^{2+} concentrations. In general, the interaction of CaM with ion channels and other targets is dominated by the C-lobe (Chagot & Chazin, 2011; Kim et al., 2010; Wang et al., 2012). We show here that K_v7.2 does not conform to this rule. In fact, the N-lobe plays the more dominant role in apo conditions and binds preferentially to helix B. On the other hand, the C-lobe is dominant in the presence of saturating Ca^{2+} concentrations and binds preferentially to helix A. Moreover, at low concentrations, our data reveals that the binding of the lobes is not independent but that cooperativity occurs between the lobes, and when the lobes are linked together also favours their binding to K_v7.2 channels. This cooperativity could be translated to a mechanism in which one lobe of CaM binds to the channel, the other lobe will be positioned in the proximity of the binding site, thereby augmenting its effective concentration and enhancing its interactions.

If we are demonstrating by fluorescence spectroscopy that the N-lobe binds with more affinity to the helix B in the absence of Ca^{2+} and that the C-lobe binds with higher affinity to the helix A in the presence of Ca^{2+}, how can we explain that the structurally characterized CaM/K_v7.2_AB complex has the inverted states in its lobes? Are both techniques contradicting each other? Not necessarily.

As previously shown in Figure 4.4, CaM complex formation alters CaM's affinity for Ca^{2+}. Many other authors have reported this behaviour too (Sachyani et al., 2014; Zhang et al., 2012). This result does not distinguish between individual lobes, but taking into the account the experiments done by individual lobes and the CaM/K_v7.2_AB complex, it is reasonable to consider that N-lobe interaction with the helix B is driven basically by its apo-state, but when this binding occurs, N-lobe's affinity for Ca^{2+} increases dramatically, possibly stabilizing the interaction. Same case scenario could append to the C-lobe, in which Ca^{2+} increases binding affinity for helix B, but once the complex is formed it weakens the C-lobe affinity for Ca^{2+}.

Is there any kind of advantage in this lock and key process? Calcium helps in the selective recognition of each lobe to helices A or B, and moreover, the calcium affinity change of each lobe after target binding only reinforces a certain productive anchoring configuration of CaM

within the C-terminus of $K_v7.2$ channels, which is likely irreversible: For example, N-lobe binds to helix B without Ca^{2+} and the protein-protein interaction increase the Ca^{2+} affinity of the N-lobe. Once Ca^{2+} is bound to the N-lobe this configuration is sealed, in terms that calcified N-lobe can not compete with another apo-N-lobe to bind a new helix B.

Nevertheless, it is difficult to know exactly which lobe is the first making contact which the counterpart as it may depend on the intracellular calcium levels. Therefore, three starting points are employed to described the mechanism by which calmodulin is anchored to the C-terminus of K_v7 channels, which is described in Figure 4.14. Depending on the calcification state of CaM, $K_v7.2$ recognition will be leaded by N-lobe in the absence of calcium towards helix B and C-lobe in presence of calcium towards helix A. These CaM-target recognitions are considered to be transient or not productive contacts. Nevertheless, after protein-protein binding, affinity for calcium changes in each lobe, leading to a calcium reorganization in the lobes and ultimately providing the active interaction form.

At the end, three CaM/$K_v7.2$_AB states are in equilibrium where, due to the changes in lobes' affinities for Ca^{2+}, the intermediate state is the most populated one.

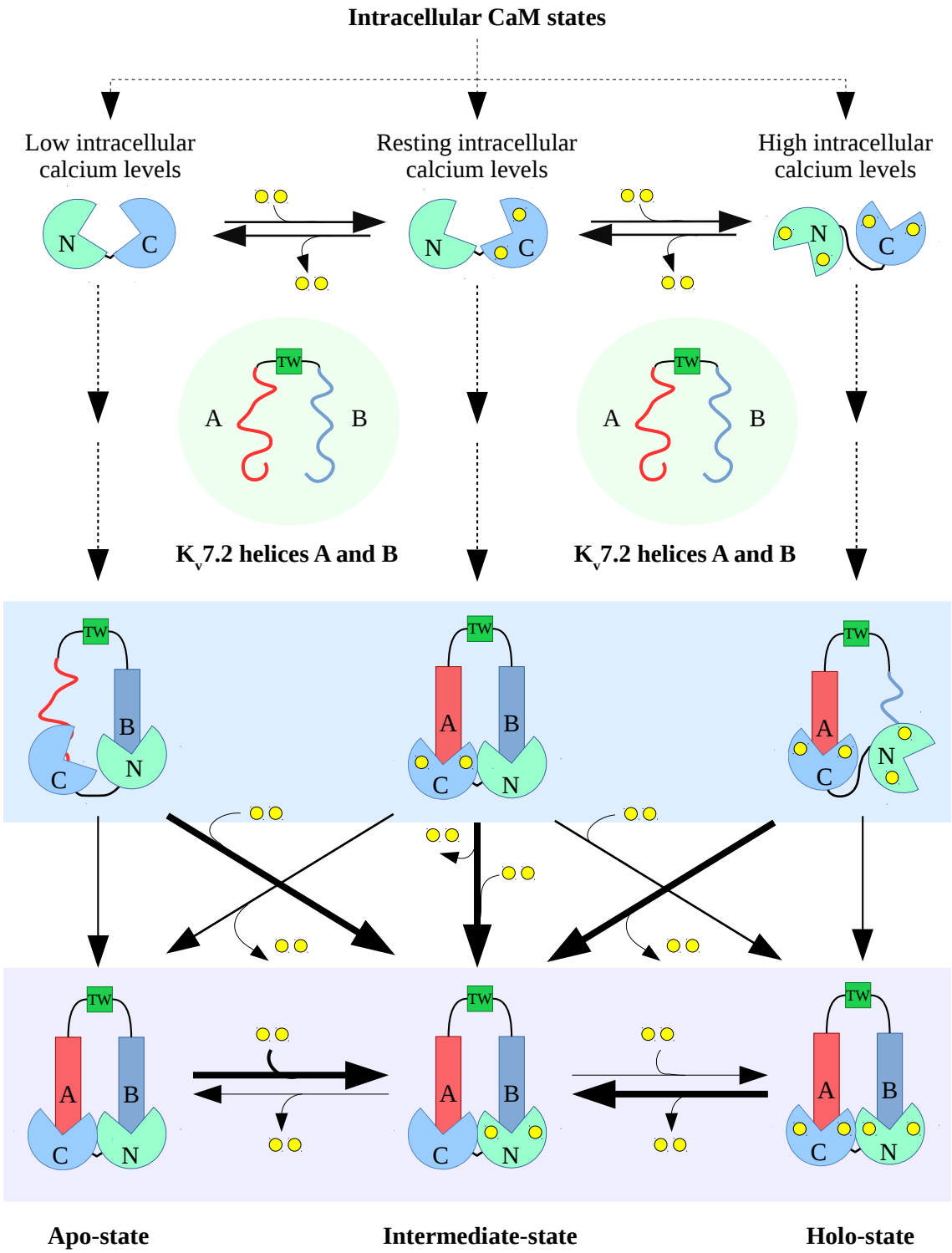


Figure 4.14. Interaction model of CaM binding to K$_v$7.2 helices A and B. Depending on the intracellular calcium concentration, CaM can be found in its Apo, Holo or intermediate state. CaM recognizes CaM binding motifs in the helices A and B, and each lobe has a preference binding site depending on CaM's calcification state. N-lobe has higher affinity to helix B in the absence of calcium, and C-lobe has higher affinity to helix A in presence of calcium. In resting intracellular levels free CaM binds helices A and B at the same time. CaM binding provokes helices A and B structuring to alpha-helices. In this first binding process there are not productive bindings (Pale blue box). Once CaM binding has done, the affinity of each lobe changes (for the N-lobe increases whereas for the C-lobe decreases). The final intermediate state is the most favourable state as it satisfies CaM's ionic reorganization. Nevertheless, a dynamic equilibrium is form between the apo and intermediate state, and intermediate and holo state (Pale purple box).

Eventually, **is there any signalling role for calcium?** And if so, **how can calcium signalling be transmitted from CaM to the channel?** On one hand we have seen that calcium is not necessary for correct M-current function, but it is considerable that calcium may have a modulation effect on it. On the other hand, we have seen that there are not structural rearrangements between any of the CaM/K_v7.2_AB states. Once the appealing mechanical switch has been discarded two other alternative mechanisms can be hypothesized.

1. Protein dynamics may change by calcium. All the experiment used in this thesis were not done with the final goal of studying protein dynamics. However, these dynamics could be perfectly reported with high accuracy by NMR experiments. The interesting fact of this idea is that even if local structure does not change after calcium addition, local dynamics may change, provoking for example in helices more flexibility or rigidity. This effect could be transferred to other contiguous regions. Taking into the account that the helix A is directly connected to the pore, it could be possible that this changes in dynamics could affect the pore creating the calcium regulation effect. In fact, recently has been reported a missense mutation in K_v7.4 channels (G321S) that may destabilize CaM binding creating a decrease in the inhibitory effects of Ca^{2+}-CaM effects on these channels (Sihn et al., 2015). Surprisingly this mutation is located in the linker between the pore and helix A which a priory does not interact directly with CaM. This article reinforces this hypothesis.

2. Signal allosterism. Complementary to previous idea, calcium signalling could be translated to other parts of the channel without the necessity of big conformational changes. As demonstrated recently by our group (Alberdi et al., 2015), disruption of coiled-coil formation by an epilepsy causing mutation decreases apparent CaM binding affinity and interrupts CaM influence on PI(4,5)P2 sensitivity. Probably helix C is orchestrating this allosteric mechanisms as it is located in between. Besides, helix C is also considered to be the “mechanical” hinge that enables the transition between close to open channel conformations (Sachyani et al., 2014). Therefore, helix C becomes an attractive target for the understanding of calcium signalling.

Overview

Summarising, CaM is a essential for correct channel trafficking, assembling and function. Here, we proposed that the different calcium affinities in the CaM's lobes may be responsible for a correct anchoring to the helices A and B of K_v7.2 channels. However, calcium signalling is not clearly understood as not conformational rearrangements are observed after calcium titration of CaM/K_v7.2_AB complex. In other to understand properly calcium regulation of the M-current it is not sufficient to draw the attention only in CaM and CaM's binding sites of the channel. Therefore, calcium effect should be studied in more extended regions of the C-terminus. In this regard, Helix C represents a good target for the further analysis as it may be implicated the allosteric mechanism where signals are translated from the pore to the more distal part of the C-terminus.

CONCLUSIONS

1. CaM is a essential protein for the correct folding and assembly of the C-terminus of K_v7.2 and K_v7.3 channels when over-expressing in *Escherichia coli*.

