Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Chimiche

CORSO DI DOTTORATO DI RICERCA IN: Scienze Molecolari
CURRICOLO: Scienze Chimiche
CICLO: XXX

Structural analysis of molecular recognition and ligand association processes

Coordinatore: Ch.mo Prof. Leonard Prins
Supervisore: Ch.mo Prof. Roberto Battistutta

Dottoranda : Elisa Costanzi

Abstract

Molecular recognition is a fundamental step in essentially any biochemical process. Detailed structural knowledge is crucial to have a better understanding of the processes in which the two interacting molecular partners are involved and can be exploited in several applied fields such as supramolecular design of new molecular assemblies, rational drug design, and enzyme engineering. In the context of molecular recognition, I have investigated (mainly by single crystal x-ray crystallography) some relevant protein-protein and protein-ligand systems in order to gain detailed structural insights on the interactions involved, at atomic level.

First, the STAS domain of prestin, an anion-dependent motor protein, and its interaction with monovalent anions and with calmodulin.

Second, the interaction between protein kinases (CDK2 and CK2) and BCL-XL, and inhibitors, for the rational design of specific drugs targeting these proteins involved in different types of cancer.

Riassunto

Il riconoscimento molecolare è uno step fondamentale nei processi biochimici. Una conoscenza strutturale dettagliata è cruciale per capire meglio i processi in cui sono coinvolti due partner molecolari interagenti e può essere sfruttata in vari campi come il disegno di nuovi assemblamenti sopramolecolari, il disegno razionale di farmaci e l'ingegnerizzazione di enzimi. In questo contesto, ho investigato (prevalentemente tramite cristallografia a raggi x su cristallo singolo) alcuni sistemi proteina-proteina e proteina-ligando per ottenere dettagli strutturali delle interazioni coinvolte, a livello atomico.

In primis, il dominio STAS di prestina, una proteina motrice anioni-dipendente, e la sua interazione con anioni monovalenti e con calmodulina.

Poi, l'interazione tra protein chinasi (CDK2 e CK2) e BCL-XL, e inibitori, per il disegno razionale di farmaci specifici nel colpire queste proteine coinvolte in vari tipi di cancro.

Preface

Molecular recognition is a fundamental step in essentially any biochemical process. Enzyme catalysis, cellular signalling, protein-protein association, protein-ligand interactions, to name only a few, involve the recognition between two or more (macro)molecules. Specific non-covalent chemical interactions are at the basis of this recognition, namely hydrogen bonds, van der Waals forces, metal coordination, polar and electrostatic effects, π-π interactions, hydrophobic interactions, halogen bonds etc. For the study of these processes, different techniques complement each other, accessing different types of information. Macromolecular x-ray crystallography is the most popular method to obtain high-resolution structural data on protein–protein and protein-ligand complexes, revealing the atomic details of the interaction that is information on the enthalpic contributions. Other techniques, such as SAXS (Small Angle X-ray Scattering), biocalorimetry and NMR can yield valuable information. Computational methods are also useful, particularly to analyse entropic contributions. This detailed structural knowledge is crucial to have a better understanding of the processes in which the two interacting molecular partners are involved and can be exploited in several applied fields such as supramolecular design of new molecular assemblies, rational drug design, and enzyme engineering. In the context of molecular recognition, I have investigated some relevant protein-protein and protein-ligand systems in order to gain detailed structural insights on the interactions involved, at atomic level. In this context my Ph.D. project was devided in two main topics, described in part 1 and part 2, and two side projects, described in part 3.

PART 1

The first main project of my Ph.D. is described in chapter 1, and it regards the STAS domain of prestin. Prestin, in mammals, is an ATP-independent molecular motor. The crystal structure of the STAS domain of rat prestin is available and it is known that intracellular anions are crucial for the functionality of prestin. I therefore studied by x-ray crystallography, their interaction with the intracellular STAS domain, identifying the nature of the most important residues and functional groups. It is also known that the STAS domain interacts with calmodulin (CaM). I started to investigate the STAS/CaM complex by SEC (Size Esclusion Chromatography), DLS, SAXS and SPR.

PART 2

Another important protein-ligand interaction that I studied, described in chapter 2, is that between enzymes and inhibitors, in particular involving some oncogenic protein kinases such as CDK2 and CK2. The knowledge of the interactions between inhibitors and deregulated enzymes is of fundamental importance for the rational design of specific drugs targeting these proteins involved in different types of cancer.

PART 3

I also had two side projects, described in chapter 3, the first one regarding the the anti-apoptotic protein Bcl-xL with the aim of developing inhibitory compounds with anti-cancer potential. The second one regarded the calcium-binding protein calsequestrin (CASQ).

PART ONE – the STAS domain of PRESTIN

Introduction

The Sulphate Permease (SulP) / Solute Linked Carrier 26 (SLC26) family

The Sulfate Permease (SulP) is a large and ubiquitous superfamily of transporters with over 200 sequenced members derived from archaea, bacteria, fungi, plants and animals. Also known as the Solute Carrier 26 (SLC26) family of anion transporters, in mammals the SulP family is composed of versatile anion exchangers, with important roles in normal physiology and human pathophysiology.

Since structural and functional information on proteins can be correlated if the proteins share a common ancestor, it is important to identify and classify protein families. The Gene Nomenclature Committee (HGNC) of the Human Genome Organization (HUGO, Povey et al, 2001) classifies human solute carriers into 47 families (Hediger et al, 2004). According to this classification, members of a family share a similar substrate and at least 20–25% sequence identity to at least one other member of the family. Sequences are however less conserved then three-dimensional structures (Schlessinger et al., 2010) therefore correlation between proteins can be shown more clearly when the crystallographic structures are available. The knowledge of the atomic structure of a protein can shed light on its mechanism of action and its function. The first crystallographic structure of a full-length transporter belonging to the SLC26 family, a fumarate transporter from *Deinococcus geothermalis*, was recently determined (Geertsma et al., 2015).

The SLC26 anion exchangers are versatile anion exchangers, with intriguing roles in normal physiology and human pathophysiology. Among the physiological processes in which the SLC26 exchangers play critical roles there are: Outer Hair Cells (OHCs) electromotility, skeletal development, synthesis of thyroid hormone, trans-epithelial Na^+-Cl^- transport, bicarbonate excretion by the distal nephron, and bicarbonate secretion by the exocrine pancreas (Mount and Romero, 2004). Members of the SLC26A family function as anion exchangers or channels in the luminal or apical membranes of epithelial tissue and are primarily involved in transport of a wide variety of monovalent and divalent anions. Each member has different anion specificity and distinctive tissue distribution; some being expressed in most organs and others with more restricted tissue expression patterns. To date, eleven human SLC26 genes have been identified, ten of which were shown to encode proteins that transport one or more substrates, including sulphate, chloride, bicarbonate, iodide, oxalate, formate, hydroxyl, mannose and fructose (Mount and Romero, 2004). SLC26A5 (prestin) was shown to act as the motor protein of cochlear OHCs (Zheng et *al.*, 2000). The SLC26 family thus exhibits an amazing variety of functions, yet the molecular basis of this diversity is poorly understood.

The SLC26 transporters are large proteins comprised of 700–1000 amino acids, and the individual family members have 21–43% amino acid identity. The evolutionary tree in figure 1-1 shows the relatedness of the family members within and among species (Dorwart et al, 2008).

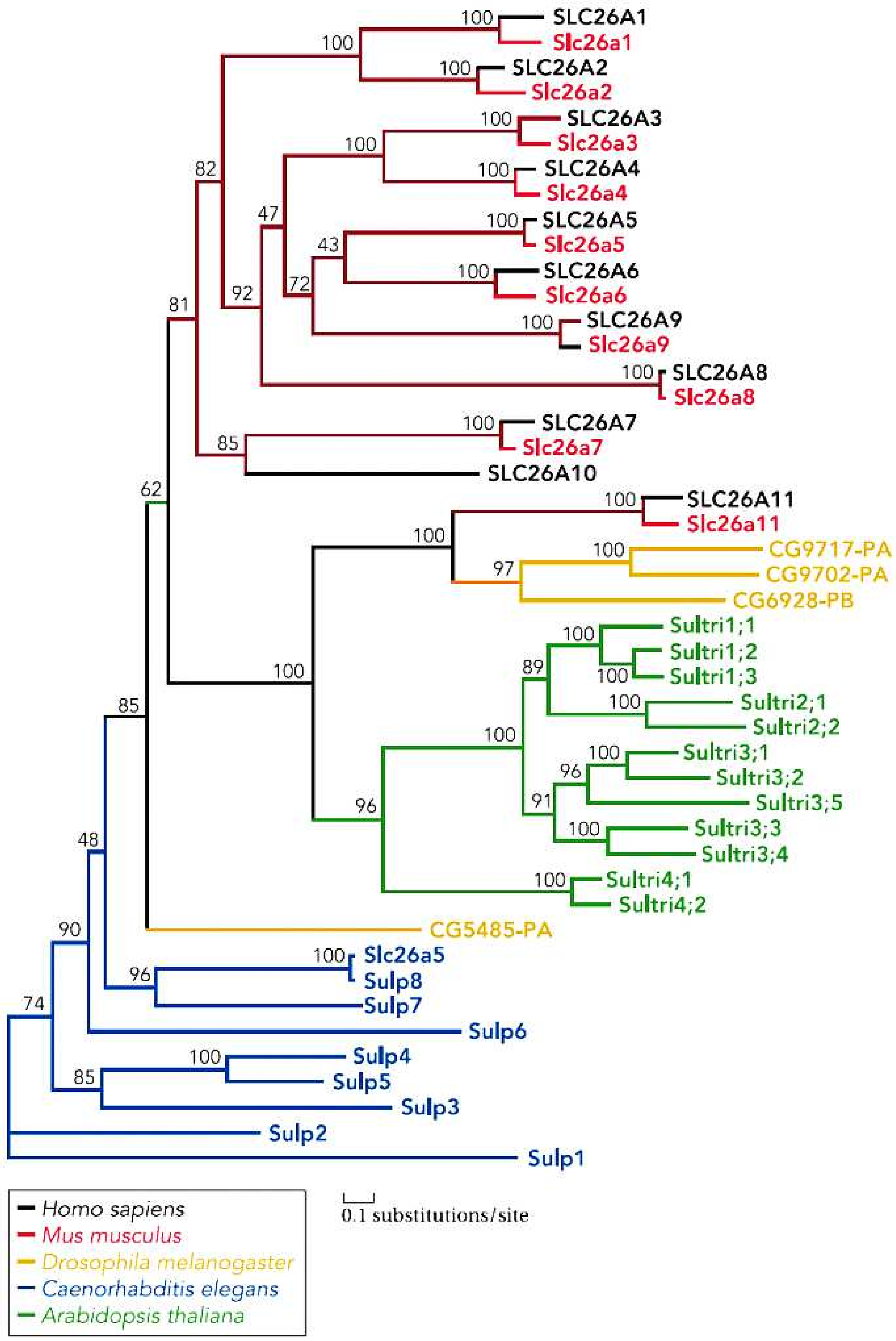


Figure 1- 1 Evolutionary tree generated using SLC26 sequences from H. sapiens and M. musculus, as well as homologous sulfate transporters from A. thaliana, C. elegans (Sultr and Sulp, respectively), and D. melanogaster The HUGO nomenclature is used here; "SLC26A-" denotes a human gene/protein, "Slc26a-" denotes a rodent ortholog.

The recently solved structure of the fumarate transporter from the bacterium *Deinococcus geothermalis* confirmed previous topological analysis of the TM domain of other members of the family, the cyanobacterial bicarbonate transporter BicA and SLC26A5 rat prestin (figure 1-2). The TM domain is composed of 14 transmembrane segments organized in a 7+7 inverted repeat fold. It is probable that all members of the family adopt such a structural organization.

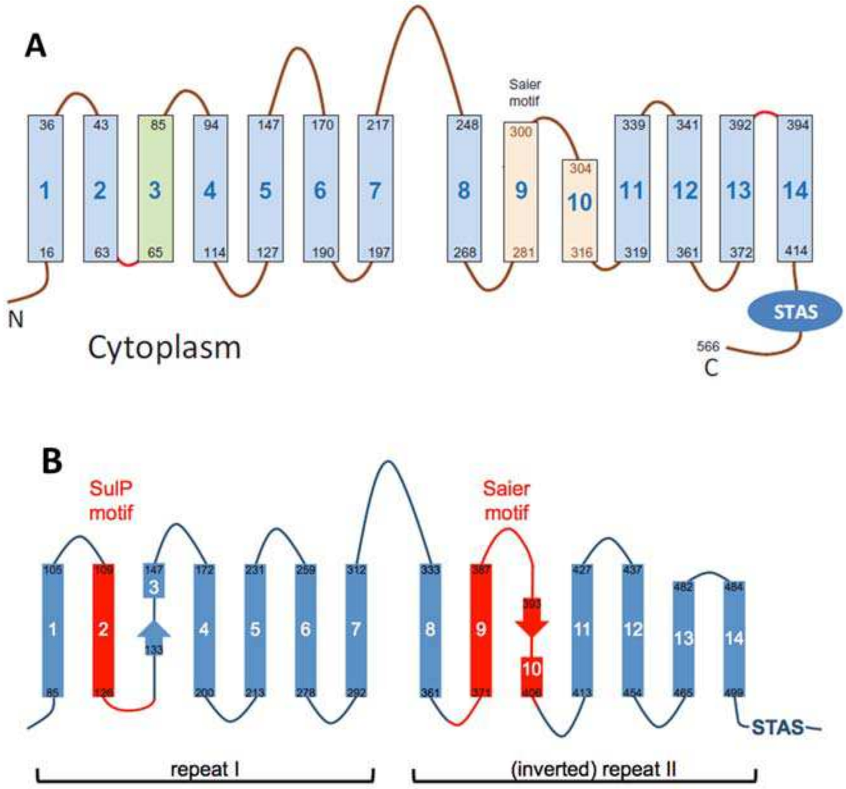


Figure 1- 2 Topology of the TMD of two SulP/SLC26 members, the cyanobacterial bicarbonate transporter BicA (panel A, Price and Howitt, 2014) and SLC26A5 rat prestin (panel B, Gorbunov et al., 2014).

Both N- and C-terminal domains are intracellular. In the TM domain there are regions of homology between SLC26 members. The first one is the 22 amino-acid "sulfate transport" consensus signature (Prosite, PS01130) Then a second cluster of invariant residues is a conserved segment defined by Saier et al. (1999), located at the C-terminal end of the hydrophobic core of the proteins. Another region of homology is the C-terminal cytoplasmic domain of all ten SLC26 proteins, the sulfate transporter and anti- sigma (STAS) domain so called due to the homology with bacterial anti-sigma factor antagonists. There is not a clear understanding of the physiological and/or mechanistic roles of the STAS domain in the SLC26 exchangers; however, the existence of disease-associated mutations in this domain suggests that it has potential importance (Aravind, 2000).

SLC26 and transport

The mode and the stoichiometry of the transport, the identity of the ion binding sites and conduction pathways, and the ionic specificity/selectivity of the transport are crucial information to understand the function of a transporter. Dorwart et al. divided the SLC26 transporters into three main groups: *group 1* members are SLC26A1 and SLC26A2 which are selective sulfate transporters; *group 2* includes SLC26A3, SLC26A4, and SLC26A6 that are coupled $Cl^-/HCO3^-$ exchangers; SLC26A7 and SLC26A9 belong to *group 3* and function as ion channel. The transport modes of SLC26A8 and SLC26A11 are unknown. SLC26A5 is an electrogenic anion exchanger in non-mammals and an incomplete transporter with motor function and piezoelectric properties in mammals.

SLC26 and pathologies

Mutations in the SLC26 family genes can cause several human pathologies (Dawson and Markovich, 2005). Among them SLC26A2 mutations are involved in chondrodysplasias, a disease that cause skeletal defects and SLC26A3 mutations are linked to congenital chloride-losing diarrhea. SLC26A4 is involved in Pendred syndrome, which is the most common form of syndromic deafness while SLC26A5 mutations were identified in individuals with non-syndromic deafness. (Liu et *al.*, 2003; Toth et *al.*, 2007).

Some of the key features of the members of the SLC26 family are summarized in table 1.1.

Gene	Protein name	Transport	Reported substrate	Tissue distribution	Disease association
SLC26A1	Sat-1	selective sulfate transporter	SO_4^{2-}, oxalate	Liver, kidney	
SLC26A2	DTDST	selective sulfate transporter	SO_4^{2-}, Cl-	Widespread	Chondrodysplasias
SLC26A3	DRA, CLD	coupled Cl^-/ $HCO3^-$ exchanger	SO_4^{2-}, Cl-, HCO_3-, OH-, oxalate, formate	Intestine, sweat gland, pancreas, prostate	Congenital chloride diarrhea
SLC26A4	Pendrin	coupled Cl^-/ $HCO3^-$ exchanger	Cl-, HCO_3-, I-, formate	Inner ear, kidney, thyroid	Pendred syndrome, deafness (DFNB4)
SLC26A5	Prestin	electrogenic anion exchanger *in non-mammals* incomplete transporter with motor function and piezoelectric properties *in mammals*	Halides, HCO_3-, Cl- oxalate, formate, SO_4^{2-}	Inner ear	Deafness?
SLC26A6	CFEX, PAT-1	coupled Cl^-/ $HCO3^-$ exchanger	SO_4^{2-}, Cl-, HCO_3-, OH-, oxalate, formate	Widespread	
SLC26A7	None	ion channel	SO_4^{2-}, Cl-, oxalate	Kidney	
SLC26A8	Tat1	unknown	SO_4^{2-}, Cl-, oxalate	Sperm, brain	
SLC26A9	None	ion channel	SO_4^{2-}, Cl-, oxalate	Lung	
SLC26A10	None	/	/	Brain	
SLC26A11	None	unknown	SO_4^{2-}	Widespread	

Table 1- 1 Main features of the SLC26 transporters

Outer Hair Cells (OHCs) and electromotility

In the choclea sound vibrations are converted in pressure waves in the fluids of the inner ear. The transduction of acoustic signals involves the cochlear hair cells that are non-neuronal epithelial cells. These cells are organized with the ones sensitive to high-pitched sounds at the basal end and those sensitive to low pitches at the apical end, and therefore in a tonotopic fashion in the basilar membrane (Géléoc and Holt, 2003).

In the choclea, there is a cellular matrix, with two types of cells, called the organ of Corti and composed by a single row of Inner Hair Cells (IHCs) and three-four rows of Outer Hair Cells (OHCs) (figure 1-3).

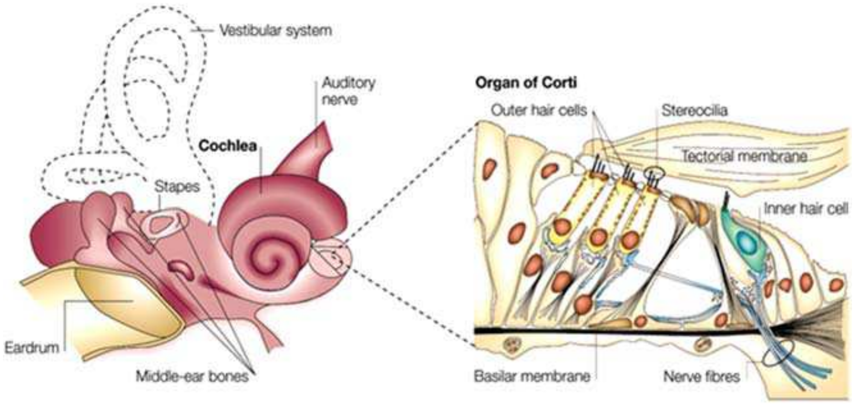


Figure 1- 3 A cross section of the cochlea illustrating the organ of Corti, the sensory epithelium of the inner ear (Dallos and Fakler, 2002).

The Inner Hair Cells (IHCs) transduce and transmit auditory information to the brain. The Outer Hair Cells (OHCs) provide local mechanical amplification of the auditory stimuli sensed by the IHCs (Dallos, 1992). The OHCs have a distinctive hair (stereocilia) bundle, which is their mechanosensory input organelle. The ciliary bundle is deflected when mechanically stimulated by incoming sound waves, triggering the opening and closing of mechanosensitive ion channels in the membrane.

The resulting changes in membrane potential are then translated in the OHCs into macroscopic changes (up to 5%) in the length of their bodies. A depolarization of the membrane is converted in a contraction of the cell, whereas hyperpolarization results in cell elongation (figure 1-4).

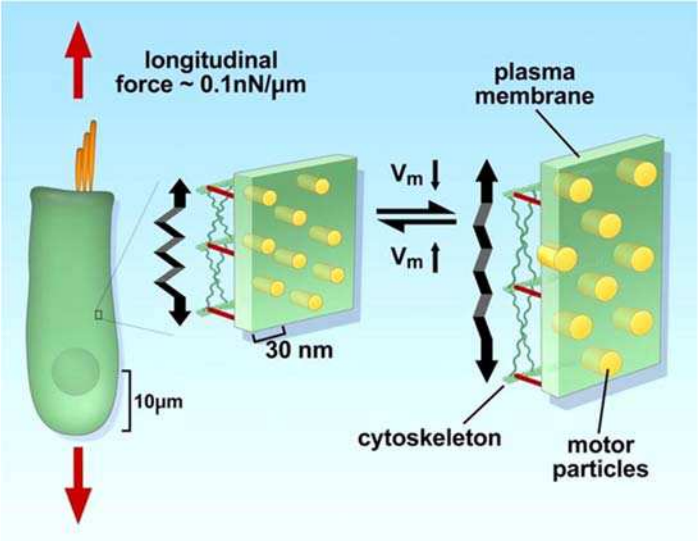


Figure 1- 4 Potential variations generate a longitudinal force in the OHCs that causes a variation of the length of the cells.

This phenomenon takes the name of "electromotility" and occurs at acoustic frequencies. The mechanical energy that is required for amplifying the sound-induced vibrations in the cochlea, is therefore generated. This whole mechanical amplification mechanism is known as the "cochlear amplifier" (Ashmore, 1987) and is fundamental for the high sensitivity and frequency selectivity of hearing in mammals.

The OHCs electromotility is ATP-independent and occurs at microsecond rates working in a cycle-by-cycle mode, up to a frequency of at least 70 kiloHertz. It is also known that Ca^{2+} ions are not required but internal Ca^{2+} levels can modulate the phenomenon.

Electromotility is tightly coupled to transmembrane voltage and this is reflected by the phenomena of gating charge movements, involving a charge dislocation across the cell membrane, similar to those observed for voltage-gated ion channels.

Cell membranes have inherent or passive electrical capacitance, which is linear and depends only on the surface area of the membrane and its dielectric properties. However, any voltage-dependent gating charge might be regarded as a change in capacitance, since capacitance is defined as the derivative of charge with respect to voltage. This nonlinear capacitance is added to the inherent linear capacitance of the cell. The charge transfer (gating current) gives rise to a voltage-dependent non-linear component of membrane capacitance, which shows a bell-shaped dependence on membrane potential (figure 1-5).

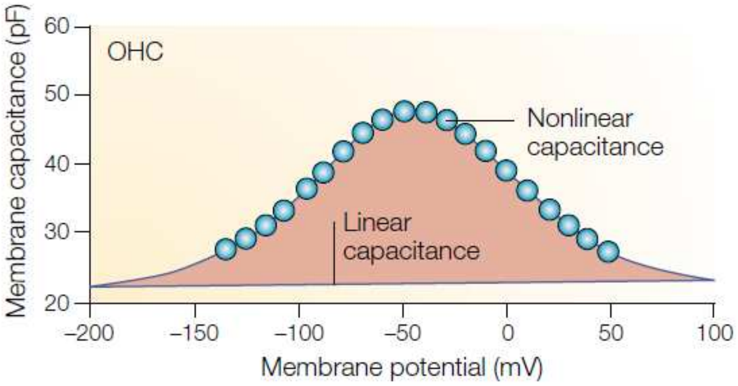


Figure 1- 5 The shape of the NLC curve reflects the probability of movable charges being translocated between two positions at opposite sides of the membrane. Therefore, capacitance peaks at the voltage that is most effective in producing a motile response.

Since the Non Linear Capacitance (NLC) is easier to measure than motility, it is widely used as a signature of electromotility (Santos-Sacchi, 1991).

All these findings led to the hypothesis of the existence of an integral membrane protein acting as a molecular motor protein in the OHCs electromotility. This molecular motor protein was then identified in SLC26A5, which was called "Prestin" from the musical notation *presto*, meaning fast (Zheng, *et al.* 2000).

Prestin (SLC26A5)

The fifth member (A5) of the Solute Linked Carrier 26 (SLC26) family of anion exchangers is Prestin. It is highly and almost exclusively expressed in the Outer Hair Cells (OHCs) of the organ of Corti in the inner ear of mammals. Although the basic function of SLC26A members is to transport anions (Mount and Romero, 2004), this is not prestin principal role in mammals. Unlike the other members of the SLC26 family, mammalian prestin undergoes voltage-dependent conformational changes and is considered the key player in the OHC somatic electromotility (Zheng et *al.*, 2000). In non-mammals, prestin acts as an electrogenic divalent/chloride anion exchanger (Schaechinger and Oliver, 2007).

The discovery of prestin

A subtractive cloning strategy to amplify transcripts expressed in OHCs but not expressed in the non-motile IHCs was used by Dallos and colleagues to identify motor protein candidates. A cDNA that is specifically expressed in OHCs was isolated and named prestin (Zheng et *al.*, 2000).

When heterologously expressed in mammalian cell lines, prestin showed all the hallmarks of the OHC motor protein. First, it endowed transfected cells with NLC, which had similar characteristics to that observed in OHCs. Second, prestin expressing cells were electromotile with motility magnitudes approaching 0.2 µm (Zheng et al., 2000). Third, actual force measurements, carried out with an atomic force microscope, showed that prestin generates significant mechanical force, which is independent of frequency up to at least 20 kHz (Ludwig et *al.*, 2001).

Mammalian prestins and electromotility

Human prestin is an 81 kDa transmembrane glycoprotein of 744 residues (Zheng et *al.*, 2000). It shares the overall structure and specific protein domains of the SLC26 family members, such as a highly conserved central core of hydrophobic amino acids, a short N-terminus and a long C-terminus. The mostly hydrophilic amino- and carboxy-termini, which flank the hydrophobic core, are located in the cytoplasm, as it was shown by epitope tagging and immunocytochemistry (Ludwig et *al.*, 2001; Zheng et *al.*, 2001).

Prestin is a highly conserved protein with 92.7% of amino acids being identical among four different mammalian species: human, mouse, rat and gerbil. Such a high degree of conservation is not common among other SLC26A members. Significant changes in prestin primary sequence occurred after the split between mammalian and avian lines, suggesting that prestin evolved in order to fit special mammalian needs (Dallos et *al.*, 2006).