2. Structurally characterized CaM/K_v7.2_AB protein-protein complex shows CaM wrapping helices A and B which are forming an anti-parallel coiled-coil. The N-lobe of CaM is binding helix B and the C-lobe of CaM is binding helix A. N-lobe is calcified whereas C-lobe is not.

3. CaM/K_v7.2_AB complex can be in three states depending on the calcification state of CaM: Apo-, intermediate-, and holo-state. Calcium interacts with CaM but do not provoke any detectable conformational change in the studied protein complex.

5 RESUMEN

Introducción

Los canales de potasio, originalmente identificados en la membrana de las células nerviosas como encargados del trasiego de iones de potasio (K^+) y generadores del potencial de acción, se encuentran en la mayoría de las células que componen los seres vivos desarrollando múltiples funciones. Estas proteínas intrínsecas de membrana forman poros que permiten la difusión pasiva de iones a través de la membrana plasmática, siendo imprescindibles en el mantenimiento y regulación del potencial de membrana. De los más de 75 genes que codifican canales de potasio en humanos, solo unos pocos han sido relacionados con enfermedades. A este respecto, la familia de canales voltaje-dependientes K_v7 (KCNQ) son una excepción: hasta la fecha, de los cinco miembros de esta familia, cuatro se han visto involucrados en diferentes enfermedades o canalopatías tales como arritmias cardíacas, sordera o epilepsia, algunas de las cuales afectan al 3 % de la población mundial. La distintas enfermedades dependen en gran medida de la isoforma afectada y de su lugar de expresión. Por ejemplo, mutaciones en las subunidades $K_v7.2$ o $K_v7.3$ que se expresan exclusivamente en el sistema nervioso central, conducen a las Convulsiones Neonatales Familiares Benignas (BFNC).

En las células neuronales se ha hallado una corriente de potasio no-inactivante, dominada "corriente M" (Brown y Adams, 1980), que fluye por los canales hererodiméricos $K_v7.2$ y $K_v7.3$ (Wang et al.,1998; Shapiro et al., 2000). Juega un papel crítico en el control de la excitabilidad porque mantiene una corriente constante en el umbral de la generación del potencial de acción, lo que hace de freno. Se ha visto involucrada en procesos filológicos y patológicos de gran relevancia. Fue la primera corriente de potasio voltaje dependiente regulada por segundos mensajeros en describirse (Marrion, 1993; Selyanko et al., 1992).

Una característica común de los canales K_v7 y otros canales operados por voltaje es la existencia de seis dominios transmembranales, en cuyo cuarto segmento se encuentra el sensor de voltaje. Además, existe un lazo exterior que conforma la abertura selectiva para el poro de K^+ (lazo P). El N-terminal intracelular es relativamente corto (~100 amino ácidos) y no está implicada en el mecanismo de inactivación (los canales K_v7 no inactivan), ni en el ensamblaje con otras subunidades o tetramerización T1. En cambio, el C-terminal intracelular es extenso, y engloba alrededor de 300-500 residuos (Figura 1). Estudios realizados en nuestro laboratorio han descrito cuatro segmentos con elevada probabilidad de formar hélices alfa (A-D)(Yus-Najera et al., 2002) que representan un lugar clave para el ensamblaje, tráfico y regulación por segundos mensajeros (Haitin & Attali, 2008)

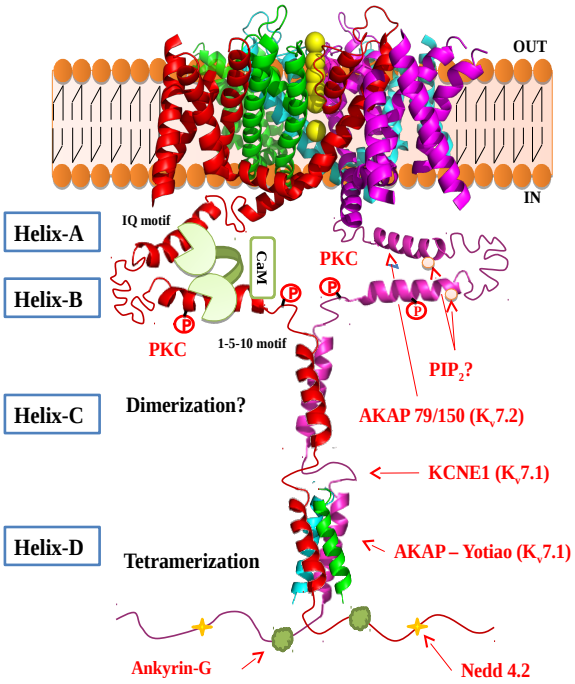


Figura 1. Esquema del los canales K_v7. Para la región intracelular solo se han representado dos subunidades. La CaM interacciona de manera constitutiva con las hélices A y B. Los sitios de unión PIP_2 pueden estar localizados en estas regiones también. Las proteínas AKAP-yotiao en $K_V7.1$, AKAP79/150 en $K_V7.2$, Ankyrin-G el $K_V7.2/3$ y Nedd4.2 en $K_V7.1$-3 también interaccionan con los canales K_v7. De la misma manera se han reportado diferentes sitios de fosforilación al canal. La hélice C puede ser la implicada en la dimerización del canal, mientras que la hélice D forma tetrámeros.

Aunque la apertura del poro es producida principalmente por los cambios de voltaje, la unión de fosfatidilinositol 4,5-bisfosfato (PIP_2) es un requisito esencial para el funcionamiento del canal. Los fosfoinosítidos son lípidos minoritarios de las membranas biológicas que son cruciales en la señalización celular.

Otro modulador de la apertura del poro es la calmodulina (CaM). Esta pequeña proteína de 148 aminoácidos es modulada por calcio y, entre otras funciones, es capaz de detectar los distintos niveles de calcio intracelular y transmitir esta información a distintas dianas. La unión de calmodulina al canal se da en la región C-terminal, concretamente gracias a los motivos de unión localizados en la hélice A (motivo IQ) y en la hélice B (motivo doble 1-5-10) (Yus-Najera et al. 2002). Se ha demostrado que la CaM es imprescindible para el ensamblaje, el trafico y la regulación de los canales K_v7 (Etxeberria et al. 2007). Mutaciones en las hélices A y B del $K_v7.2$, que dificultan la unión de la CaM y, por tanto, disminuyen su trafico a la membrana del canal, están asociadas a la enfermedad BFNC.

Objetivos

Como cabe de esperar, el estudio de las moléculas y mecanismos que están involucrados en la comunicación celular es un paso crítico para el entendimiento de estas enfermedades y la mejora de los tratamientos de patologías relacionadas. A este respecto, mientras que la comprensión de su regulación y de sus propiedades electrofisiológicas ha alcanzado niveles razonables, poco se sabe sobre la estructura y conformación que adoptan los canales de potasio K_v7, básicamente porque las técnicas que se emplean para los estudios estructurales como la cristalografía o la resonancia magnética nuclear (RMN) requieren de grandes cantidades de proteína purificada que son difíciles de alcanzar.

El propósito de la presente tesis es el aumentar los conocimientos estructurales de los canales K_v7.2 con la meta de mejorar el entendimiento de cómo se da la regulación por calcio de estos canales, y a su vez mejorar el conocimiento de los mecanismos por los que ciertas patologías ocurren. Para ello la tesis presenta tres fases: 1) Mejora de la expresión recombinante del C-terminal del canal, 2) Caracterización estructural del complejo CaM/Hélices_AB mediante RMN, y 3) Estudio del impacto del calcio sobre el complejo y los posibles mecanismos por el que se da la modulación del canal.

1. Mejora de la expresión recombinante del C-terminal del canal

Esta fase ha estado enfocada en optimizar la expresión de esta proteína de manera recombinante en *Escherichia coli*. Al igual que ocurre con otras proteínas de membrana, la producción de ésta resultaba en agregados proteicos no funcionales e insolubles. A fin de solventar el problema, se han tanteado distintas estrategias de los que destacan los siguientes: evitar los codones “raros”, emplear distintas cepas de *E. coli*, añadir a las proteínas etiquetas, búsqueda de ortólogos de los canales K_v7 que sean más sean más solubles, diseñar distintas deleciones del dominio intracelular del canal y finalmente co-expresar proteínas que interactúen con el C-terminal del canal como la calmodulina. Gracias a la combinación de todas ellas, pero en especial a la co-expresión de la CaM, se ha conseguido suficientes cantidades de proteína soluble para abordar aproximaciones estructurales.

2. Caracterización estructural del complejo CaM/K_v7.2_AB

El complejo proteico que hemos estudiado mediante RMN está compuesto por la CaM y por una construcciones de las hélices A y B que alberga tres deleciones que han sido realizadas para mejorar la solubilidad y la asignación de los desplazamientos químicos: ΔF316_R325, ΔR374_T501 y ΔP533_H546. Tal y como se describe en la tesis, la deleción del enlazador flexible que une la hélice A con la B (ΔR374_T501) no afecta a la función normal de los canales.

Utilizando distintos experimentos de RMN se ha llegado a asignar alrededor del 95 % de los residuos que forman el complejo compuesto por 264 aminoácidos. Mediante los desplazamientos químicos de la cadena central del complejo se ha identificado la estructura secundaria. Mediante el análisis del acoplamiento dipolar residual se ha obtenido información de la orientación relativa de los componentes del complejo. Las restricciones de corta-, media- y larga-distancia fueron identificadas gracias a la información proporcionada por el efecto Overhauser nuclear (NOE). Toda esta serie de restricciones han sido procesadas para determinar la estructura de cada componente del complejo utilizando el programa ARIA2 (Rieping et al., 2007), y el complejo proteico final ha sido definido mediante el programa especializado en el acoplamiento de proteínas. HADDOCK2.2 (de Vries et al., 2007).