To exert its function, prestin should comprise at least two essential functional domains: the voltage sensor that detects changes in the transmembrane potential of the cell, and the actuator that undergoes a conformational change and thereby facilitates cell contraction or elongation in response to depolarization and hyperpolarization, respectively. The nature of the voltage sensor was determined by Oliver and colleagues (Oliver et al., 2001). To identify the sensor, all the charged residues that are not conserved between prestin and SLC26A6, the SLC protein with the closest homology to prestin, were neutralized either alone or in combination and the resulting mutated proteins were probed for NLC. As NLC was not abolished in any of these cases, the concept that charged particles that are extrinsic to the protein act as the voltage sensor was tested. Cations nor external anions had any effect on NLC. However, NLC and electromotility were reversibly eliminated by removing Cl^- ions from the cytoplasm of cells containing either wild-type prestin or any of the prestin mutants tested. Detailed analysis showed that the half-activating Cl^- concentration was 6 mM, matching the normal intracellular amount of this anion. Other monovalent anions were also found to be effective in promoting NLC, in the order $I^- \approx Br^- > NO_3^- > Cl^- > HCO_3^- > F^-$. Carboxylic acids could also substitute for Cl^-, although their effectiveness is inversely proportional to their chain length. The voltage required for moving butyrate across the plasma membrane is therefore more than twice that necessary for translocating the halides, mentioned above, or formate. Moreover, divalent anions (for example, SO_4^{2-}) were ineffective in functionally replacing their monovalent counterparts. On the basis of these results, it was concluded that prestin uses an extrinsic voltage sensor, monovalent anions, available in the cytoplasm. After binding to a site with millimolar affinity, these anions are translocated across the membrane in response to changes in the transmembrane voltage. They move towards the extra-cellular surface following hyperpolarization, and towards the cytoplasmic side in response to depolarization (figure 1-6).

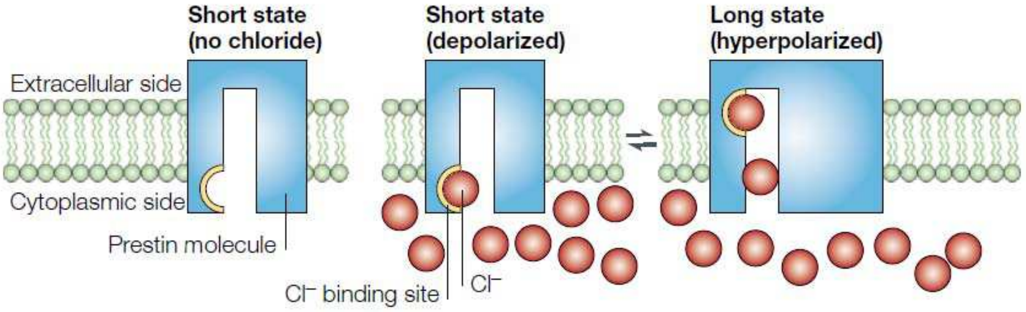


Figure 1- 6 Model of the control of prestin by internal Cl– ions. Left: In the absence of internal Cl–, the molecule is in its 'short' state. Middle: The cell membrane is depolarized. Cl– is bound to the molecule but remains at the cytoplasmic face of the membrane. Right: The cell membrane is hyperpolarized. Bound Cl– is translocated across the molecule towards the external face. When bound Cl– is in the outside position, the molecule assumes its 'long' conformation. Conversely, when the bound Cl– is displaced back towards the cytoplasm (middle), the molecule becomes 'short'. At the normal outer hair cell resting potential, the molecule is more likely to be in its long state. Consequently, the cell contraction caused by depolarization is greater than the cell elongation caused by hyperpolarization (Dallos and Fakler, 2002).

As a consequence, this translocation triggers conformational changes in the protein that ultimately alter its surface area in the plane of the plasma membrane. The area decreases when the anion is near the cytoplasmic face of the membrane ('short state'), and increases when the ion has crossed the membrane to the outer surface ('long state'). Prestin therefore acts as an incomplete transporter swinging anions across the membrane, but not allowing these anions to dissociate and escape to the extracellular space. When monovalent anions are not present in the cytoplasm, all prestin molecules are in their ‘short’ state, as the OHC is maximally contracted.

Oliver *et al.* also showed that salicylate, the active component of aspirin, acts as a competitive antagonist at the anion-binding site of prestin. The anionic salicylate is able to induce NLC (that is, it is translocated by prestin in a very similar manner to the translocation of large carboxylic acids). The amplitude of this NLC, however, is more than an order of magnitude lower than that found for the small Cl^- ion. Nevertheless, the binding affinity of salicylate is around 300-fold higher than that found for Cl^-. This result provides a possible explanation for the significantly reduced OHC electromotility that probably underlies the hearing loss induced by large doses of aspirin (Shehata et al., 1991).

Non-mammalian prestins

Non-mammalians prestins do not appear to be motor proteins as their mammalian orthologs. In Oliver’s lab prestin orthologs from chicken and zebrafish, non-mammalian vertebrates that presumably lack electromotility in their auditory systems, were cloned. Using patch-clamp experiments, they showed that these prestin orthologs generate robust transport currents in the presence of the divalent anions sulfate or oxalate. Transport is blocked by salicylate, an inhibitor of electromotility generated by mammalian prestin. The dependence of transport equilibrium potentials on sulfate and chloride concentration gradients shows that the prestin orthologs are electrogenic antiporters, exchanging sulfate or oxalate for chloride in a strictly coupled manner with a 1:1 stoichiometry (Schaechinger and Oliver, 2007).

Mammalian and non-mammalian prestins comparison and model of their mechanisms of action

Several aspects of the function of mammalian prestins show striking similarities when compared with SO_4^{2-}/Cl^- antiport by chickenPrestin/zebrafishPrestin. First, voltage sensitivity and the resultant conformational rearrangements of mammalian prestin depend on the presence of millimolar Cl^-. Binding of the monovalent anion could be to the same binding sites involved in Cl^- transport by nonmammalian prestin. Second, salicylate block of mammalian prestin quantitatively equals the inhibition of nonmammalian transport. Third, zPres generates voltage-dependent charge movements resembling the gating charge associated with electromotility. Finally, mammalian and chicken isoforms share a substantial degree of sequence conservation. These similarities strongly suggest that the unique function of mammalian prestin evolved as a modification of an anion exchange mechanism. Oliver and collegues proposed that electromotility might arise from an incomplete anion transport cycle (figure 1-7).

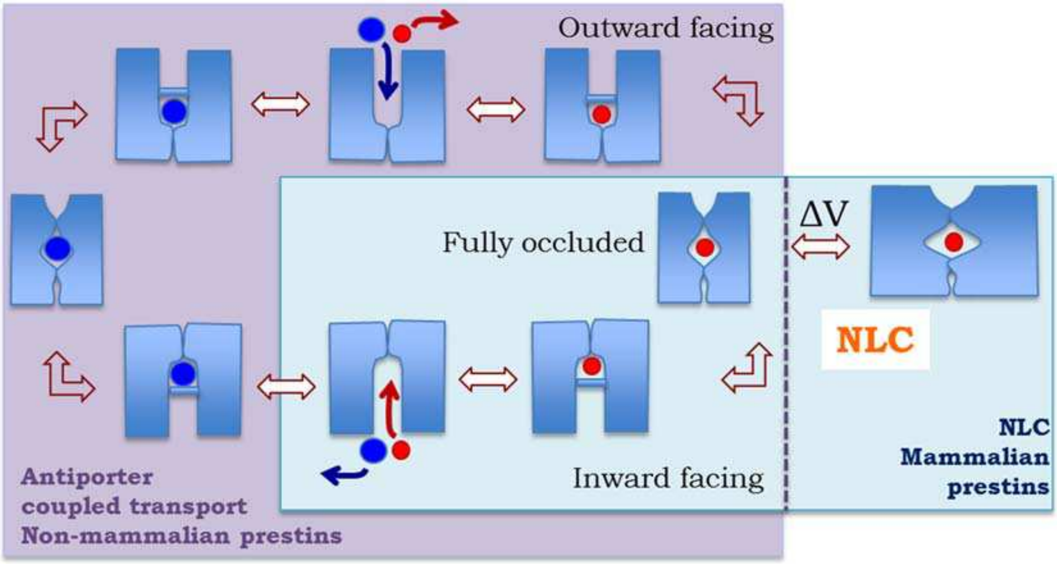


Figure 1- 7 Alternating-access model for prestin function. In purple, the antiporter mechanism for non-mammalian orthologs, in light blu, the stastes accessible to mammalian orthologs.

In non-mammals, stoichiometric antiport is realized by transitions between the two major conformations with the substrate binding site facing either the cytosol or the extracellular space; transitions between these two states only occur upon substrate binding. The distribution between the various states depends on transport rates and substrate concentrations. In mammals, the anion can reach the TM domain only from the cystoplasm explaining the necessity of intracellular anions for the generation of NLC. The model also explains the suppressive effect of intracellular oxalate on NLC generated by rPres: because NLC is completely independent of extracellular Cl^- (Oliver et al, 2001), dissociation from and binding of Cl^- to the outward-facing conformation must be impossible or greatly disfavoured. This is also consistent with the lack of transport activity in rPres (Schaechinger and Oliver, 2007).

Structural studies

The study of the structural properties of prestin is of paramount importance for its central role in the hearing process of mammals and hence to counteract hearing impairments caused by prestin mutations and malfunctioning. Despite a great deal of interest in the structural properties of prestin, at present the experimental 3D structure of the full-length protein is still lacking. This is due to the notorious difficulties in producing and structurally characterize membrane proteins, in particular those with large conformational variations typical for instance of transporter proteins, that cycle though inward-open and outward open conformational states. However, structural information is available for the individual, isolated, TM domain and STAS domain of mammalian ratPres.

The STAS domain

The STAS domain is the C-terminal cytoplasmic domains of SulP and SLC26 anion transport proteins, playing a fundamental role in the function of SulP/SLC26 transporters. Deletion mutants demonstrated that the domain is necessary for correct membrane localization of prestin and for the generation of Non Linear Capacitance (Zheng et al., 2005).

The intracellular STAS domain derives its name from a remote but significant sequence similarity with bacterial spoIIAA (ASA, Anti-Sigma factor Antagonist) proteins (STAS = Sulfate Transporters and Anti-sigma factor Antagonist; Aravind et al., 2000). These anti-anti-σ proteins indirectly stimulate bacterial RNA polymerase by inactivating inhibitory anti-σ kinases, liberating σ factors to direct specific transcription of target genes or operons (Sharma et al., 2011).

Regarding STAS domains, structural data are available both for bacterial domains (Sharma et al., 2011; Babu et al., 2010) and for the *R. norvegicus* domain of prestin (Pasqualetto et al., 2010). The crystal structure of the STAS domain of rat prestin was determined at 1.57 Å resolution (figure 1-8) and shows the presence of 6 β-strands and 5 α-helixes.

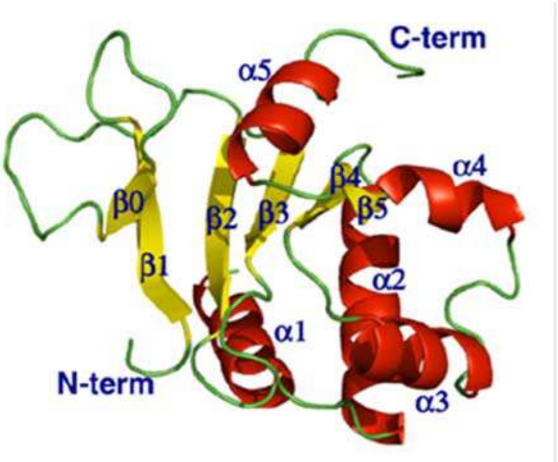


Figure 1- 8 RatSTAS domain of prestin ([505-563]GlySer[637-718]) with its secondary structure elements highlighted

This structure revealed that the prestin STAS significantly deviates from those of the related bacterial ASA proteins, redefining the boundaries of the domain, particularly at the N-terminal end. It was found that the N-terminal region from residue 505 to residue 525 (rat prestin numbering), previously considered merely a generic linker region between the last transmembrane region and the STAS domain, is indeed fully part of the STAS domain from a structural point of view. This implies that, unexpectedly, the STAS domain lies just beneath the plasma membrane, most probably being able to interact with the lipid bilayer and/or with portions of the transmembrane domain of the protein in a functionally relevant manner. The similarity with the ASA proteins in the secondary structure and topology of the remaining part of the structure justifies the notion of STAS domain, as a domain in common to bacterial anti-sigma factor antagonists and SulP anion transporters. The structure-function analysis of the rat prestin STAS domain indicates that this model can be a general template for most SLC26 and SulP anion transporters. The STAS domain can be implicated in molecular interactions involving different types of possible partners: the lipid bilayer; the transmembrane domain of the same protein; other intracellular portions of the transporter for the correct assembly of the oligomeric state of prestin; small molecules that can regulate protein function); other proteins.

Very recently it was found by our group that ratSTAS has a pocket that can directly bind anions (figure 1-9).

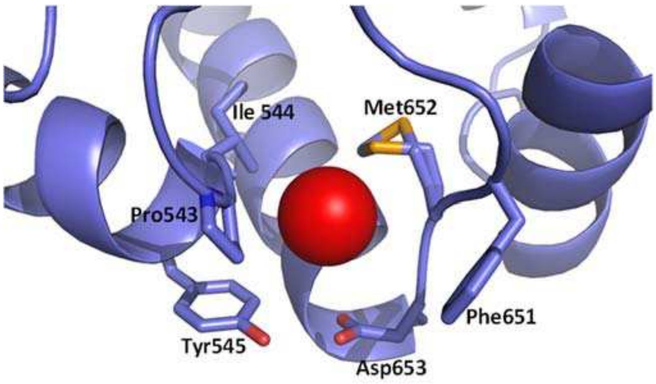


Figure 1- 9 Main residues surrounding the bound anion (red sphere).

This suggests a fundamental role of the domain in the steps preceding the transport of anionic species across the membrane through the prestin TM domain (figure 1-10).