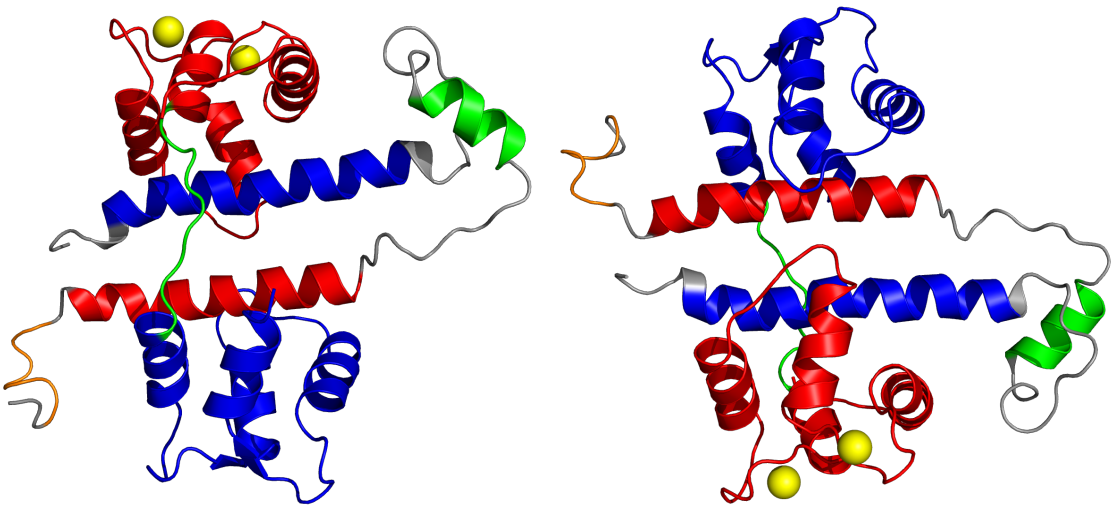


Figura 2. Modelo estructural del complejo CaM/K_v7.2_AB. La calmodulina esta abrazada a las hélices A y B del canal. Concretamente, el lóbulo N (rojo) de la CaM esta en contacto con la hélice B (azul). En cambio, el lóbulo C (azul) esta en contacto con la hélice A (rojo). Las hélices A y B están situadas de manera antiparalela. La hélice TW de esta región del C-terminal, en verde, no está en contacto con la calmodulina. En naranja, en el extremo amino, la región GILGSGQ esta coloreada en naranja. El enlazador flexible de la CaM está coloreado en verde. Nótese que solo dos iones de calcio (esferas amarillas) están unidos a la calmodulina en el lóbulo N.

El complejo muestra a la CaM íntimamente abrazando a las hélices AB del C-terminal de los canales K_v7.2. Este complejo es monomérico. La zona del C-terminal estudiada por RMN presenta cuatro regiones de interés biológico: 1) La región GILGSGQ es la responsable de unir la región intracelular con el poro situado en la membrana plasmática. De este modo, gran parte de las modulaciónes del canal originadas en el el C-terminal son transmitidas por esta región. 2) La hélice A es la región con el motivo de unión IQ a calmodulina. 3) La secuencia de la hélice TW muestra cierta similitud con otros motivos de unión a CaM. Ciertas mutaciones en esta región han sido descritas que disminuyen la afinidad de la CaM hacia el canal (Gomis-Perez et al., 2015). La estructura, en cambio, no revela ningún contacto con dicha región lo cual no quiere decir que esas mutaciones afecten de manera indirecta a la CaM. 4. La hélice B donde se alojan dos motivos de unión a CaM solapados (1-5-10). Las hélices A y B están orientadas de manera antiparalela donde hay interacciones Van der Waals que estabilizan el complejo.

La CaM muestra su típica arquitectura con dos lóbulos casi idénticos conectados por un enlazador flexible. Dos iones de calcio están presentes en el modelo, concretamente en el lóbulo N de la calmodulina. El lóbulo C, en cambio, no está calcificado.

La manera en el la formación del complejo se da es selectiva, ya que el lóbulo C tiene mayor afinidad por la hélice A en presencia de calcio y el lóbulo N por la hélice B en ausencia de calcio. Una vez que ocurre la interacción, la afinidad de los lóbulos cambia de polaridad convirtiendo a el lóbulo N en el calcificado y el lóbulo C en el libre. Posiblemente esto sea el fruto de un mecanismo con el propósito de estabilizar dicha unión.

3. Efecto del calcio y la señalización celular

Esta fase tiene como meta determinar el impacto del calcio sobre la estructura del complejo. Para ello se realiza una titración por calcio sobre el complejo y es analizado mediante RMN. En dicho experimento se pueden distinguir tres poblaciones estructurales. 1) Un estado en ausencia de calcio, 2) un estado intermediario y 3) un estado totalmente calcificado.

Las diferencias de desplazamientos químicos en los tres estados se remiten básicamente a las zonas de la calmodulina que unen calcio (manos EF). Teniendo como referencia el estado en ausencia de calcio, se observa que el calcio se une primero al lóbulo N, y después al lóbulo C. Los cambios entre estos estados son secuenciales, siempre pasando por el estado intermedio. Sorprendentemente no se aprecia ningún cambio conformacional relevante en ninguna de las regiones del C-terminal del canal. Esto descarta completamente la hipótesis en el que se cree que la calmodulina sufre un cambio conformacional drástico que conlleva la apertura o cierre del canal.

Estos datos sugieren que el mecanismo de modulación que tiene el calcio sobre el canal debe de ser algo más sutil, con lo que se hipotetizan dos posibles mecanismos.

1. Las dinámicas moleculares del complejo varían en función del calcio. Aunque no se observen cambios conformacionales aparentes las dinámicas locales pueden variar, por ejemplo rigidificando o flexibilizando ciertas regiones. Este efecto se puede transmitir a regiones contiguas transfiriendo la señal al poro.

2. Mecanismos alostéricos. Complementario con la hipótesis anterior, la señalización por calcio puede ser transmitida a otras regiones tanto del C-terminal como del poro sin necesidad de cambios conformacionales grandes. Ejemplo de ello es el articulo recientemente publicado (Alberdi et al., 2015), donde se muestra que la interrupción de la tetramerización de la hélice D del C-terminal influye en la afinidad de CaM hacia el canal y además afecta a la unión de PIP_2.

Conclusiones finales

La CaM es esencial en el trafico, ensamblaje y función de los canales $K_v7.2$. El complejo proteico caracterizado en esta tesis sirve para dar explicación a una serie de mutantes causantes de enfermedades como BFNC. Estas mutaciones suelen estar implicadas en residuos específicos en los que se da el reconocimiento de la CaM hacia el canal o en los que se da la estabilización del complejo. Mediante la titración de calcio se han podido distinguir tres estados en la calcificación del complejo. Aún así no se han observado cambios drásticos grandes en el complejo, con lo que se puede deducir que la señalización por calcio ha de ser un mecanismo más sutil, posiblemente mediante cambios en las dinámicas locales o mediante mecanismos alostéricos. Con lo cual, con el fin de entender completamente el mecanismo es necesario seguir con la investigación de lo que ocurre en regiones más distales a los de la zona de unión a CaM.

BIBLIOGRAPHY

Acton, T. B., Xiao, R., Anderson, S., Aramini, J., Buchwald, W. A., Ciccosanti, C., … Montelione, G. T. (2011). Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Methods in Enzymology*, *493*, 21–60. http://doi.org/10.1016/B978-0-12-381274-2.00002-9

Adaixo, R., Harley, C. A., Castro-Rodrigues, A. F., & Morais-Cabral, J. H. (2013). Structural properties of PAS domains from the KCNH potassium channels. *PloS One*, *8*(3), e59265. http://doi.org/10.1371/journal.pone.0059265

Aivar, P., Fernández-Orth, J., Gomis-Perez, C., Alberdi, A., Alaimo, A., Rodríguez, M. S., … Villarroel, A. (2012). Surface expression and subunit specific control of steady protein levels by the Kv7.2 helix A-B linker. *PloS One*, *7*(10), e47263. http://doi.org/10.1371/journal.pone.0047263

Alaimo, A., Alberdi, A., Gomis-Perez, C., Fernańdez-Orth, J., Bernardo-Seisdedos, G., Malo, C., … Villarroel, A. (2014). Pivoting between calmodulin lobes triggered by calcium in the Kv7.2/calmodulin complex. *PLoS ONE*, *9*(1).

Alberdi, A., Gomis-Perez, C., Bernardo-Seisdedos, G., Alaimo, A., Malo, C., Aldaregia, J., … Villarroel, A. (2015). Uncoupling PIP2-calmodulin regulation of Kv7.2 channels by an assembly de-stabilizing epileptogenic mutation. *Journal of Cell Science*. http://doi.org/10.1242/jcs.176420

Antz, C., Geyer, M., Fakler, B., Schott, M. K., Guy, H. R., Frank, R., … Kalbitzer, H. R. (1997). NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature*, *385*(6613), 272–5. http://doi.org/10.1038/385272a0

Arechaga, I., Miroux, B., Runswick, M. J., & Walker, J. E. (2003). Over-expression of Escherichia coli F1F(o)-ATPase subunit a is inhibited by instability of the uncB gene transcript. *FEBS Letters*, *547*(1-3), 97–100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12860393

Austin, C. (2003). Novel approach to obtain biologically active recombinant heterodimeric proteins in Escherichia coli. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *786*(1-2), 93–107. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12651005

Babu, Y. S., Bugg, C. E., & Cook, W. J. (1988). Structure of calmodulin refined at 2.2 A resolution. *Journal of Molecular Biology*, *204*(1), 191–204. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3145979

Banerjee, A., Lee, A., Campbell, E., & Mackinnon, R. (2013). Structure of a pore-blocking toxin in complex with a eukaryotic voltage-dependent K(+) channel. *eLife*, *2*, e00594. http://doi.org/10.7554/eLife.00594

Barhanin, J., Lesage, F., Guillemare, E., Fink, M., Lazdunski, M., & Romey, G. (1996). KvLQT1 and IsK (minK) proteins associate to form the IKS cardiac potassium current. *Nature*, *384*(6604), 78–80. http://doi.org/10.1038/384078a0

Barros, F., Domínguez, P., & de la Peña, P. (2012). Cytoplasmic Domains and Voltage-Dependent Potassium Channel Gating. *Frontiers in Pharmacology*, *3*, 49. http://doi.org/10.3389/fphar.2012.00049

Batulan, Z., Haddad, G. a, & Blunck, R. (2010). An intersubunit interaction between S4-S5 linker and S6 is responsible for the slow off-gating component in Shaker K+ channels. *The Journal of Biological Chemistry*, *285*(18), 14005–19. http://doi.org/10.1074/jbc.M109.097717

Berjanskii, M., & Wishart, D. S. (2006). NMR: prediction of protein flexibility. *Nature Protocols*, *1*(2), 683–688.

Berjanskii, M. V, Neal, S., & Wishart, D. S. (2006). PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Research*, *34*(Web Server issue), W63–9. http://doi.org/10.1093/nar/gkl341

Berjanskii, M. V., & Wishart, D. S. (2005). A simple method to predict protein flexibility using secondary chemical shifts. *Journal of the American Chemical Society*, *127*(43), 14970–14971.