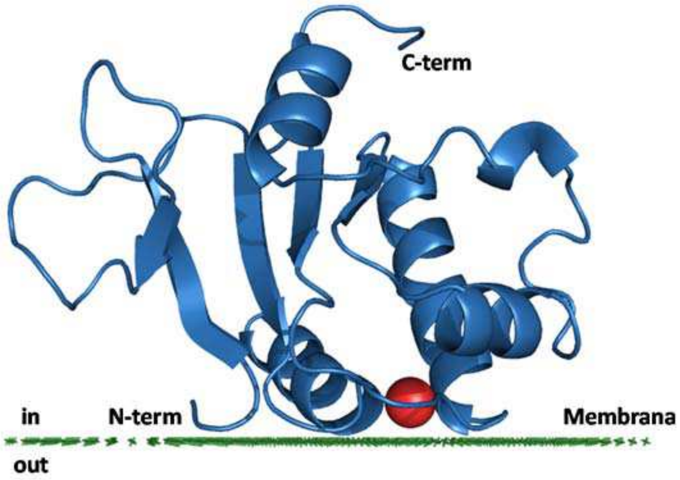


Figure 1- 10 In blue the structure of wtSTAS (PDB: 3LLO); the red sphere corresponds to a Cl^- anion, the green crosses show the hypothetical location of the TMD/membrane bilayer.

The STAS domain and calmodulin

It was also shown that calmodulin binds directly to the IDR (intrinsically disordered region) in the C-terminal STAS domain of prestin in a calcium-obligate manner (Keller et al., 2014) as shown in figure 1-11. Using both isolated murine outer hair cells (OHCs) and a heterologous expression system, it was found that this calcium-obligate CaM binding shifts the operating point of the protein to more hyperpolarized potentials with consequent alteration of the function of the prestin. These results suggest possible regulation mechanisms of the function of prestin, and, maybe, of other members of the SLC26 family.

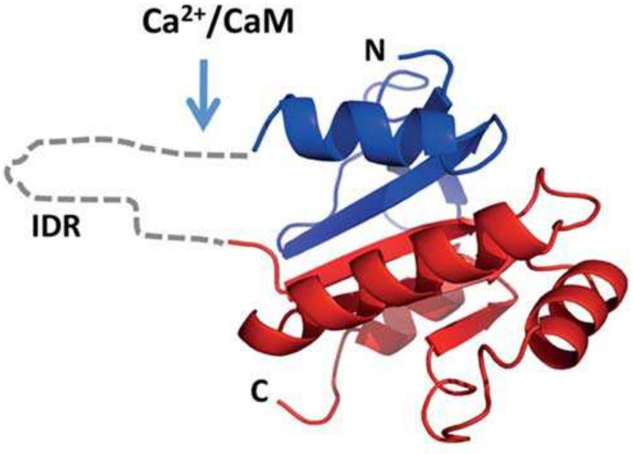


Figure 1- 11 CaM binding site in the C-terminal domain of rat prestin (Keller et al., 2014).

SECs were performed both in presence of calcium and in presence of EGTA using human calmodulin and a 30-aa long peptide from gerbil prestin corresponding to the binding region. Fluorescence polarization was also exploited to calculate the Kd, that were 86.5 nM in presence of calcium and 73.4 μM in presence of EDTA (figure 1-12).

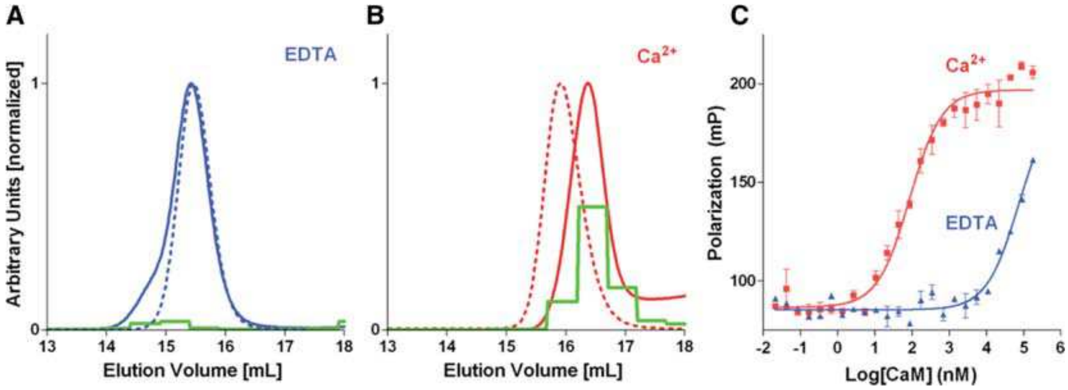


Figure 1- 12 Biochemical assays of the CaM–SLC26 interaction. A, B, SEC of CaM in the presence (blue and red solid lines) and absence (blue and red broken lines) of fluorescein-labeled prestin peptide under EDTA (A) or CaCl2 (B) conditions. Fluorescence intensities are shown with solid green lines. C, Fluorescence polarization assay ofCaMversus labeled prestin peptide in the presence (red) or absence (blue) of calcium. Error bars represent SDs of three measurements, with fits to the Hill equation superimposed on data points. (Keller et al., 2014).

The transmembrane domain of prestin

Recently, a reliable structural model of the TMD of prestin a has been derived by combining homology modeling and molecular dynamics (MD) simulations, revealing that the transmembrane core region of prestin comprises 14 TM segments organized in a 7 + 7 inverted repeat architecture (figure 1-13, Gorbunov et al, 2014). Using domain swapping between rPres and zPres together with measurements of transport currents and capacitance Oliver et al., had already identified structural elements required for piezoelectricity: two distinct stretches in the rPres polypeptide, one comprising transmembrane domains 1 and 2 and the adjacent linker (aa 93–136; NLC-domain 1), the other consisting of aa 381–438; (NLC-domain 2).

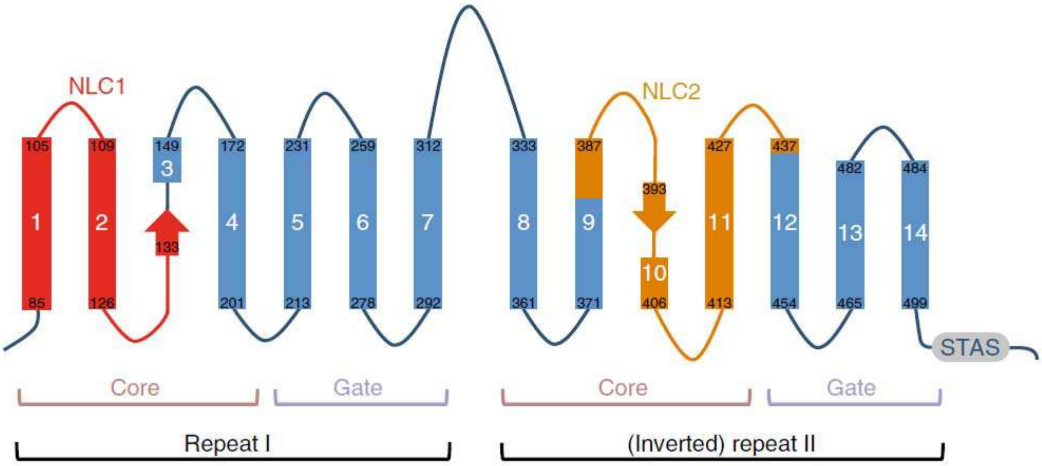


Figure 1- 13 Schematic representation of the topology and structural motifs of rPres. Cylinders represent helices, arrows represent the two b-strands b3 and b10 in TM3 and TM10. Protein regions NLC1 and NLC2 are highlighted in red and orange, respectively. Topological localization of repeats I and II and of core and gate domains are indicated. Cytoplasmic C and N termini not drawn to scale.

Using a substituted cysteine accessibility method (SCAM) scan covering all predicted intra- and extra-cellular domains, the theoretical model was experimentally validated. It was also possible to identify a central cavity (figure 1-14) as the major anion-binding site that controls the electromotile activity of prestin, explaining the known anion dependence of the electromechanical activity of the protein.

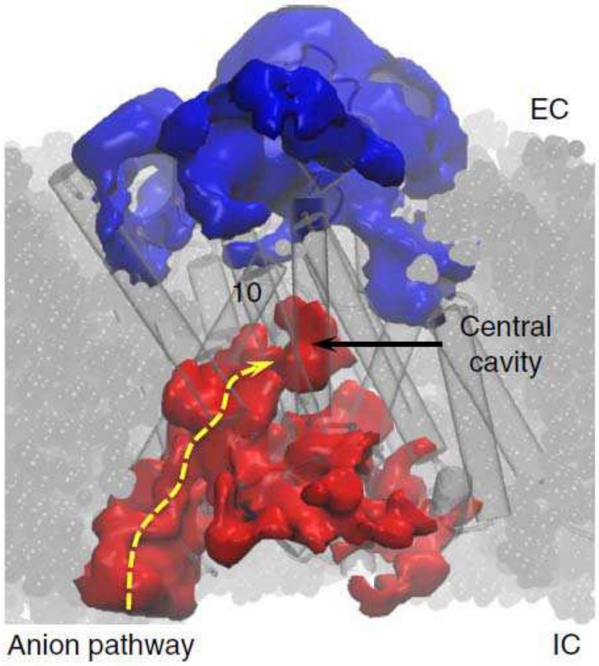


Figure 1- 14 TMD cavities identified by analysis of solvent trajectories are shown in red (cavities accessible from cytoplasm, IC) and blue (accessible from extracellular solution, EC). The intracellular access pathway to the central cavity (yellow dashed line) was identified as a continuous solvated channel with a diameter compatible with passage of oxalate.

Aim of the work

Prestin is the anion-dependent motor protein responsible for the outer hair cells electromotility, at the basis of the increased sensitivity and frequency selectivity of the hearing process in mammals. The crystal structure of the STAS domain of rat prestin is known and regarding the TM domain, an experimentally validated structural model has been recently proposed. It is known that intracellular anions are crucial for the functionality of prestin. Over the past few years in our group were collected diffraction data from crystals of the STAS domain of rat prestin (a mammal) in which the anions known to be necessary for the functionality of the whole protein were present, either from co-crystallization or from soaking experiments. We also recently determined the crystal structure of the STAS domain from chicken prestin, the first one from a non-mammalian vertebrate. Given that these datasets were processed with different methods over the time and to take advantage of the most recent improvements in data processing and structure refinement, and in order to see the fine details of the structures and of the interactions of the different anions, the first aim of my work was to start from the diffraction images and develop a homogeneous procedure of data reduction, structure solving and refinement for all of the six structures.

It is also known that the STAS domain interacts with calmodulin (CaM) with residues of the flexible variable loop, which is normally removed for crystallization purposes. I intend to investigate the STAS/CaM complex by SEC (Size Exclusion Chromatography), DLS, X-ray Crystallography and/or SAXS. The first step is the recombianant production of the two proteins: calmodulin and the STAS domain of rat prestin, including the variable loop.

Experimental procedures

Recombinant protein expression

ratPrestin STAS domain [505-718]

The coding sequences of the STAS domain of prestin of *Rattus norvegicus* [505-718] had already been cloned in a pETSUMO and expression vector. The construct was previously used to transform *E. coli* BL21(DE3) cells. Bacterial clones were then conserved at -80 °C. Cells were grown in liquid medium LB (Luria-Bertani; Sambrook and Russel, 2001): 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl. The medium was sterilized by autoclaving it for 20 minutes at 121 °C and implemented with 50 µg/ml kanamycin, to which bacterial cells transformed with the pETSUMO vector are resistant. *E. coli* cells were grown over-night (ON) in LB medium containing 50 µg/ml kanamycin at 37°C in a suitable shaker. LB medium supplied with 50 µg/ml kanamycin was then inoculated with this over-night culture in the ratio 1:10. The culture was kept at 37°C in a suitable shaker and the growth was followed measuring the OD_{600} (Optical Dispersion at 600 nm, directly related to the amount of cells in suspension). The temperature was slowly decrease until 23°C when an OD_{600} of 0.6 was reached, then protein expression was induced by adding IPTG to a final concentration of 1 mM when an OD_{600} of 0.8 was reached. After induction, bacteria were growth over-night at 20 °C. Cells were then harvested by centrifugation (20 min, 4000 g, 4°C). The harvested cells were resuspended in buffer A: 50 mM Na_2HPO_4, 300 mM NaCl, 10 mM β-mercaptoethanol, pH 8.0 supplemented with 2% of buffer B: 50 mM Na_2HPO_4, 500 mM Imidazole, 300 mM NaCl, 10 mM β-mercaptoethanol, pH 8.0 and EDTA-free protease inhibitors and lysed with a “French Press” cell disrupter at 1000 psi. Cell debris were then removed from the soluble components extracted by centrifugation of the lysate (45 min, 12000 g, 4°C). After this centrifugation step, the supernatant was filtered using 0.22 µm filters before being loaded onto the column for the first step of the purification procedure.

calmodulin

The coding sequences of human had already been cloned in a pET28 and expression vector. The construct was previously used to transform *E. coli* BL21(DE3) cells. Bacterial clones were then conserved at -80 °C. Cells were grown in liquid medium LB (Luria-Bertani; Sambrook and Russel, 2001): 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl. The medium was sterilized by autoclaving it for 20 minutes at 121 °C and implemented with 50 µg/ml kanamycin, to which bacterial cells transformed with the pET28 vector are resistant. *E. coli* cells were grown over-night (ON) in LB medium containing 50 µg/ml kanamycin at 37°C in a suitable shaker. LB medium supplied with 50 µg/ml kanamycin was then inoculated with this over-night culture in the ratio 1:10. The culture was kept at 37°C in a suitable shaker and the growth was followed measuring the OD_{600} (Optical Dispersion at 600 nm, directly related to the amount of cells in suspension). The temperature was slowly decrease until 20°C when an OD_{600} of 0.6 was reached, then protein expression was induced by adding IPTG to a final concentration of 1 mM when an OD_{600} of 0.8 was reached. After induction, bacteria were growth over-night at 20 °C. Cells were then harvested by centrifugation (20 min, 4000 g, 4°C). The harvested cells were resuspended in 50 mM Tris-HCl pH 7.5, supplemented with EDTA-free protease inhibitors and lysed with a “French Press” cell disrupter at 1000 psi. Cell debris were then removed from the soluble components extracted by centrifugation of the lysate (45 min, 12000 g, 4°C). After this centrifugation step, the supernatant was filtered using 0.45 µm filters before the beginning of the purification procedure.