Bezanilla, F. (2000). The voltage sensor in voltage-dependent ion channels. *Physiological Reviews*, *80*(2), 555–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10747201

Biervert, C. (1998). A Potassium Channel Mutation in Neonatal Human Epilepsy. *Science*, *279*(5349), 403–406. http://doi.org/10.1126/science.279.5349.403

Bradding, P., & Wulff, H. (2013). Ion channels. *Thorax*, *68*(10), 974–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23564342

Brown, D. A., & Adams, P. R. (1980). Muscarinic suppression of a novel voltage-sensitive K+ current in a vertebrate neurone. *Nature*, *283*(5748), 673–676. http://doi.org/10.1038/283673a0

Cabantous, S., & Waldo, G. S. (2006). In vivo and in vitro protein solubility assays using split GFP. *Nature Methods*, *3*(10), 845–854. http://doi.org/10.1038/nmeth932

Campbell, T. N., & Choy, F. Y. M. (2001). The Effect of pH on Green Fluorescent Protein: a Brief Review, *2*, 1–4. Retrieved from http://www.horizonpress.com/mbt/v/v2/01.pdf

Carafoli, E. (2002). Calcium signaling: a tale for all seasons. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1115–22. http://doi.org/10.1073/pnas.032427999

Carrió, M. M., & Villaverde, A. (2001). Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Letters*, *489*(1), 29–33. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11231008

Cavalli, A., Salvatella, X., Dobson, C. M., & Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9615–20. http://doi.org/10.1073/pnas.0610313104

Chagot, B., & Chazin, W. J. (2011). Solution NMR structure of Apo-calmodulin in complex with the IQ motif of human cardiac sodium channel NaV1.5. *Journal of Molecular Biology*, *406*(1), 106–19. http://doi.org/10.1016/j.jmb.2010.11.046

Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Research*, *33*(Web Server issue), W72–6. http://doi.org/10.1093/nar/gki396

Cooper, E. C., Aldape, K. D., Abosch, A., Barbaro, N. M., Berger, M. S., Peacock, W. S., … Jan, L. Y. (2000). Colocalization and coassembly of two human brain M-type potassium channel subunits that are mutated in epilepsy. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(9), 4914–9. http://doi.org/10.1073/pnas.090092797

Cornvik, T., Dahlroth, S.-L., Magnusdottir, A., Herman, M. D., Knaust, R., Ekberg, M., & Nordlund, P. (2005). Colony filtration blot: a new screening method for soluble protein expression in Escherichia coli. *Nature Methods*, *2*(7), 507–509. http://doi.org/10.1038/nmeth767

Cruzblanca, H., Koh, D. S., & Hille, B. (1998). Bradykinin inhibits M current via phospholipase C and Ca2+ release from IP3-sensitive Ca2+ stores in rat sympathetic neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(12), 7151–6. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22770&tool=pmcentrez&rendertype=abstract

de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., … Bonvin, A. M. J. J. (2007). HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, *69*(4), 726–33. http://doi.org/10.1002/prot.21723

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3), 277–293.

Delmas, P., & Brown, D. a. (2005). Pathways modulating neural KCNQ/M (Kv7) potassium channels. *Nature Reviews. Neuroscience*, *6*(11), 850–62. http://doi.org/10.1038/nrn1785

Delmas, P., Coste, B., Gamper, N., & Shapiro, M. S. (2005). Phosphoinositide lipid second messengers: new paradigms for calcium channel modulation. *Neuron*, *47*(2), 179–82. http://doi.org/10.1016/j.neuron.2005.07.001

Doreleijers, J. F., Mading, S., Maziuk, D., Sojourner, K., Yin, L., Zhu, J., … Ulrich, E. L. (2003). BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. *Journal of Biomolecular NMR*, *26*(2), 139–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12766409

Dosset, P., Hus, J.-C., Marion, D., & Blackledge, M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *Journal of Biomolecular NMR*, *20*(3), 223–231. http://doi.org/10.1023/A:1011206132740

Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics (Oxford, England)*, *21*(16), 3433–4. http://doi.org/10.1093/bioinformatics/bti541

Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., … MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science (New York, N.Y.)*, *280*(5360), 69–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9525859

Drew, D., Lerch, M., Kunji, E., Slotboom, D., & Gier, J. De. (2006). Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods*, *3*(4), 303–313. http://doi.org/10.1038/nmeth0406-303

Ducat, T., Declerck, N., Gostan, T., Kochoyan, M., & Déméné, H. (2006). Rapid Determination of Protein Solubility and Stability Conditions for NMR Studies Using Incomplete Factorial Design. *Journal of Biomolecular NMR*, *34*(3), 137–151. http://doi.org/10.1007/s10858-006-0003-0

Dunker, A. K., Lawson, J. D., Brown, C. J., Romero, P., Oh, J. S., Oldfield, C. J., … Plaza, B. (2001). Intrinsically Disordered Protein.

Dzivenu, O. K., Park, H. H., & Wu, H. (2004). General co-expression vectors for the overexpression of heterodimeric protein complexes in Escherichia coli. *Protein Expression and Purification*, *38*(1), 1–8. http://doi.org/10.1016/j.pep.2004.07.016

Eghbalnia, H. R., Wang, L., Bahrami, A., Assadi, A., & Markley, J. L. (2005). Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *Journal of Biomolecular NMR*, *32*(1), 71–81. http://doi.org/10.1007/s10858-005-5705-1

Etxeberria, A., Aivar, P., Rodriguez-Alfaro, J. A., Alaimo, A., Villacé, P., Gómez-Posada, J. C., … Villarroel, A. (2008). Calmodulin regulates the trafficking of KCNQ2 potassium channels. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *22*(4), 1135–43. http://doi.org/10.1096/fj.07-9712com

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, *10*(6), 1141–9. http://doi.org/10.1002/pmic.200900258

Falkenburger, B. H., Dickson, E. J., & Hille, B. (2013). Quantitative properties and receptor reserve of the DAG and PKC branch of G(q)-coupled receptor signaling. *The Journal of General Physiology*, *141*(5), 537–55. http://doi.org/10.1085/jgp.201210887

Frishman, D., & Argos, P. (1997). Seventy-five percent accuracy in protein secondary structure prediction. *Proteins*, *27*(3), 329–35. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9094735

Gamper, N., Li, Y., & Shapiro, M. S. (2005). Structural requirements for differential sensitivity of KCNQ K+ channels to modulation by Ca2+/calmodulin. *Molecular Biology of the Cell*, *16*(8), 3538–51. http://doi.org/10.1091/mbc.E04-09-0849

Geiger, J., Weber, Y. G., Landwehrmeyer, B., Sommer, C., & Lerche, H. (2006). Immunohistochemical analysis of KCNQ3 potassium channels in mouse brain. *Neuroscience Letters*, *400*(1-2), 101–4. http://doi.org/10.1016/j.neulet.2006.02.017

Ghosh, S., Nunziato, D. A., & Pitt, G. S. (2006). KCNQ1 assembly and function is blocked by long-QT syndrome mutations that disrupt interaction with calmodulin. *Circulation Research*, *98*(8), 1048–54. http://doi.org/10.1161/01.RES.0000218863.44140.f2

Gilling, M., Rasmussen, H. B., Calloe, K., Sequeira, A. F., Baretto, M., Oliveira, G., … Tommerup, N. (2013). Dysfunction of the Heteromeric KV7.3/KV7.5 Potassium Channel is Associated with Autism Spectrum Disorders. *Frontiers in Genetics*, *4*, 54. http://doi.org/10.3389/fgene.2013.00054

Göbl, C., & Tjandra, N. (2012). Application of Solution NMR Spectroscopy to Study Protein Dynamics. *Entropy*, *14*(12), 581–598. http://doi.org/10.3390/e14030581

Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., & Lian, L.-Y. (2004). A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society*, *126*, 8933–8939. http://doi.org/10.1021/ja049297h

Gomis-Perez, C., Alaimo, A., Fernandez-Orth, J., Alberdi, A., Aivar-Mateo, P., Bernardo-Seisdedos, G., … Villarroel, A. (2015). An unconventional calmodulin-anchoring site within the AB module of Kv7.2 channels. *Journal of Cell Science*, *128*(16), 3155–63. http://doi.org/10.1242/jcs.174128

Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in Escherichia coli. *The Protein Journal*, *32*(6), 419–25. http://doi.org/10.1007/s10930-013-9502-5

Grabarek, Z. (2006). Structural basis for diversity of the EF-hand calcium-binding proteins. *Journal of Molecular Biology*, *359*(3), 509–25. http://doi.org/10.1016/j.jmb.2006.03.066

Guerini, D., & Krebs, J. (1983). Influence of temperature and denaturing agents on the structural stability of calmodulin. *FEBS Letters*, *164*(1), 105–110. http://doi.org/10.1016/0014-5793(83)80029-5

Guerry, P., & Herrmann, T. (2012). Automated Protein Structure Determination Methods. In *NMR of Biomolecules* (pp. 536–546). Wiley-VCH Verlag GmbH & Co. KGaA. http://doi.org/10.1002/9783527644506.ch33

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, *18*(15), 2714–23. http://doi.org/10.1002/elps.1150181505

Gutman, G. A., Chandy, K. G., Grissmer, S., Lazdunski, M., McKinnon, D., Pardo, L. A., … Wang, X. (2005). International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacological Reviews*, *57*(4), 473–508.

Habibi, N., Mohd Hashim, S. Z., Norouzi, A., & Samian, M. R. (2014). A review of machine learning methods to predict the solubility of overexpressed recombinant proteins in Escherichia coli. *BMC Bioinformatics*, *15*, 134. http://doi.org/10.1186/1471-2105-15-134

Haitin, Y., & Attali, B. (2008). The C-terminus of Kv7 channels: a multifunctional module. *The Journal of Physiology*, *586*(7), 1803–10. http://doi.org/10.1113/jphysiol.2007.149187

Haitin, Y., Wiener, R., Shaham, D., Peretz, A., Cohen, E. B.-T., Shamgar, L., … Attali, B. (2009). Intracellular domains interactions and gated motions of I(KS) potassium channel subunits. *The EMBO Journal*, *28*(14), 1994–2005. http://doi.org/10.1038/emboj.2009.157

Hart, D. J., & Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in Escherichia coli. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(Pt 1), 19–26. http://doi.org/10.1107/S0907444905036097

Hattori, M., Hibbs, R. E., & Gouaux, E. (2012). A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. *Structure (London, England : 1993)*, *20*(8), 1293–9. http://doi.org/10.1016/j.str.2012.06.009

Hernandez, C. C., Zaika, O., Tolstykh, G. P., & Shapiro, M. S. (2008). Regulation of neural KCNQ channels: signalling pathways, structural motifs and functional implications. *The Journal of Physiology*, *586*(7), 1811–21. http://doi.org/10.1113/jphysiol.2007.148304

Hille, B., Dickson, E. J., Kruse, M., Vivas, O., & Suh, B.-C. (2015). Phosphoinositides regulate ion channels. *Biochimica et Biophysica Acta*, *1851*(6), 844–856. http://doi.org/10.1016/j.bbalip.2014.09.010

Hirose, S., & Noguchi, T. (2013). ESPRESSO: a system for estimating protein expression and solubility in protein expression systems. *Proteomics*, *13*(9), 1444–56. http://doi.org/10.1002/pmic.201200175

Howard, R. J., Clark, K. A., Holton, J. M., & Minor, D. L. (2007). Structural insight into KCNQ (Kv7) channel assembly and channelopathy. *Neuron*, *53*(5), 663–75. http://doi.org/10.1016/j.neuron.2007.02.010

Hung, L.-H., & Samudrala, R. (2003). Accurate and automated classification of protein secondary structure with PsiCSI. *Protein Science*, *12*(2), 288–295. http://doi.org/10.1110/ps.0222303

Hwang, P. M., Pan, J. S., & Sykes, B. D. (2014). Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli. *FEBS Letters*, *588*(2), 247–52. http://doi.org/10.1016/j.febslet.2013.09.028

Ishida, H., & Vogel, H. J. (2006). Protein-peptide interaction studies demonstrate the versatility of calmodulin target protein binding. *Protein and Peptide Letters*, *13*(5), 455–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16800798

Jentsch, T. J. (2000). Neuronal KCNQ potassium channels: physiology and role in disease. *Nature Reviews. Neuroscience*, *1*(1), 21–30. http://doi.org/10.1038/35036198

Jespersen, T., Grunnet, M., & Olesen, S.-P. (2005). The KCNQ1 potassium channel: from gene to physiological function. *Physiology (Bethesda, Md.)*, *20*, 408–416. http://doi.org/10.1152/physiol.00031.2005

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, *292*(2), 195–202. http://doi.org/10.1006/jmbi.1999.3091

Kawasaki, H., & Kretsinger, R. H. (1994). Calcium-binding proteins. 1: EF-hands. *Protein Profile*, *1*(4), 343–517. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8528904

Keeler, J. (2010). Understanding NMR Spectroscopy, 2nd Edition. Retrieved November 1, 2015, from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470746092.html

Kharkovets, T., Hardelin, J. P., Safieddine, S., Schweizer, M., El-Amraoui, A., Petit, C., & Jentsch, T. J. (2000). KCNQ4, a K+ channel mutated in a form of dominant deafness, is expressed in the inner ear and the central auditory pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4333–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18242&tool=pmcentrez&rendertype=abstract

Kim, E. Y., Rumpf, C. H., Van Petegem, F., Arant, R. J., Findeisen, F., Cooley, E. S., … Minor, D. L. (2010). Multiple C-terminal tail Ca(2+)/CaMs regulate Ca(V)1.2 function but do not mediate channel dimerization. *The EMBO Journal*, *29*(23), 3924–38. http://doi.org/10.1038/emboj.2010.260

Kincaid, R. L., Vaughan, M., Osborne, J. C., & Tkachuk, V. A. (1982). Ca2+-dependent interaction of 5-dimethylaminonaphthalene-1-sulfonyl-calmodulin with cyclic nucleotide phosphodiesterase, calcineurin, and troponin I. *The Journal of Biological Chemistry*, *257*(18), 10638–43. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6286665

Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta*, *1814*(8), 942–68. http://doi.org/10.1016/j.bbapap.2010.10.012

Kosenko, A., Kang, S., Smith, I. M., Greene, D. L., Langeberg, L. K., Scott, J. D., & Hoshi, N. (2012). Coordinated signal integration at the M-type potassium channel upon muscarinic stimulation. *The EMBO Journal*, *31*(14), 3147–56. http://doi.org/10.1038/emboj.2012.156

Kremer, W., Weyand, M., Winklmeier, A., Schreier, C., & Kalbitzer, H. R. (2013). 1.2 Å X-ray Structure of the Renal Potassium Channel Kv1.3 T1 Domain. *The Protein Journal*, *32*(7), 533–542. http://doi.org/10.1007/s10930-013-9513-2

Kreusch, A., Pfaffinger, P. J., Stevens, C. F., & Choe, S. (1998). Crystal structure of the tetramerization domain of the Shaker potassium channel. *Nature*, *392*(6679), 945–8. http://doi.org/10.1038/31978

Kubisch, C., Schroeder, B. C., Friedrich, T., Lütjohann, B., El-Amraoui, A., Marlin, S., … Jentsch, T. J. (1999). KCNQ4, a novel potassium channel expressed in sensory outer hair cells, is mutated in dominant deafness. *Cell*, *96*(3), 437–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10025409

Kwon, Y., Hofmann, T., & Montell, C. (2007). Integration of phosphoinositide- and calmodulin-mediated regulation of TRPC6. *Molecular Cell*, *25*(4), 491–503. http://doi.org/10.1016/j.molcel.2007.01.021

Labudde, D., Leitner, D., Krüger, M., & Oschkinat, H. (2003). Prediction algorithm for amino acid types with their secondary structure in proteins (PLATON) using chemical shifts. *Journal of Biomolecular NMR*, *25*(1), 41–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12566998

LAEMMLI, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, *227*(5259), 680–685. http://doi.org/10.1038/227680a0

Lange, A., Giller, K., Hornig, S., Martin-Eauclaire, M.-F., Pongs, O., Becker, S., & Baldus, M. (2006). Toxin-induced conformational changes in a potassium channel revealed by solid-state NMR. *Nature*, *440*(7086), 959–62. http://doi.org/10.1038/nature04649

Lee, J. J., & Berns, D. S. (1968). Protein aggregation. The effect of deuterium oxide on large protein aggregates of C-phycocyanin. *The Biochemical Journal*, *110*(3), 465–70. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1187373&tool=pmcentrez&rendertype=abstract

Lemercier, G., Bakalara, N., & Santarelli, X. (2003). On-column refolding of an insoluble histidine tag recombinant exopolyphosphatase from Trypanosoma brucei overexpressed in Escherichia coli. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *786*(1-2), 305–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12651027

Lerche, C., Scherer, C. R., Seebohm, G., Derst, C., Wei, A. D., Busch, A. E., & Steinmeyer, K. (2000). Molecular cloning and functional expression of KCNQ5, a potassium channel subunit that may contribute to neuronal M-current diversity. *The Journal of Biological Chemistry*, *275*(29), 22395–400. http://doi.org/10.1074/jbc.M002378200

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure (London, England : 1993)*, *11*(11), 1453–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14604535

Long, S. B., Campbell, E. B., & Mackinnon, R. (2005a). Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science (New York, N.Y.)*, *309*(5736), 897–903. http://doi.org/10.1126/science.1116269

Long, S. B., Campbell, E. B., & Mackinnon, R. (2005b). Voltage sensor of Kv1.2: structural basis of electromechanical coupling. *Science (New York, N.Y.)*, *309*(5736), 903–8. http://doi.org/10.1126/science.1116270

López-Méndez, B., & Güntert, P. (2006). Automated protein structure determination from NMR spectra. *Journal of the American Chemical Society*, *128*(40), 13112–22. http://doi.org/10.1021/ja061136l

Lorimer, G. H. (1996). A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *10*(1), 5–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8566548

Lutz, S. (2010). Beyond directed evolution--semi-rational protein engineering and design. *Current Opinion in Biotechnology*, *21*(6), 734–43. http://doi.org/10.1016/j.copbio.2010.08.011

Maeda, Y., Ueda, T., & Imoto, T. (1996). Effective renaturation of denatured and reduced immunoglobulin G in vitro without assistance of chaperone. *Protein Engineering*, *9*(1), 95–100. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9053908

Maljevic, S., Lerche, C., Seebohm, G., Alekov, A. K., Busch, A. E., & Lerche, H. (2003). C-terminal interaction of KCNQ2 and KCNQ3 K+ channels. *The Journal of Physiology*, *548*(Pt 2), 353–60. http://doi.org/10.1113/jphysiol.2003.040980

Maljevic, S., Wuttke, T. V, & Lerche, H. (2008). Nervous system KV7 disorders: breakdown of a subthreshold brake. *The Journal of Physiology*, *586*(7), 1791–1801.

Maljevic, S., Wuttke, T. V, Seebohm, G., & Lerche, H. (2010). KV7 channelopathies. *Pflügers Archiv : European Journal of Physiology*, *460*(2), 277–88. http://doi.org/10.1007/s00424-010-0831-3

Marshall, C. B., Nishikawa, T., Osawa, M., Stathopulos, P. B., & Ikura, M. (2015). Calmodulin and STIM proteins: Two major calcium sensors in the cytoplasm and endoplasmic reticulum. *Biochemical and Biophysical Research Communications*, *460*(1), 5–21. http://doi.org/10.1016/j.bbrc.2015.01.106

Maxwell, K. L., Mittermaier, A. K., Forman-Kay, J. D., & Davidson, A. R. (1999). A simple in vivo assay for increased protein solubility. *Protein Science : A Publication of the Protein Society*, *8*(9), 1908–11. http://doi.org/10.1110/ps.8.9.1908

Maylie, J., Bond, C. T., Herson, P. S., Lee, W.-S., & Adelman, J. P. (2004). Small conductance Ca2+-activated K+ channels and calmodulin. *The Journal of Physiology*, *554*(Pt 2), 255–61. http://doi.org/10.1113/jphysiol.2003.049072

Miceli, F., Soldovieri, M. V., Martire, M., & Taglialatela, M. (2008). Molecular pharmacology and therapeutic potential of neuronal Kv7-modulating drugs. *Current Opinion in Pharmacology*, *8*(1), 65–74. http://doi.org/10.1016/j.coph.2007.10.003

Miranda, P., Cadaveira-Mosquera, A., González-Montelongo, R., Villarroel, A., González-Hernández, T., Lamas, J. A., … Giraldez, T. (2013). The neuronal serum- and glucocorticoid-regulated kinase 1.1 reduces neuronal excitability and protects against seizures through upregulation of the M-current. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(6), 2684–96. http://doi.org/10.1523/JNEUROSCI.3442-12.2013

Morais Cabral, J. H., Lee, A., Cohen, S. L., Chait, B. T., Li, M., & Mackinnon, R. (1998). Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain. *Cell*, *95*(5), 649–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9845367

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., & Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology*, *20*(1), 87–90. http://doi.org/10.1038/nbt0102-87