Purification protocols

An AKTApurifier 10 with a multiple-wavelength detector was used for the chromatographic steps of the purification procedures. The chromatographic runs were followed at 280 nm and chromatograms were recorded using the UNICORN software (GE healthcare).

ratPrestin STAS domain [505-718]

The filtered supernatant obtained after the mechanical lysis of the bacterial cells was loaded onto a 1 ml *HisTrap HP* (GE healthcare) affinity column, exploiting the presence of the $(His)_6$-tag on the protein. The column was equilibrated with 2% buffer B and the sample was loaded. The fusion protein was eluted with a step gradient procedure with increasing concentrations of imidazole, first 20 mM, then 150 mM, 300 mM and finally 500 mM using a flow rate of 1 ml/min.

The $(His)_6$-SUMO-STAS[505-718] protein was eluted with 150 mM imidazole, and since the SUMO protease (life sensors) is sensitive to the presence of imidazole, it was necessary to remove it with a desalting chromatographic step exploiting a HiPrep 26/10 desalting coloumn. The proteolytic cleavage was then performed over-night at 4°C. After the proteolysis, the sample was reloaded onto the affinity column used before. This time the STAS protein elutes in the flow through while uncleaved $(His)_6$-SUMO-STAS and the SUMO protease (that also has an $(His)_6$-tag) are eluted using imidazole, first 150 mM and then 500 mM.

The sample was then concentrated by ultrafiltration, using Vivaspin concentrators. Then it was loaded onto a HiLoad 16/60 preparation grade Superdex 75 column equilibrated with 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5, to eliminate possible aggregate forms and place the protein in the most appropriate buffer system for the following studies. The fractions corresponding to ratSTAS[505-718] were collected and concentrated to 10-20 mg/ml. The aliquots were then flash-frozen using liquid nitrogen and conserved at -80 °C.

calmodulin

The filtered supernatant obtained after the mechanical lysis of the bacterial cells was mixed with 5 ml of Phenyl FF resin (GE Healthcare) in ice for 1 h and 15 minutes to remove highly idrophobic proteins. The mixture was then centrifuged for 15 min at 5000 g at 4 °C. The supernatant was filtered with 0.45 and 0.2 µm filters and $CaCl_2$ was added at a final concentration of 5 mM. The supernatant was then loaded onto a HiTrap Phenyl FF(LS) (GE Healthcare) 5 ml column, equilibrated with buffer A: 50 mM Tris-HCl, 1 mM $CaCl_2$, pH 7.5. After the elution of the Flow through the column was washed with buffer B: 50 mM Tris-HCl, 1 mM $CaCl_2$, 100 mM NaCl, pH 7.5 and then calmodulin was eluted with buffer C: 50 mM Tris-HCl, 1 mM EDTA, pH 7.5. The sample was then concentrated by ultrafiltration, using Vivaspin concentrators. Then it was loaded onto a HiLoad 16/60 preparation grade Superdex 75 column equilibrated with 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, to eliminate possible aggregate forms and place the protein in the most appropriate buffer system for the following studies. The fractions corresponding to calmodulin were collected and concentrated to 20-30 mg/ml. The aliquots were then flash-frozen using liquid nitrogen and conserved at -80 °C.

SDS-PAGE

The purification procedures were followed by SDS-PAGE. Samples were heated for 5 minutes at 95 °C to denature them in the sample-loading buffer: 60 mM Tris pH 6.8, 10% glycerol, 2% SDS, 0.01% bromophenol blue, 2% β-mercaptoethanol. Then they were loaded with Low molecular weight See Blue® Plus2 Pre-Stained Standard (*Invitrogen)* markers on SDS-PAGE gels. A voltage of 180 V was used to performed the electrophoretic run with the gels mounted vertically on the electrophoretic apparatus and in contact with the running buffer: 25 mM Tris pH 8.3, 250 mM glycin, 0.1% SDS. The gels were colored for 20 minutes under soft shaking with the staining solution: 40% EtOH, 10% AcOH, 0.1% w/v Coomassie G-250 and then the destaining solution: 40% EtOH, 10% AcOH and a final wash in water were used to remove unspecific staining.

Dynamic Light Scattering

The dynamic light scattering measurements were performed using a Zetasizer Nano™ (*Malvern instrument*) and a low-volume quartz cuvette. Samples were centrifuged at 12000 rpm for 10 minutes to remove dust particles and high molecular weight aggregates; then the measurements were performed at 20 °C.

Analytical Size Exclusion Chromatography

SEC runs were perforemed using a Superdex 75 10 300 GL column (GE Healthcare) equilibrated with a buffer consisting of 20 mM tris pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$ or of 20 mM tris pH 7.4, 150 mM NaCl, 1 mM EGTA.

Protein crystallization

Crystallization trials on the ratSTAS/CaM complex were performed with the Oryx8 (Douglas instruments) automatic system using commercially available crystallization screens.

Single Crystal X-ray Crystallography

X-ray diffraction datasets were collected at the XRD1 beamline of the ELETTRA Synchrotron (Basovizza, TS) at 100 K.

Diffraction data were initially processed (indexing, integration and scaling) with the XDS package (Kabsch, 2010). Integrated reflections were then reduced and merged with AIMLESS (Evans, 2006) (Evans, An introduction to data reduction: space group determination, scaling and intensity statistics, 2011) from the CCP4 suite (Winn, et al., 2011). The structures were solved by molecular replacement using the previously deposited model with the PDB ID 3LLO as search model and the program PHASER (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The initial structures were then refined alternating manual steps of real space refinement using the program COOT (Emsley & Cowtan, 2004) (Emsley, Lohkamp, Scott, & Cowtan, 2010) and automatic steps of reciprocal space refinement using the suite PHENIX (Adams, et al., 2010). PYMOL was used to make all the images and the superposition analysis.

Small Angle X-ray Scattering

For the structural characterization of the STAS/CaM complex in solution SAXS data were collected at the ESRF BM29 BioSAXS beamline (Pernot et al, 2013), both in batch and in the SEC-SAXS mode, with a SEC column coupled to SAXS measurements. For batch SAXS measurements, 10 frames on a solution of 3.6 mg/ml STAS/CaM complex in 20 mM tris pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$ were collected. As solvent background, 10 frames were collected before and after sample measurements and subtracted to sample scattering curves. The STAS/CaM complex was previously isolated with a SEC run on a Superdex 75 10 300 GL column (GE Healthcare) column. For SEC/SAXS measurements, 100 µl of STAS/CaM complex at 5 mg/ml in 20 mM tris pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$ were loaded onto a Superdex 75 10/300 size exclusion column (GE Healthcare), extensively equilibrated in the same buffering system, at a flow rate of 0.5

ml/min, using the beamline Shimadzu HPLC system. Solution eluting from the column was redirected to the sample flow path for immediate SAXS data collection. All scattering data (batch and SEC/SAXS) were collected using a Pilatus 1M 2D detector from Dectris (detector distance 2.8720 m), corresponding to a momentum transfer s (4πsinθ/λ) range of 0.08-4.5 nm−1, with 2 s exposure time per frame (approximately 2000 frames were collected per SEC run), wavelength 0.99190 Å, temperature 293 K. Initial data processing was performed automatically using the pipeline for bioSAXS data implemented at the ESRF BM29 beamline using the EDNA framework (Brennich et al, 2016) and analysed with the laboratory-information management system ISPyB (De Maria Antolinos et al, 2015). Further primary data reduction and analysis on selected SAXS data was performed manually with PRIMUS (Konarev et al, 2003) from the ATSAS 2.8.0 program package (Petoukhov et al, 2012). Useful experimental data ranges were determined with SHANUM (Konarev and Svergun, 2015), discarding data at higher angles containing no useful information. For SEC/SAXS, the first 1000 frames were automatically averaged by the pipeline to yield an "average buffer" solvent profile that was then subtracted from frames classified as "sample" (Brennich et al, 2016). To avoid possible interference of the single components, analysis on the STAS/CaM complex was performed on the first part of the SEC elution profile, averaging frames in the 1200-1220 range, according to the invariance of the values for the radius of gyration RG, the QR parameter (Rambo et al, 2013) and the molecular mass. Scattering curves are reported in nm-1, s=4πsin(θ)/λ. Ab initio modeling with the single phase dummy atom model approach was performed with DAMMIN/DAMMIF (Svergun, 1999; Franke and Svergun, 2009). Quaternary structure rigid body modeling of the STAS/CaM complex using the crystal structures of the isolated STAS and CaM was performed with SASREF (Petoukhov and Svergun, 2005) and with CORAL for the addiction of missing fragments (Petoukhov et al, 2012).

Results and Discussion

The STAS domain and the anion binding site

Rat prestin was chosen as a model for mammalian prestin and diffraction data from crystals of the STAS domain of in which the different monovalent anions known to be necessary for the functionality of the whole protein were present were collected in my group over the past few years. These datasets were collected on crystals obtained either from co-crystallization or from soaking experiments. Also, the crystal structure of the STAS domain from chicken prestin, the first one from a non-mammalian vertebrate, was recently determined. Given that these datasets were processed with different methods over the time and to take advantage of the most recent improvements in data processing and structure refinement, I started working on them from the diffraction images and developed a homogeneous procedure of data reduction structure solving and refinement for all of the six structures, allowing us to see the fine details of the structures and of the interactions of the different anions.

The final structures ended up in a paper titled: "The STAS domain of mammalian SLC26A5 prestin harbors an anion-binding site." published on the "Biochemical Journal" (Lolli et al, 2016) which was also object of a recent commentary published on the same journal (Birke and Javelle, 2016).

As regards the five crystal structures of the mammalian ratSTAS domain in presence of the monovalent anions (chloride, iodide, bromide, thiocyanate and nitrate), the main feature is the presence of the anion-binding site. Compairing the structures with the apo one (PDB: 3LLO), it is evident that the cavity remains unchanged upon anion binding indicating that the binding site is preformed, ready to accommodate and release anions without the need of conformational rearrangements (figure 1-15).

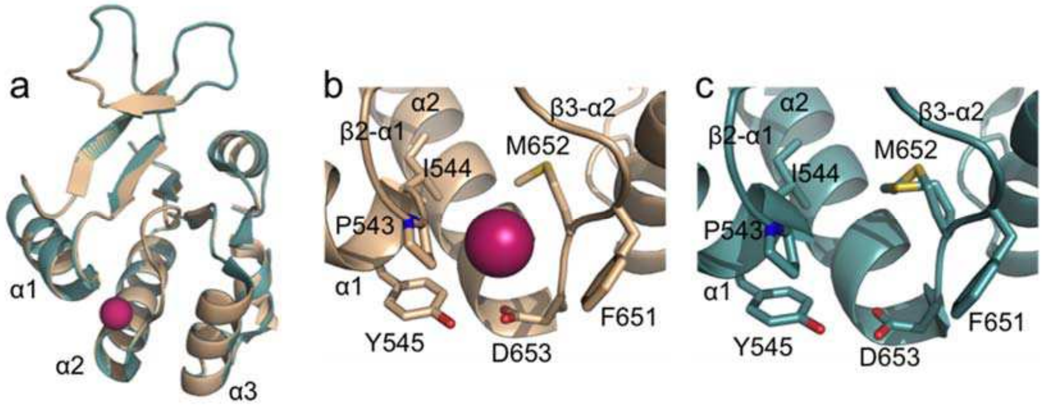


Figure 1- 15 The anion-binding site in rPres (rat prestin) STAS. (a) Superposition of rPres STAS in presence (wheat) and absence (cyan) of bound chloride (magenta). Secondary structure elements and the overall 3D architecture of the domain are conserved. Close-up view of the anion-binding site in presence (b) and absence (c) of chloride, showing that binding does not alter the local structure (Lolli et. al, 2016).

The anion-binding site is a relatively large and deep cavity, situated in a region proposed to be facing the TM domain. The cavity has a positive potential, mainly due to the dipoles of backbone NHs from the N-termini of helices α1 and α2 (figure 1-16).

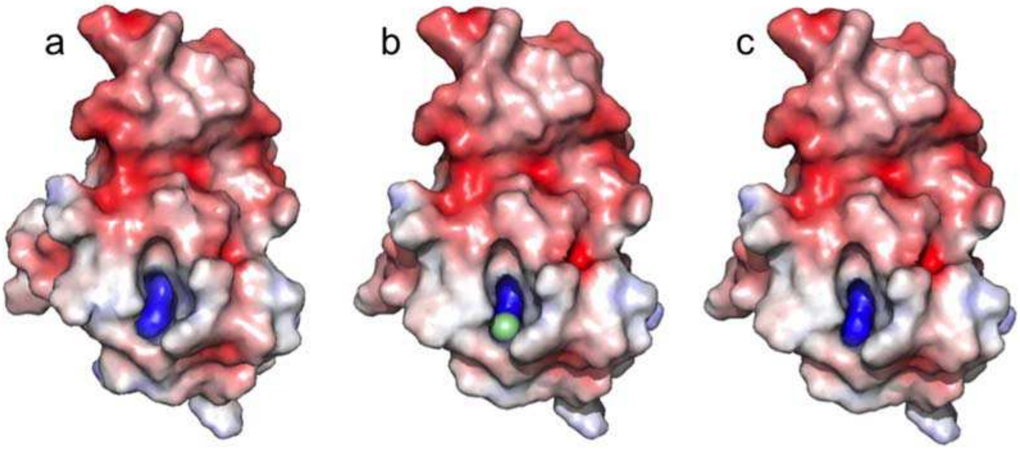


Figure 1- 16 Molecular surface of rPres STAS colored according to electrostatics. In blue positive and in red negative potentials. (a) apo structure (PDB 3LLO), (b) with bound chloride (green sphere), (c) as in (b) without the anion. The binding site is a deep cavity with a positive potential that is not altered by anion binding.