Neal, S., Berjanskii, M., Zhang, H., & Wishart, D. S. (2006). Accurate prediction of protein torsion angles using chemical shifts and sequence homology. *Magnetic Resonance in Chemistry : MRC*, *44 Spec No*, S158–67. http://doi.org/10.1002/mrc.1832

Neubauer, B. A., Waldegger, S., Heinzinger, J., Hahn, A., Kurlemann, G., Fiedler, B., … Sander, T. (2008). KCNQ2 and KCNQ3 mutations contribute to different idiopathic epilepsy syndromes. *Neurology*, *71*(3), 177–83. http://doi.org/10.1212/01.wnl.0000317090.92185.ec

Ng, C. A., Hunter, M. J., Perry, M. D., Mobli, M., Ke, Y., Kuchel, P. W., … Vandenberg, J. I. (2011). The N-terminal tail of hERG contains an amphipathic α-helix that regulates channel deactivation. *PloS One*, *6*(1), e16191. http://doi.org/10.1371/journal.pone.0016191

Nishihara, K., Kanemori, M., Yanagi, H., & Yura, T. (2000). Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Applied and Environmental Microbiology*, *66*(3), 884–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=91917&tool=pmcentrez&rendertype=abstract

Ogden, D., & Stanfield, P. (1981). Patch clamp techniques for single channel and whole-cell recording. *Currents*, *2*(7), 53–78. Retrieved from http://www.utdallas.edu/~tres/microelectrode/microelectrodes_ch04.pdf

Okada, M., Zhu, G., Hirose, S., Ito, K. I., Murakami, T., Wakui, M., & Kaneko, S. (2003). Age-dependent modulation of hippocampal excitability by KCNQ-channels. *Epilepsy Research*, *53*(1-2), 81–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12576170

Petsko, G. a., & Ringe, D. (2004). Protein Structure and Function. *New Science Press Ltd*, 195. http://doi.org/10.1021/ja209464f

Phizicky, E. M., & Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiological Reviews*, *59*(1), 94–123. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=239356&tool=pmcentrez&rendertype=abstract

Pioletti, M., Findeisen, F., Hura, G. L., & Minor, D. L. (2006). Three-dimensional structure of the KChIP1-Kv4.3 T1 complex reveals a cross-shaped octamer. *Nature Structural & Molecular Biology*, *13*(11), 987–95. http://doi.org/10.1038/nsmb1164

Piserchio, A., Pellegrini, M., Mehta, S., Blackman, S. M., Garcia, E. P., Marshall, J., & Mierke, D. F. (2002). The PDZ1 domain of SAP90. Characterization of structure and binding. *The Journal of Biological Chemistry*, *277*(9), 6967–73. http://doi.org/10.1074/jbc.M109453200

Price, W. N., Chen, Y., Handelman, S. K., Neely, H., Manor, P., Karlin, R., … Hunt, J. F. (2009). Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nature Biotechnology*, *27*(1), 51–7. http://doi.org/10.1038/nbt.1514

Puigbò, P., Guzmán, E., Romeu, A., & Garcia-Vallvé, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research*, *35*(Web Server issue), W126–31. http://doi.org/10.1093/nar/gkm219

Rai, M., & Padh, H. (2001). Expression systems for production of heterologous proteins, *80*(9).

Rett, A., & Teubel, R. (1964). Neugeborenen Krämpfe im Rahmen einer epileptisch belasteten Familie. *Wiener Klinische Wochenschrift*, *76*, 609–613.

Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., & Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics (Oxford, England)*, *23*(3), 381–2. http://doi.org/10.1093/bioinformatics/btl589

Roberts, G. (2011). Protein NMR Spectroscopy: Practical Techniques and Applications - Gordon Roberts, Lu-Yun Lian. Retrieved November 1, 2015, from http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470721936.html

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E., & Dunker, a. K. (1997). Identifying disordered regions in proteins from amino acid sequence. *Proceedings of International Conference on Neural Networks (ICNN'97)*, *1*, 90–95. http://doi.org/10.1109/ICNN.1997.611643

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*, 172. http://doi.org/10.3389/fmicb.2014.00172

Roura-Ferrer, M., Etxebarria, A., Solé, L., Oliveras, A., Comes, N., Villarroel, A., & Felipe, A. (2009). Functional implications of KCNE subunit expression for the Kv7.5 (KCNQ5) channel. *Cellular Physiology and Biochemistry : International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, *24*(5-6), 325–34. http://doi.org/10.1159/000257425

Rückert, M., & Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *Journal of the American Chemical Society*, *122*(32), 7793–7797. http://doi.org/10.1021/ja001068h

Sachyani, D., Dvir, M., Strulovich, R., Tria, G., Tobelaim, W., Peretz, A., … Hirsch, J. A. (2014). Structural basis of a Kv7.1 potassium channel gating module: studies of the intracellular. *Structure (London, England : 1993)*, *22*(11), 1582–94. http://doi.org/10.1016/j.str.2014.07.016

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., … Ito, Y. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature*, *458*(7234), 102–5. http://doi.org/10.1038/nature07814

Sanguinetti, M. C., Curran, M. E., Zou, A., Shen, J., Spector, P. S., Atkinson, D. L., & Keating, M. T. (1996). Coassembly of K(V)LQT1 and minK (IsK) proteins to form cardiac I(Ks) potassium channel. *Nature*, *384*(6604), 80–3. http://doi.org/10.1038/384080a0

Sansom, M. S. P., Shrivastava, I. H., Bright, J. N., Tate, J., Capener, C. E., & Biggin, P. C. (2002). Potassium channels: structures, models, simulations. *Biochimica et Biophysica Acta*, *1565*(2), 294–307. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12409202

Santagata, S., Boggon, T. J., Baird, C. L., Gomez, C. A., Zhao, J., Shan, W. S., … Shapiro, L. (2001). G-protein signaling through tubby proteins. *Science (New York, N.Y.)*, *292*(5524), 2041–50. http://doi.org/10.1126/science.1061233

Scannevin, R. H., Wang, K., Jow, F., Megules, J., Kopsco, D. C., Edris, W., … Rhodes, K. J. (2004). Two N-terminal domains of Kv4 K(+) channels regulate binding to and modulation by KChIP1. *Neuron*, *41*(4), 587–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14980207

Schmidt, C., Irausquin, S. J., & Valafar, H. (2013). Advances in the REDCAT software package. *BMC Bioinformatics*, *14*(1), 302. http://doi.org/10.1186/1471-2105-14-302

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. http://doi.org/10.1038/nmeth.2089

Schoenmakers, T. J., Visser, G. J., Flik, G., & Theuvenet, A. P. (1992). CHELATOR: an improved method for computing metal ion concentrations in physiological solutions. *BioTechniques*, *12*(6), 870–4, 876–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1642895

Schroeder, B. C., Hechenberger, M., Weinreich, F., Kubisch, C., & Jentsch, T. J. (2000). KCNQ5, a novel potassium channel broadly expressed in brain, mediates M-type currents. *The Journal of Biological Chemistry*, *275*(31), 24089–95. http://doi.org/10.1074/jbc.M003245200

Schroeder, B. C., Kubisch, C., Stein, V., & Jentsch, T. J. (1998). Moderate loss of function of cyclic-AMP-modulated KCNQ2/KCNQ3 K+ channels causes epilepsy. *Nature*, *396*(6712), 687–690. http://doi.org/10.1038/25367

Schwake, M., Pusch, M., Kharkovets, T., & Jentsch, T. J. (2000). Surface expression and single channel properties of KCNQ2/KCNQ3, M-type K+ channels involved in epilepsy. *The Journal of Biological Chemistry*, *275*(18), 13343–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10788442

Schwarz, J. R., Glassmeier, G., Cooper, E. C., Kao, T.-C., Nodera, H., Tabuena, D., … Bostock, H. (2006). KCNQ channels mediate IKs, a slow K+ current regulating excitability in the rat node of Ranvier. *The Journal of Physiology*, *573*(Pt 1), 17–34. http://doi.org/10.1113/jphysiol.2006.106815

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Marius Clore, G. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, *160*(1), 65–73. http://doi.org/10.1016/S1090-7807(02)00014-9

Selyanko, A. A., & Brown, D. A. (1996). Intracellular calcium directly inhibits potassium M channels in excised membrane patches from rat sympathetic neurons. *Neuron*, *16*(1), 151–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8562079

Sen, T. Z., Jernigan, R. L., Garnier, J., & Kloczkowski, A. (2005). GOR V server for protein secondary structure prediction. *Bioinformatics (Oxford, England)*, *21*(11), 2787–8. http://doi.org/10.1093/bioinformatics/bti408

Shamgar, L., Ma, L., Schmitt, N., Haitin, Y., Peretz, A., Wiener, R., … Attali, B. (2006). Calmodulin is essential for cardiac IKS channel gating and assembly: impaired function in long-QT mutations. *Circulation Research*, *98*(8), 1055–63. http://doi.org/10.1161/01.RES.0000218979.40770.69

Shapiro, M. S., Roche, J. P., Kaftan, E. J., Cruzblanca, H., Mackie, K., & Hille, B. (2000). Reconstitution of muscarinic modulation of the KCNQ2/KCNQ3 K(+) channels that underlie the neuronal M current. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(5), 1710–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10684873

Shen, Y., Delaglio, F., Cornilescu, G., & Bax, a. (2009). TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR*, *44*(4), 213–223. http://doi.org/10.1007/s10858-009-9333-z

Sihn, C.-R., Kim, H. J., Woltz, R. L., Yarov-Yarovoy, V., Yang, P.-C., Xu, J., … Yamoah, E. N. (2015). Mechanisms of Calmodulin Regulation of Different Isoforms of Kv7.4 K+ Channels. *The Journal of Biological Chemistry*. http://doi.org/10.1074/jbc.M115.668236

Sivanandam, V. N., & Millet, O. (2015). NMR Spectroscopy for Monitoring Functional Dynamics in Solution. In *eLS*. John Wiley & Sons, Ltd. http://doi.org/10.1002/9780470015902.a0003104.pub2

Smialowski, P., Doose, G., Torkler, P., Kaufmann, S., & Frishman, D. (2012). PROSO II--a new method for protein solubility prediction. *The FEBS Journal*, *279*(12), 2192–200. http://doi.org/10.1111/j.1742-4658.2012.08603.x

Soldovieri, M. V., Boutry-Kryza, N., Milh, M., Doummar, D., Heron, B., Bourel, E., … Lesca, G. (2014). 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. *Human Mutation*, *35*(3), 356–67. http://doi.org/10.1002/humu.22500