As regards the interaction that the anions make upon binding, NHs establish ion-dipole interactions with Cl^-, Br^- and I^-. Thiocyanate and nitrate make hydrogen bonds with I544 and D653 NHs (figure 1-17).

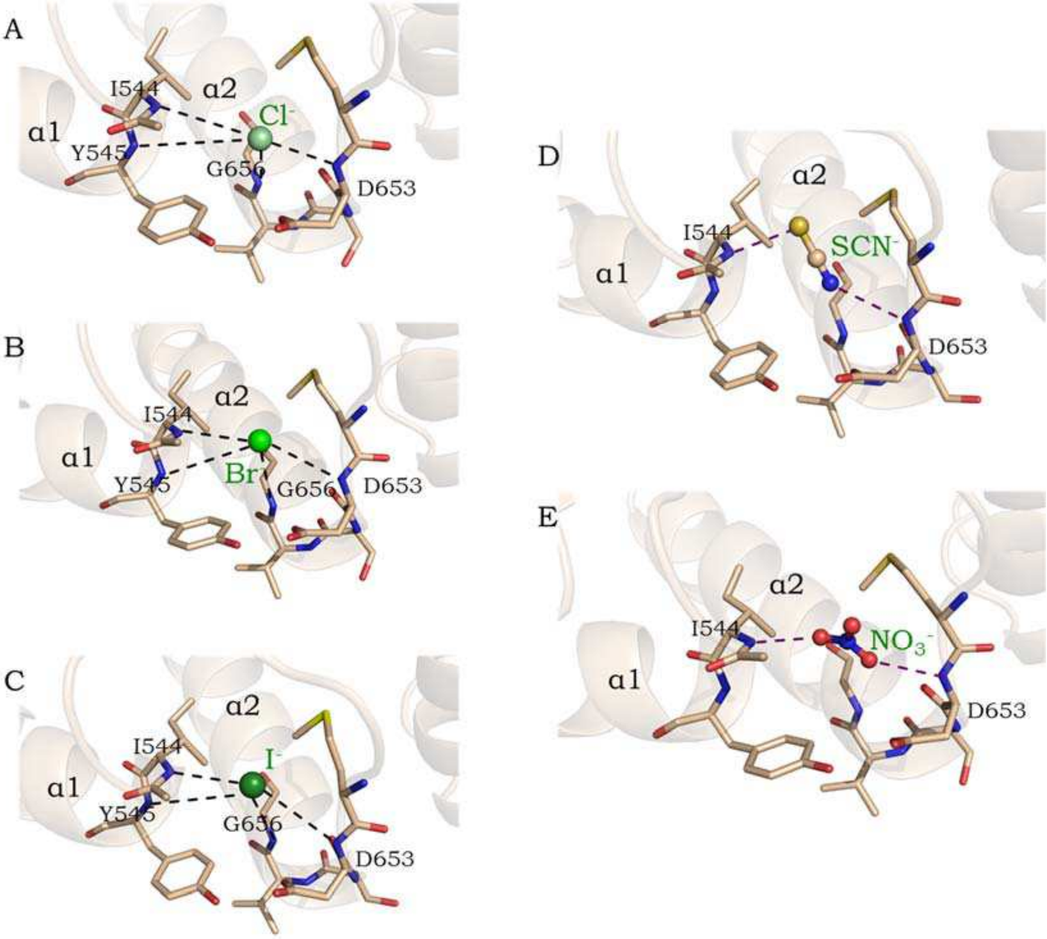


Figure 1- 17 Main interactions of anions bound to rPres STAS. Ion-dipole interactions with the N-terminal NHs of helix α1 (I544 and Y545) and helix α2 (D653 and G656) are indicated by black dashed lines for chloride (a), bromide (b) and iodide (c). Hydrogen-bonds with NHs of I544 and D653 (magenta dashed lines) are shown for thiocyanate (d) and nitrate (e).

The anion-binding site is specific for the monovalent anions that are necessary for the function of the full-length prestin. In fact, even though the crystallization condition (1.8 M Ammonium sulfate, 5% (v/v) PEG400, 0.1 M MES (2-*N*-morpholino-ethanesulfonic acid) pH 6.5, 0.1% octyl-β-D-glucopyranoside) contains sulphate at high concentration, this divalent anion is never present in the cavity, which therefore seems specific for monovalent anions. This is probably due to the lack of positive charges in the cavity, that can therefore accommodate only monovalent anions though ion-dipole interactions.

In order to prove the presence of the different anions, when possible, the diffraction experiment was carried out using x-ray wavelengths at which it was possible to assess the anomalous signal of the element. Anomalous scattering, or anomalous dispersion, takes place when using an x-ray beam with an energy near the absorption maximum of an element (typically with high atomic number) present in the crystal. Therefore, diffraction data were collected at 2 Å to see the anomalous signal of chloride, iodide and the sulfur atom in thiocianate and at 0.915 Å for bromide. Figure 1-8 shows the values of the components f' and f' in function of the x-ray wavelength used for the diffraction experiment.

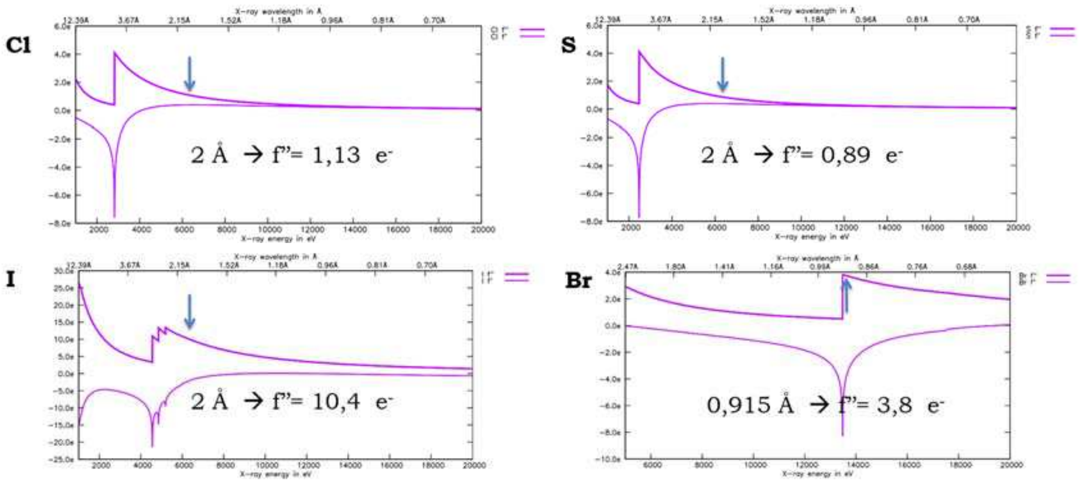


Figure 1- 18 Anomalous dispersion for the atoms under study in function of the x-ray wavelength. The arrows indicate the entity of f” at the wavelengths used in the diffraction experiments.

Figure 1-19 shows the anomalous maps contoured at 3σ for the anions bound in the binding site.

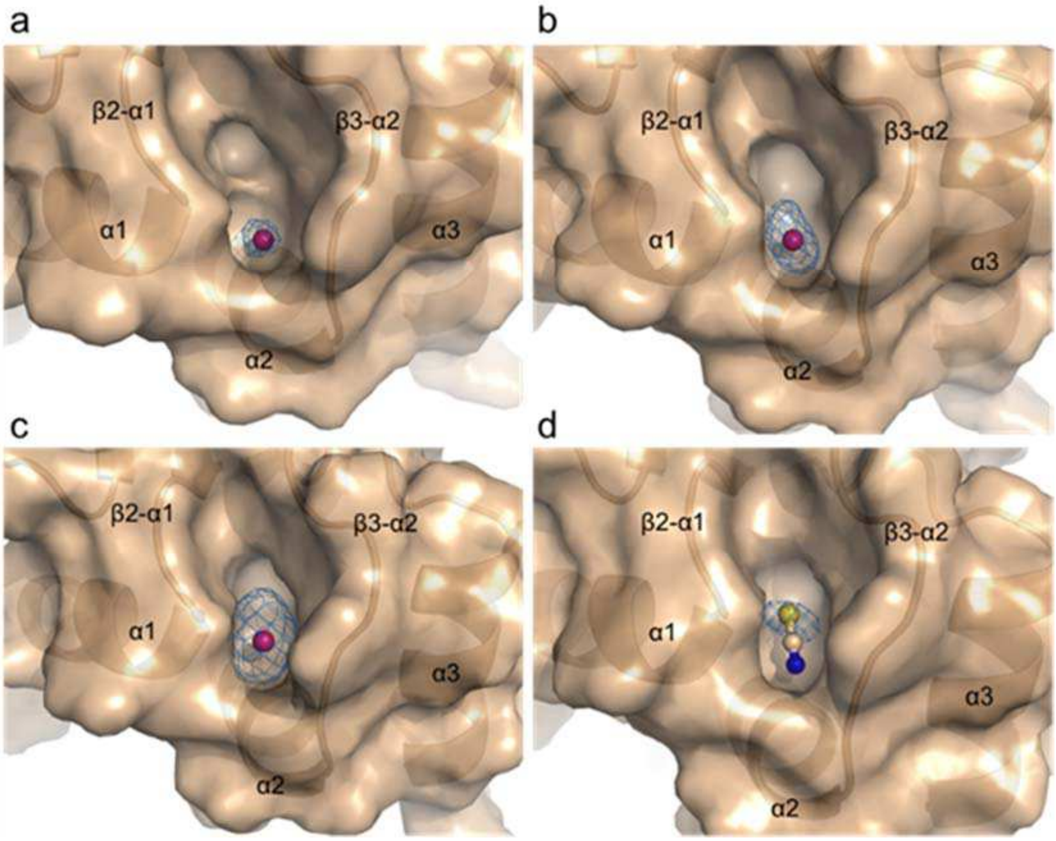


Figure 1- 19 Anomalous maps contoured at 3σ for chloride (a), bromide (b), iodide (c) and sulphur of thiocyanate (d) bound to rPres STAS. Anomalous data were collected at 2 Å for chloride, iodide and sulphur, and at 0.915 Å for bromide (Lolli et al, 2016).

Usually X-ray diffraction experiments are performed using wavelengths in the range of 0.9-1.2 Å. The use of the 2.0 Å wavelength implied some technical difficulties compared to a "standard" diffraction experiment, in the preparation (setting) of the beamline, in the data collection and in the data processing. For instance, with 2.0 Å wavelength there is a non negligible absorption of the diffracted X-rays by the air, therefore the crystal-to-detector distance was set at the minimum possible value, i.e. 75 mm.

The diffraction data were initially processed (indexing, integration and scaling) with the XDS program package, optimizing several parameters of processing in order to obtain the best reflection intensities statistics given the high background and the low signal/noise ratio (figure 1-20).

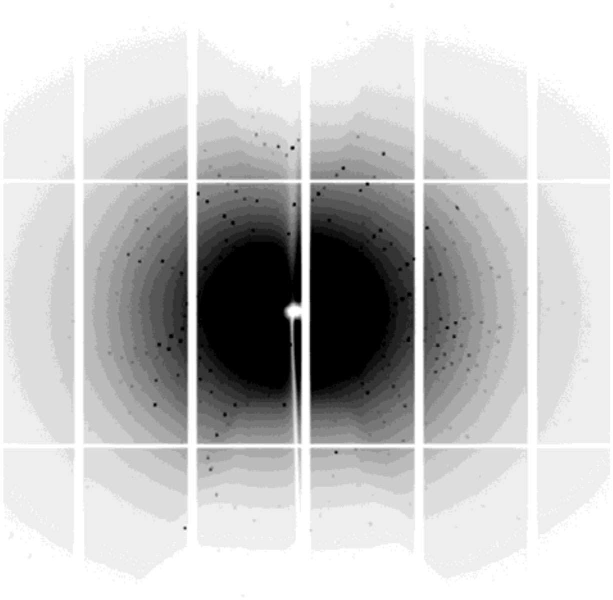


Figure 1- 20 A representative diffraction image of the ratSTAS+Cl data collection. Note the high background near the image center and the low signal/noise ratio due to the use of the 2 Å X-ray radiation necessary to detect the (weak) anomalous signal of chloride.

All of the structures were solved by molecular replacement using the apo rat STAS structure (PDB: 3LLO) as search model. The refinement procedure was optimized for the 6 structures, leading to the crystallographic statistics summarized in table 1-2.

	rPres_Cl	rPres_Br	rPres_I	rPres_SCN	rPres_NO$_3$	cPres
Data collection						
Wavelenght (Å)	2	0,915	2	2	1,2	0,915
Space group	P3(1)21	P3(1)21	P3(1)21	P3(1)21	P3(1)21	I2
Cell dimensions						
a b c (Å)	62,13 62,13 66,59	61,41 61,41 66,42	61,91 61,91 66,49	62,37 62,37 66,75	62,22 62,22 66,87	81,6 56,9 101,1
α β γ (°)	90 90 120	90 90 120	90 90 120	90 90 120	90 90 120	90 96,19 90
Resolution (Å)	41,85-1,87 (1,91-1,87)	43,51-1,83 (1,87-1,83)	41,74-2,4 (2,49-2,40)	41,99-2,04 (2,09-2,04)	41,96-1,81 (1,85-1,81)	46,59-2,3 (2,38-2,3)
R_{sym}	0,071 (0,749)	0,078 (1,998)	0,161 (1,733)	0,081 (1,138)	0,075 (1,143)	0,069 (0,665)
R_{meas}	0,075 (0,861)	0,082 (2,111)	0,172 (1,861)	0,085 (1,219)	0,082 (1,249)	0,076 (0,725)
< I/σI >	26,2 (2,2)	25,0 (1,7)	14,9 (1,7)	22,2 (2,1)	15,0 (1,7)	18,9 (2,8)
Completeness (%)	98,8 (84,2)	99,9 (98,4)	99,8 (98,1)	97,4 (87,3)	99,9 (98,5)	96,1 (96,9)
Multiplicity	15,6 (7,0)	19,5 (17,9)	14,7 (13,9)	17,4 (14,2)	6,4 (6,2)	6,4 (6,3)
Refinement						
Resolution (Å)	31,01-1,87	41,51-1,83	41,74-2,4	41,99-2,04	41,96-1,81	46,59-2,30
Number of reflections	23464	24555	11066	18020	14033	19851
R_{work}/R_{free}	15,07/19,6	15,3/19,7	18,8/22,1	16,3/19,8	16,4/19,0	21,0/25,7
B-factors (Å^2)						
protein	34,0	41,4	50,6	48,4	34,1	50,8
ligands/ion	53,5/43	60,6/69,4	58,0/69,4	69,6/44,5	59,4/31,9	62,1/62,0
r.m.s. deviations						
bond lenght (Å)	0,003	0,011	0,003	0,016	0,008	0,016
bond angles (°)	0,764	1,219	0,793	1,259	1,057	1,432
Ramachandran plot						
favoured (%)	100	99	99	99	99	96
outliers (%)	0	0	0	0	0	0

Table 1- 2 Data collection and refinement statistics

As regards the structure of the STAS domain of chicken prestin, the overall architecture resembles that of the rat homologue.