Soldovieri, M. V., Cilio, M. R., Miceli, F., Bellini, G., Miraglia del Giudice, E., Castaldo, P., … Taglialatela, M. (2007). Atypical gating of M-type potassium channels conferred by mutations in uncharged residues in the S4

region of KCNQ2 causing benign familial neonatal convulsions. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(18), 4919–28. http://doi.org/10.1523/JNEUROSCI.0580-07.2007

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*(1), 1. http://doi.org/10.1186/1475-2859-4-1

Sormanni, P., Aprile, F. A., & Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. *Journal of Molecular Biology*, *427*(2), 478–90. http://doi.org/10.1016/j.jmb.2014.09.026

Spratt, D. E., Newman, E., Mosher, J., Ghosh, D. K., Salerno, J. C., & Guillemette, J. G. (2006). Binding and activation of nitric oxide synthase isozymes by calmodulin EF hand pairs. *The FEBS Journal*, *273*(8), 1759–71. http://doi.org/10.1111/j.1742-4658.2006.05193.x

Steinlein, O. K., Conrad, C., & Weidner, B. (2007). Benign familial neonatal convulsions: always benign? *Epilepsy Research*, *73*(3), 245–9. http://doi.org/10.1016/j.eplepsyres.2006.10.010

Strong, P. N. (1990). Potassium channel toxins. *Pharmacology & Therapeutics*, *46*(1), 137–162. http://doi.org/10.1016/0163-7258(90)90040-9

Suh, B.-C., & Hille, B. (2002). Recovery from muscarinic modulation of M current channels requires phosphatidylinositol 4,5-bisphosphate synthesis. *Neuron*, *35*(3), 507–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12165472

Tam, S. W., & Zaczek, R. (1995). Linopirdine. A depolarization-activated releaser of transmitters for treatment of dementia. *Advances in Experimental Medicine and Biology*, *363*, 47–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7618529

Thompson, A. N., Kim, I., Panosian, T. D., Iverson, T. M., Allen, T. W., & Nimigean, C. M. (2009). Mechanism of potassium-channel selectivity revealed by Na(+) and Li(+) binding sites within the KcsA pore. *Nature Structural & Molecular Biology*, *16*(12), 1317–24. http://doi.org/10.1038/nsmb.1703

Tobin, M. B., Gustafsson, C., & Huisman, G. W. (2000). Directed evolution: the “rational” basis for “irrational” design. *Current Opinion in Structural Biology*, *10*(4), 421–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10981629

Van Petegem, F., Clark, K. A., Chatelain, F. C., & Minor, D. L. (2004). Structure of a complex between a voltage-gated calcium channel β-subunit and an α-subunit domain. *Nature*, *429*(6992), 671–675. http://doi.org/10.1038/nature02588

Vernon, R., Shen, Y., Baker, D., & Lange, O. F. (2013). Improved chemical shift based fragment selection for CS-Rosetta using Rosetta3 fragment picker. *Journal of Biomolecular NMR*, *57*(2), 117–127.

Villarroel, A., Taglialatela, M., Bernardo-Seisdedos, G., Alaimo, A., Agirre, J., Alberdi, A., … Areso, P. (2014b). The ever changing moods of calmodulin: How structural plasticity entails transductional adaptability. *Journal of Molecular Biology*. Academic Press.

Waldo, G. S., Standish, B. M., Berendzen, J., & Terwilliger, T. C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*, *17*(7), 691–5. http://doi.org/10.1038/10904

Wang, C., Chung, B. C., Yan, H., Lee, S., & Pitt, G. S. (2012). Crystal Structure of the Ternary Complex of a NaV C-Terminal Domain, a Fibroblast Growth Factor Homologous Factor, and Calmodulin. *Structure/Folding and Design*, *20*(7), 1167–1176. http://doi.org/10.1016/j.str.2012.05.001

Wang, C.-C., Chen, J.-H., Lai, W.-C., & Chuang, W.-J. (2007). 2DCSi: identification of protein secondary structure and redox state using 2D cluster analysis of NMR chemical shifts. *Journal of Biomolecular NMR*, *38*(1), 57–63. http://doi.org/10.1007/s10858-007-9146-x

Wang, H. (1998). KCNQ2 and KCNQ3 Potassium Channel Subunits: Molecular Correlates of the M-Channel. *Science*, *282*(5395), 1890–1893. http://doi.org/10.1126/science.282.5395.1890

Wang, Q., Curran, M. E., Splawski, I., Burn, T. C., Millholland, J. M., VanRaay, T. J., … Keating, M. T. (1996). Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nature Genetics*, *12*(1), 17–23. http://doi.org/10.1038/ng0196-17

Wang, Y., & Jardetzky, O. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Science : A Publication of the Protein Society*, *11*(4), 852–61. http://doi.org/10.1110/ps.3180102

Wassenaar, T. A., van Dijk, M., Loureiro-Ferreira, N., van der Schot, G., de Vries, S. J., Schmitz, C., … Bonvin, A. M. J. J. (2012). WeNMR: Structural Biology on the Grid. *Journal of Grid Computing*, *10*(4), 743–767.

Weber, C., Mello de Queiroz, F., Downie, B. R., Suckow, A., Stühmer, W., & Pardo, L. A. (2006). Silencing the activity and proliferative properties of the human EagI Potassium Channel by RNA Interference. *The Journal of Biological Chemistry*, *281*(19), 13030–7. http://doi.org/10.1074/jbc.M600883200

Weckhuysen, S., Mandelstam, S., Suls, A., Audenaert, D., Deconinck, T., Claes, L. R. F., … de Jonghe, P. (2012). KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. *Annals of Neurology*, *71*(1), 15–25. http://doi.org/10.1002/ana.22644

Wen, H., & Levitan, I. B. (2002). Calmodulin is an auxiliary subunit of KCNQ2/3 potassium channels. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(18), 7991–8001. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12223552

Wiener, R., Haitin, Y., Shamgar, L., Fernández-Alonso, M. C., Martos, A., Chomsky-Hecht, O., … Hirsch, J. a. (2008). The KCNQ1 (Kv7.1) COOH terminus, a multitiered scaffold for subunit assembly and protein interaction. *The Journal of Biological Chemistry*, *283*(9), 5815–30. http://doi.org/10.1074/jbc.M707541200

Wishart, D. S. (2011). Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *58*(1-2), 62–87. http://doi.org/10.1016/j.pnmrs.2010.07.004

Wishart, D. S., Arndt, D., Berjanskii, M., Tang, P., Zhou, J., & Lin, G. (2008). CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Research*, *36*(Web Server issue), W496–502. http://doi.org/10.1093/nar/gkn305

Wissmann, R., Bildl, W., Oliver, D., Beyermann, M., Kalbitzer, H.-R., Bentrop, D., & Fakler, B. (2003). Solution structure and function of the “tandem inactivation domain” of the neuronal A-type potassium channel Kv1.4. *The Journal of Biological Chemistry*, *278*(18), 16142–50. http://doi.org/10.1074/jbc.M210191200

Wuttke, T. V, Penzien, J., Fauler, M., Seebohm, G., Lehmann-Horn, F., Lerche, H., & Jurkat-Rott, K. (2008). Neutralization of a negative charge in the S1-S2 region of the KV7.2 (KCNQ2) channel affects voltage-dependent activation in neonatal epilepsy. *The Journal of Physiology*, *586*(2), 545–55. http://doi.org/10.1113/jphysiol.2007.143826

Xiao, H., Bao, Z., & Zhao, H. (2015). High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Industrial & Engineering Chemistry Research*, *54*(16), 4011–4020. http://doi.org/10.1021/ie503060a

Xie, Y., & Wetlaufer, D. B. (1996). Control of aggregation in protein refolding: the temperature-leap tactic. *Protein Science : A Publication of the Protein Society*, *5*(3), 517–23. http://doi.org/10.1002/pro.5560050314

Xu, C., Watras, J., & Loew, L. M. (2003). Kinetic analysis of receptor-activated phosphoinositide turnover. *The Journal of Cell Biology*, *161*(4), 779–791. http://doi.org/10.1083/jcb.200301070

Xu, D., Jaroszewski, L., Li, Z., & Godzik, A. (2014). FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics (Oxford, England)*, *30*(5), 660–7. http://doi.org/10.1093/bioinformatics/btt578

Xu, Q., Chang, A., Tolia, A., & Jr, D. L. M. (2013). Structure of a Ca2+/CaM:Kv7.4 (KCNQ4) B-Helix Complex Provides Insight into M Current Modulation. *Journal of Molecular Biology*, *425*(2), 378–394. http://doi.org/10.1016/j.jmb.2012.11.023

Xu, Q., Chang, A., Tolia, A., & Minor, D. L. (2013). Structure of a Ca(2+)/CaM:Kv7.4 (KCNQ4) B-helix complex provides insight into M current modulation. *Journal of Molecular Biology*, *425*(2), 378–94. http://doi.org/10.1016/j.jmb.2012.11.023

Yang, W. P., Levesque, P. C., Little, W. A., Conder, M. L., Ramakrishnan, P., Neubauer, M. G., & Blanar, M. A. (1998). Functional expression of two KvLQT1-related potassium channels responsible for an inherited

idiopathic epilepsy. *The Journal of Biological Chemistry*, *273*(31), 19419–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9677360

Yu, F. H., Yarov-Yarovoy, V., Gutman, G. A., & Catterall, W. A. (2005). Overview of molecular relationships in the voltage-gated ion channel superfamily. *Pharmacological Reviews*, *57*(4), 387–95. http://doi.org/10.1124/pr.57.4.13

Yus-Najera, E., Santana-Castro, I., & Villarroel, A. (2002). The identification and characterization of a noncontinuous calmodulin-binding site in noninactivating voltage-dependent KCNQ potassium channels. *The Journal of Biological Chemistry*, *277*(32), 28545–53. http://doi.org/10.1074/jbc.M204130200

Zaika, O., Lara, L. S., Gamper, N., Hilgemann, D. W., Jaffe, D. B., & Shapiro, M. S. (2006). Angiotensin II regulates neuronal excitability via phosphatidylinositol 4,5-bisphosphate-dependent modulation of Kv7 (M-type) K+ channels. *The Journal of Physiology*, *575*(Pt 1), 49–67. http://doi.org/10.1113/jphysiol.2006.114074

Zanier, K., Nominé, Y., Charbonnier, S., Ruhlmann, C., Schultz, P., Schweizer, J., & Travé, G. (2007). Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species. *Protein Expression and Purification*, *51*(1), 59–70. http://doi.org/10.1016/j.pep.2006.07.029

Zaydman, M. A., & Cui, J. (2014). PIP2 regulation of KCNQ channels: biophysical and molecular mechanisms for lipid modulation of voltage-dependent gating. *Frontiers in Physiology*, *5*, 195. http://doi.org/10.3389/fphys.2014.00195

Zhang, M., Abrams, C., Wang, L., Gizzi, A., He, L., Lin, R., … Zhang, J. (2012). Structural basis for calmodulin as a dynamic calcium sensor. *Structure (London, England : 1993)*, *20*(5), 911–23. http://doi.org/10.1016/j.str.2012.03.019

Zweckstetter, M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nature Protocols*, *3*(4), 679–690. http://doi.org/10.1038/nprot.2008.36

PUBLICATIONS

Alberdi, A., Gomis-Perez, C., Bernardo-Seisdedos, G., Alaimo, A., Malo, C., Aldaregia, J., … Villarroel, A. (2015). Uncoupling PIP2-calmodulin regulation of Kv7.2 channels by an assembly de-stabilizing epileptogenic mutation. *Journal of Cell Science*. http://doi.org/10.1242/jcs.176420

Gomis-Perez, C., Alaimo, A., Fernandez-Orth, J., Alberdi, A., Aivar-Mateo, P., Bernardo-Seisdedos, G., … Villarroel, A. (2015). An unconventional calmodulin-anchoring site within the AB module of Kv7.2 channels. *Journal of Cell Science*, *128*(16), 3155–63. http://doi.org/10.1242/jcs.174128

Villarroel, A., Taglialatela, M., Bernardo-Seisdedos, G., Alaimo, A., Agirre, J., Alberdi, A., … Areso, P. (2014). The ever changing moods of calmodulin: How structural plasticity entails transductional adaptability. *Journal of Molecular Biology*. Academic Press.

Alaimo, A., Alberdi, A., Gomis-Perez, C., Fernańdez-Orth, J., Bernardo-Seisdedos, G., Malo, C., … Villarroel, A. (2014). Pivoting between calmodulin lobes triggered by calcium in the Kv7.2/calmodulin complex. *PLoS ONE*, *9*(1).

APPENDIX

Due to the length of the following files can be found attached in electronic format:

- Chemical shifts for CaM in BMRB format.

- Chemical shifts for K_v7.2_AB in BMRB format.

- Residual Dipolar Couplings in XEASY format.

- ARIA2 output.

- PDB file for CaM.

- PDB file for K_v7.2_AB.

- HADDOCK output.

ABBREVIATIONS

AA	Amino-acid
AKAP	A-kinase anchor proteins
Amp	Ampicillin
AP	Action Potential
ApoCaM	Apo-calmodulin
ATP	Adenosine 5'-triphosphate
BFNC	Benign familial neonatal convulsions
CaMBD	Calmodulin binding domain
CaM	Calmodulin
cAMP	cyclic adenosine 5’-monophosphate
cDNA	Complementary DNA
CVs	Colum volumes
DAG	Diacylglicerol
D-CaM	Dansyl-calmodulin
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
Ec_{50}	Half maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid

GB116	Helices A and B region with ΔF316_R325, ΔR374_T501 and ΔP533_H546.
GST	Glutathione S-transferase
h	Hour
HoloCaM	holo-calmodulin
HSQC	Heteronuclear Single-Quantum Coherence
Ic_{50}	Half maximal inhibitory concentration
IP3	Inositol tri-phosphate
IPTG	Isopropyl-beta-thio galactopyranoside
IQ	Calmodulin binding domain motive
Kan	Kanamycin
K_v	Voltage-dependent potassium channel
LB	Luria Bertani Broth
LQTS	Long QT-syndrome
MBP	Maltose-Binding Protein
mVenus	monomeric circularly permutated venus
min	Minutes
mRNA	Messenger RNA
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
O/N	over night
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PDB	Protein Data Banck
PIC	Protease Inhibitor Cocktail
PIP_2	Phosphatidylinositol 4,5-bisphosphate
PMSF	phenylmethanesulfonylfluoride
Q2ABc	K_v7.2_ABc
RCI	Random Coil Index
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
RPM	Revolutions per minute
RT	Room temperature

FIGURE LIST

Figure 1.1. Resting membrane potential.

Figure 1.2. Ideal representation of an action potential with its different phases.

Figure 1.3. Phylogeny tree of VGPCs

Figure 1.4. General architecture of a voltage-gated ion channel.
Figure 1.5. Maximum likehood tree of human K_v channels.
Figure 1.6. Structure of KcsA.
Figure 1.7. Structure of K_v1.2.
Figure 1.8. Schematic representation of the intracellular domains of K_v channels.
Figure 1.9. Neuronal response to a stimulus in absence and in presence of muscarine.
Figure 1.10. Scheme of the C-terminal of K_v7 channels.
Figure 1.11. Putative PIP_2 interaction sites on K_v7 channels.
Figure 1.12. M-current inhibition in sympathetic neurons.
Figure 1.13. Localization of BFNC-linked mutations.
Figure 1.14. Calcium homeostasis in cells.
Figure 1.15. Net of proteins regulated by CaM.
Figure 1.16. CaM structure.
Figure 1.17. EF-hands structure.
Figure 1.18. Interaction preference of each CaM residue with amino acids from target peptides.
Figure 1.19. CaM's averaged contact surface with targets.
Figure 1.20. CaM complexing with its target.
Figure 1.21. CaM complexing with its target.
Figure 1.22. CaM sequence and disease-associated mutations.
Figure 2.1. Workflow proposed to improve recombinant protein's solubility.
Figure 2.2. Secondary structure prediction in silico.
Figure 2.3. Intrinsically Disordered Regions prediction in silico.
Figure 2.4. Over-expression and purification of His-Q2ABc protein (28 kDa).
Figure 2.5. Over-expression and purification of GST-Q2ABc protein (28 kDa).
Figure 2.6. Graphical representation of the impact of optimized and not-optimized genes over humanized and not-humanized strains.
Figure 2.7. Q2ABc protein solubility using different fusion proteins.
Figure 2.8. Protein purification and fluorometric assay.
Figure 2.9. Example of CaM-fused proteins.
Figure 2.10. Thermostability assay of CaM-fused proteins.
Figure 2.11. His-tag location effect in protein solubility.
Figure 2.12. Screening of protein solubility introducing point mutations.
Figure 2.13. Solubility screening of Kv7 channels orthologous fused to CaM.
Figure 2.14 K_v7.2 and K_v7.3 C-terminus solubility using different fusion proteins.
Figure 3.1. Schematic view of different structure determination techniques.
Figure 3.2. Spin precession and energetic diagram.
Figure 3.3. Magnetic moment perturbation and relaxation.

Figure 3.4. ^{15}N-HSQC spectra of COX17 protein.
Figure 3.5. Magnetization transfers in three dimensional spectrum.
Figure 3.6. Average values for each residue.
Figure 3.7. RDCs graphical representation.
Figure 3.8. De novo stimation of the axial component and the rhombicity value of the alignment tensor.
Figure 3.9. Overview of macromolecular structure determination by NMR.
Figure 3.10. PONDR VL-XT disorder prediction and designed amino acid sequences trying to minimize these regions.
Figure 3.11. Combinations of tested deletions.
Figure 3.12. Thermostability assay for different deletions.
Figure 3.13. Δ6L deletion give rise to functional K_v7.2 channels.
Figure 3.14. Buffer screening for K_v7.2 C-terminus with Δ6L fused to Venus and co-expressed with CaM.
Figure 3.15. Alternative constructs' concentrations.
Figure 3.16. Protein solubility percentage of selected constructs.
Figure 3.17. Protein complex ^{15}N-HSQC.
Figure 3.18. HSQC overlay of independently labelled proteins forming CaM/K_v7.2_AB.
Figure 3.19. Random Coil Index (RCI) and Root Mean Square Deviation (RMSD) values for GB116 and CaM.
Figure 3.20. Secondary structure of GB116 and CaM.
Figure 3.21. Phi (Φ) and Psi (Ψ) torsion angles restrains for GB116 and CaM.
Figure 3.22. Relative orientation of protein elements forming the complex.
Figure 3.23. K_v7.2_AB contact map.
Figure 3.24. K_v7.2_AB structural model.
Figure 3.25. Ramachandran plot of best model for helices AB.
Figure 3.26. CaM contact map.
Figure 3.27. CaM structural model.
Figure 3.28. Rammachandran plot of best model for CaM.
Figure 3.29. Structure of CaM/K_v7.2_AB complex.
Figure 3.30. Overlay of CaM/K_v7.2_AB complex and CaM/K_v7.1_AB complex.
Figure 3.31. Diagram of inter-helical interaction in the anti-parallel coiled coil helices.
Figure 4.1. Free-energy landscape of proteins.
Figure 4.2. Protein dynamics over different time ranges.
Figure 4.3. NMR detection of protein interactions.
Figure 4.4. Q2ABc binding weakens the Ca^{2+}-CaM interaction.

Figure 4.5. Q2ABc binds preferentially to the N-lobe in the absence and to the C-lobe in the presence of Ca^{2+}.
Figure 4.6. CaM's lobes preference for each helix (A and B).
Figure 4.7. Similarities between calmodulin lobes complexed with $K_v7.2$ and database forming 34 different calmodulins.
Figure 4.8. H-N chemical shift comparison of #7 and #16 entries towards reference.
Figure 4.9. Size-Exclusion Chromatography.
Figure 4.10. ^{15}N-HSQC overlay of three complex states.
Figure 4.11. CSPs induced by calcium in CaM.
Figure 4.12. CSPs induced by calcium in $K_v7.2$ helices AB region.
Figure 4.13. Structure characterized CaM/$K_v7.2$_AB complex with corresponding Ca^{2+} ions.
Figure 4.14. Interaction model of CaM binding to $K_v7.2$ helices A and B.

TABLE LIST

Table 1.1. Potassium channels and human diseases.
Table 1.2. Sequence alignment of K_v7 channels' helix A and helix B.
Table 1.3. Examples of proteins regulated by CaM.
Table 1.4. Overall motif for CaM recognition.
Table 2.1. Table of most used *E. coli* strands.
Table 2.2. Antibiotics.
Table 2.3. Replicon types and combination.
Table 2.4. Table of most used promoters and their characteristics.
Table 2.5. Table of most used tags and their characteristics.
Table 2.6. Bulk protein expression comparative to BL21(DE3) strain.
Table 3.1. Buffer screening preparation table.
Table 4.1. Summary of binding parameters for $K_v7.2$AB.
Table 4.1. Summary of binding parameters.
Table 4.2. Summary of binding parameters for each CaM lobe.