The major structural difference between chicken and rat STAS domains is at the level of the β2-α1 loop and the following α1 helix. The β2-α1 loop is closer to the β3-α2 loop, with the consequence that the anion-binding site is absent in chicken STAS indicating that anion-binding is a property peculiar to the mammalian domain. The interaction between loops β2-α1 and β3-α2 in chicken is stabilized by the formation of a short parallel β-sheet. The sequence around loop β2-α1 is fully conserved in mammals (INAPI), while is different and variable in non-mammals (figure 1-21).

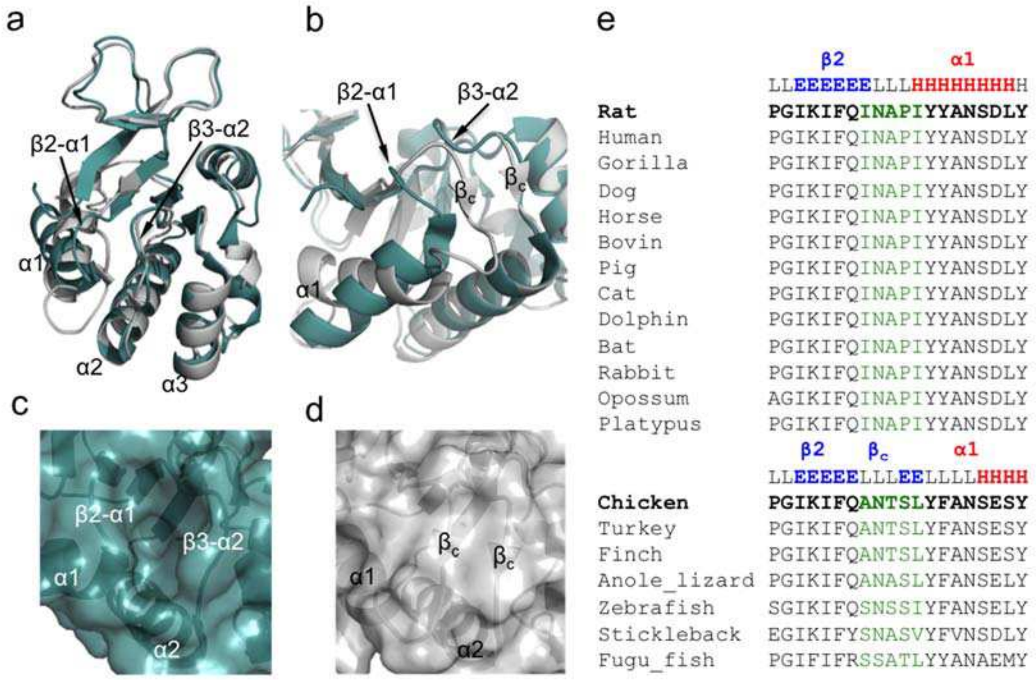


Figure 1- 21 Comparison of rPres and cPres STAS domains. (a) Superposition of rat (cyan) and chicken (grey) STAS shows that the secondary structure elements and the overall 3D architecture are generally conserved. The most relevance differences are the orientation of helix α1 and the position of loop β2-α1 that in chicken interacts with loop β3-α2 with the formation of a short β-sheet (βC) (b). The consequence is that the rPres anion-binding site (c) is absent in cPres (d). (e) Alignment of residues around the β2-α1 segment showing conservation of the INAPI stretch in mammals and sequence variation in non-mammals. Secondary structures (L = loop, E = β-strand, H = helix) are indicated.

All of these results led us to the following conclusion. The STAS domain of mammalian prestin, unlike the non-mammalian one, harbors a unique anion-binding site, the first interaction site of the anion before the translocation in the TM domain. While the TM domain is the main responsible for NLC/electromotility in mammals and transport in non-mammals, the STAS domain is responsible for the fine tuning of these properties. We therefore proposed that the anion-binding site in the STAS domain of mammalian prestins represents a reservoir of ready- to-use anions, increasing their local concentration, for the fast rate at which this molecular motor must operate (up to 120 kHz in the ultrasonic hearing of some cetaceans). This rapidness is not necessary for transport in non-mammalian prestin for which the presence of a STAS anion-binding site does not seem essential.

STAS and calmodulin – protein expression and purification

RatPrestin STAS domain [505-718] and calmodulin were recombinantly produced in bacteria and purified using chromatographic methods as described in the experimental procedures.

ratPrestin STAS domain [505-718]

As regards the STAS domain of ratPrestin, the (His)6-tag was exploited for the purification procedure, which consisted of 4 steps: 1) Immobilized Metal-ion Affinity Chromatography – IMAC, using a Ni^{2+}-NTA resin; 2) proteolytic cleavage of the His-SUMO-tag with the specific SUMO-protease; 3) removal of the His-SUMO-tag from the solution by Ni^{2+}-NTA IMAC; 4) SEC (Size Exclusion Cromatography). Figure 1-22 shows the chromatogram of the first step of the purification procedure.

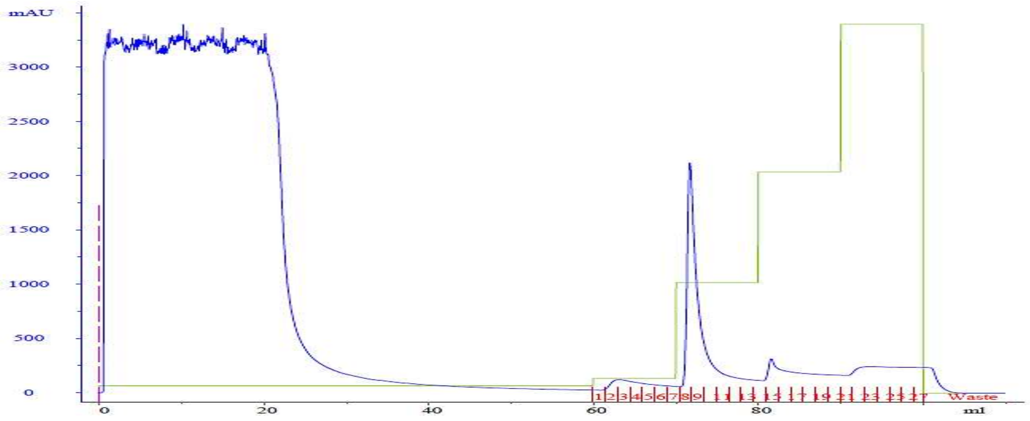


Figure 1- 22 ratPrestin STAS domain [505-718] IMAC affinity chromatogram; in blue the elution profile (Abs at 280 nm), in green the step gradient of imidazole, in red the collected fractions.

The majority of the recombinant protein was eluted in the following step, with 30% buffer B, and therefore 150 mM imidazole, along with some impurities (F8, F9 and F10). Fractions containing the protein (F8, 9 and 10) were pooled, concentrated by ultrafiltration and the buffer was replaced with buffer A, removing the imidazole used for the elution using a HiPREP 26/10 desalting column (GE Healthcare). This is necessary to ensure the optimal SUMO protease activity. The sample was incubated with the SUMO protease (Life sensors) overnight at 4 °C for the proteolytic cleavage of the $(His)_6$-SUMO-tag. The resulting hydrolyzed material was applied onto the His-Trap column and the flow-through (containing the $(His)_6$-SUMO-tag free STAS domain) was recovered while the uncleaved fusion protein and the protease (fused to an $(His)_6$ tag) remain bound to the column. Figure 1-23 shows the chromatogram of this second purification step, where the protein of interest does not interact with the stationary phase, having lost the $(His)_6$-tag, while the uncleaved residues, the $(His)_6$-SUMO, and the SUMO protease are eluted afterwards using increasing quantities of imidazole.

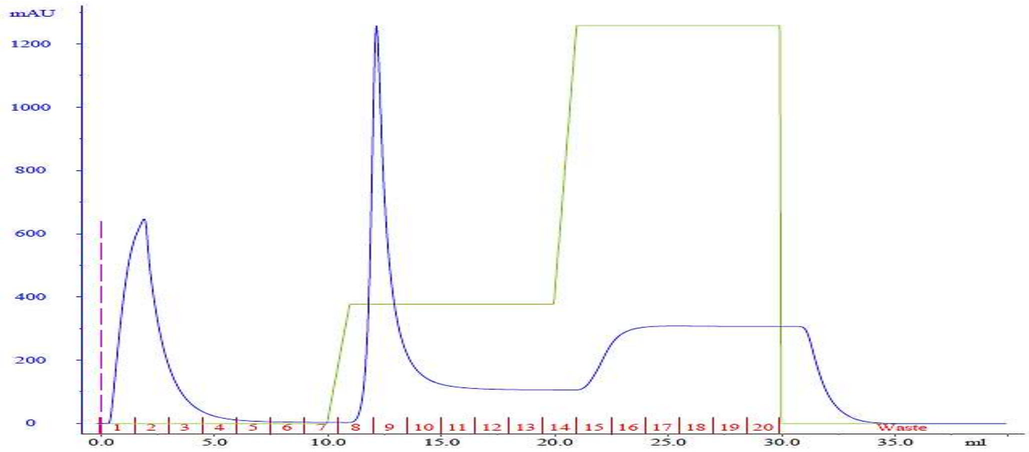


Figure 1- 23 Second IMAC affinity chromatogram for ratPrestin STAS domain [505-718], after proteolytic removal of the (His)6-SUMO-tag with sumo protease, in blue the elution profile (Abs at 280 nm), in green the step gradient of imidazole, in red the collected fractions.

The eluate was further purified by size exclusion chromatography, using a HiLoad 16/60 preparation grade Superdex 75 column (GE

Healthcare), equilibrated with the buffer 20 mM Tris-HCl, 150 mM NaCl, 5 mM DTT, pH 7.5 at 1 ml/min. Figure 1-24 shows the chromatograms for this final purification step.

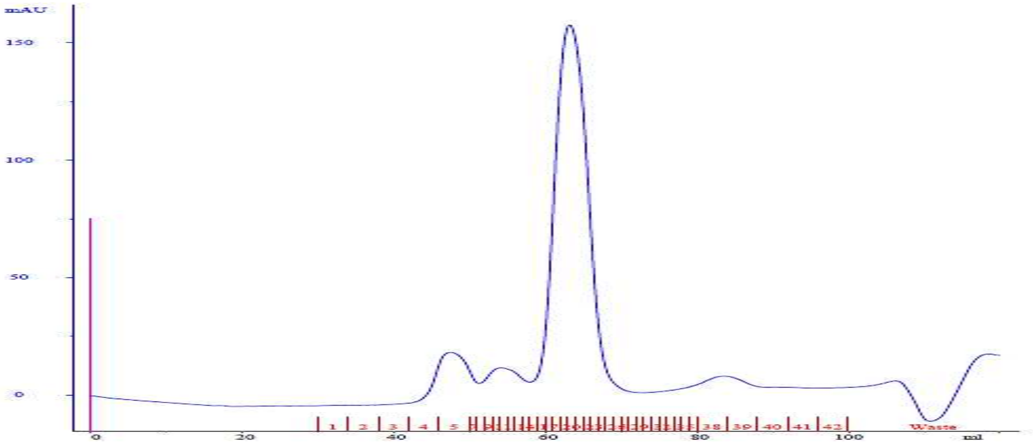


Figure 1- 24 ratPrestin STAS domain [505-718] Size exclusion chromatography chromatogram.

The proceeding of the purification procedure was followed by SDS-PAGE as shown in figure 1-25.

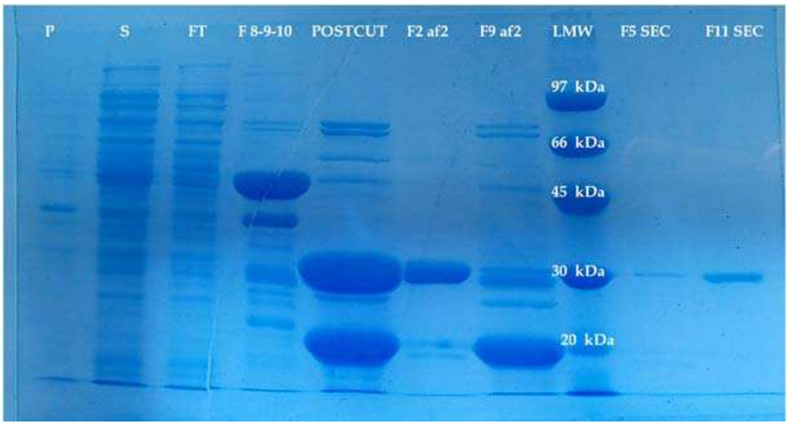


Figure 1- 25 Coomassie stained SDS-page of the fractions of the purification procedure for ratPrestin STAS domain [505-718]. LMW: markers of known MW; P: lysate pellet (insoluble fraction); S: lysate supernatant (soluble fraction); FT: "flow through", unbound proteins; FX: fraction n° X along the elution gradients.

As can be seen in the gel, the fusion protein was eluted from the first affinity chromatography in fractions F8-9-10, then the cut with the sumo protease was successful (POSTCUT lane). The cleaved STAS and the (His)6-SUMO-tag were separated in the second affinity chromatography (F2 af2 and F9 af2). In the last two lane, some aggregates forms of the STAS domain present in the first two peaks of the SEC. Fractions of the peak corresponding to the monomeric protein in the SEC chromatography where also collected and analyzed by SDS-page (figure 1-26) and the band corresponding to the STAS domain is the only one visible in the gel.

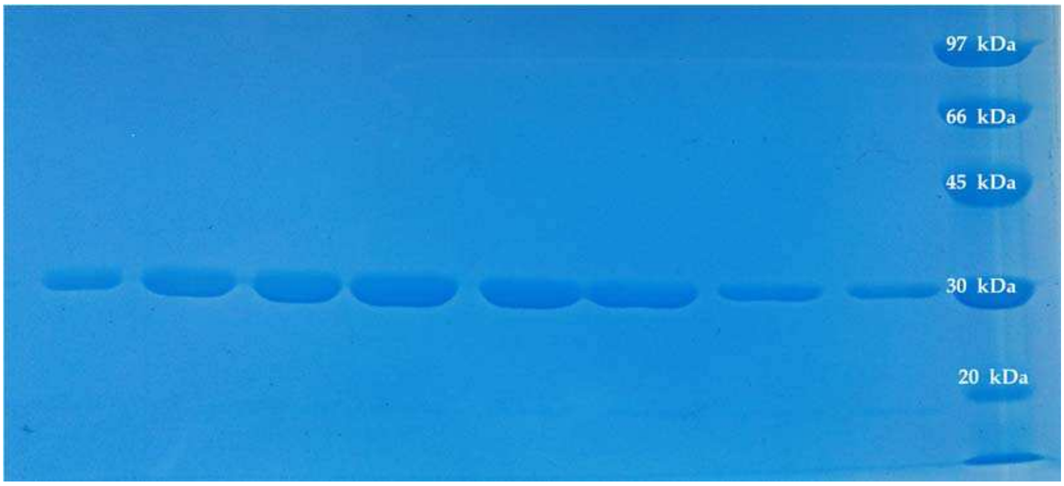


Figure 1- 26 Coomassie stained SDS-page of the fractions F16-F23 of the final size exclusion chromatography.

calmodulin

As regards the purification of calmodulin, the property of the protein to expose hydrophobic residues in presence of calcium was exploited for the purification procedure, that started with an in batch removal of highly hydrophobic proteins by incubation of the supernatant with 5 ml of Phenyl FF resin (GE Healthcare). Calcium was then added to the resulting supernatant, allowing calmodulin to bind to the hydrophobic resin of a HiTrap Phenyl FF(LS) (GE Healthcare) 5 ml column. After the elution of the flow through, the column was washed with buffer B, containing 100 mM NaCl, to remove potential proteins bound unspecifically through ionic interactions. Then calmodulin was eluted with buffer C, containing 1 mM EDTA. Figure 1-27 shows the chromatogram of this hydrophobic interaction step of the purification.

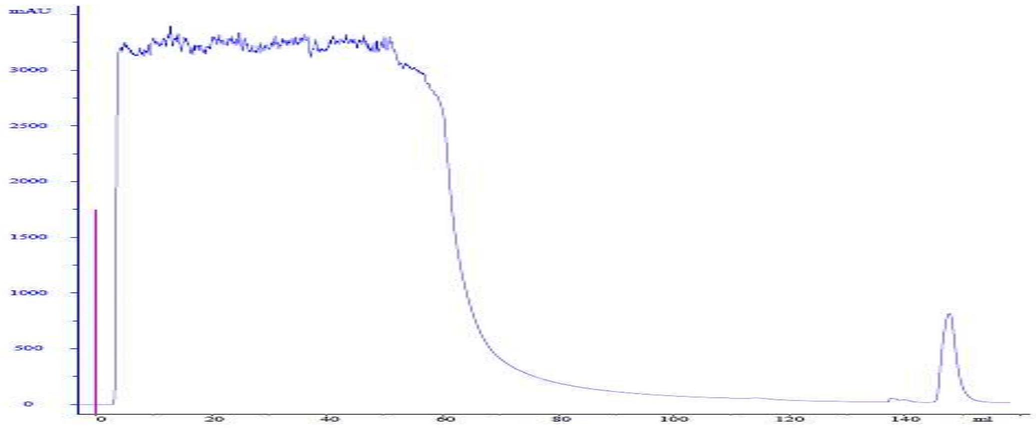


Figure 1- 27 calmodulin hydrophobic interaction chromatogram. In blue the elution profile (Abs at 280 nm).

The eluted peak was further purified by Size Exclusion chromatography, as shown in figure 1-28. The most pure fractions of the peak were pooled, concentrated and flash-frozen in liquid nitrogen for storage at -80 °C.

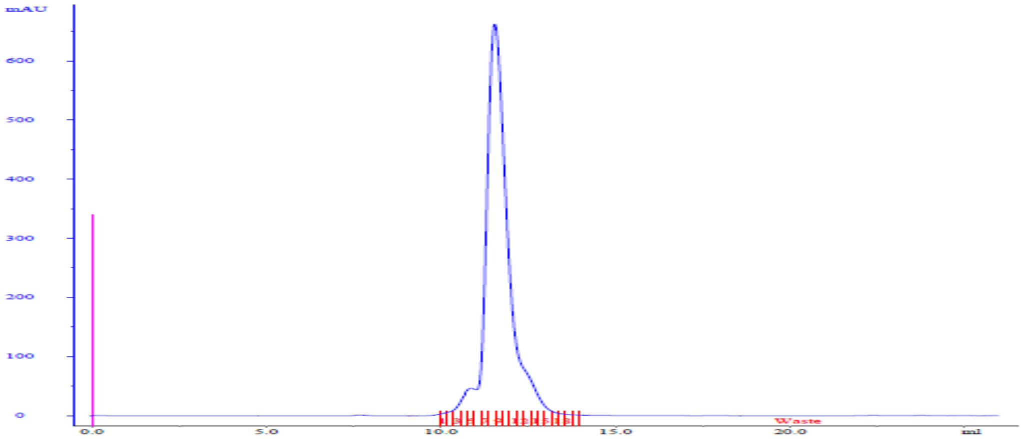


Figure 1- 28 calmodulin size exclusion chromatography chromatogram

The proceeding of the purification procedure was followed by SDS-PAGE (figure 1-29).

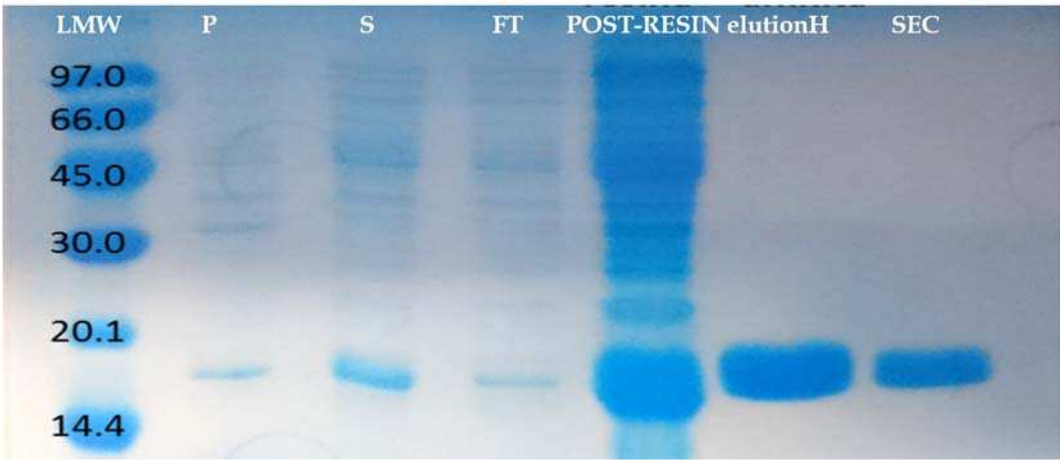


Figure 1- 29 Coomassie stained SDS-page of the fractions of the purification procedure for calmodulin. LMW: markers of known MW; P: lysate pellet (insoluble fraction); S: lysate supernatant (soluble fraction); FT: "flow through", unbound proteins; POST-RESIN: Sn after the in batch purification step; elutionH: protein eluted from the Hydrophobic interaction column, SEC: final purified protein after the SEC column.

STAS and calmodulin – Size Exclusion Chromatography

SEC and DLS experiments were performed by mixing the two proteins at a 1:1 molar ratio, in the presence of 5 mM calcium with a 30 minutes long incubation before the experiments.

SEC runs were perforemed using a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with a buffer consisting of 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM $CaCl_2$. Figure 1-30 shows the elution profiles of the two proteins alone and of the complex formed in the presence of 5 mM calcium.

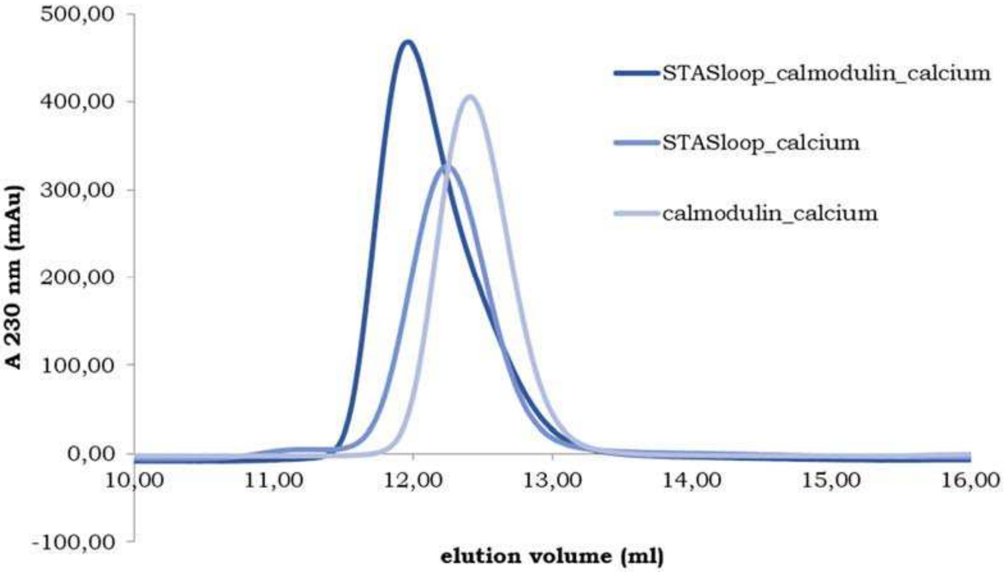


Figure 1- 30 SEC elution profiles (Abs at 230 nm). Absorption was followed at 230 nm due to the low molar extinction coefficient of calmodulin at 280 nm.

As a negative control, a run was also performed using a different construct of the STAS domain [505-563]GS[637-718], lacking the variable loop and therefore the region of the interaction. In this case the complex is not formed and the two proteins elute separately as can be seen in figure 1-31, confirming that the region of the interaction is in the variable loop.

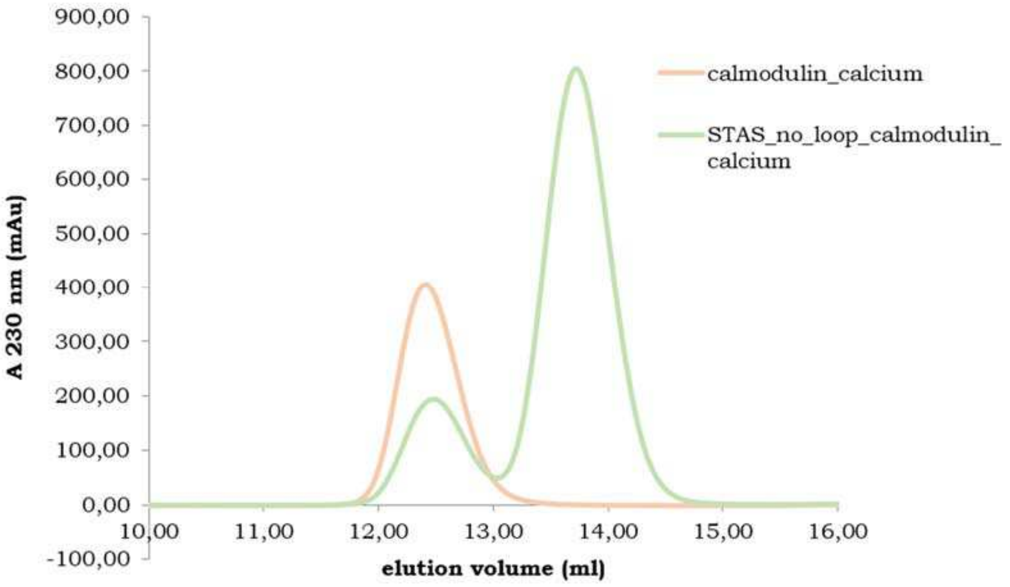


Figure 1- 31 SEC elution profiles (Abs at 230 nm)

These results prove that the complex between the STAS domain and calmodulin is formed in solution, and that the variable loop is fundamental for this formation.

The fractions of the peak of the complex were analyzed by SDS-PAGE as shown in figure 1-32.

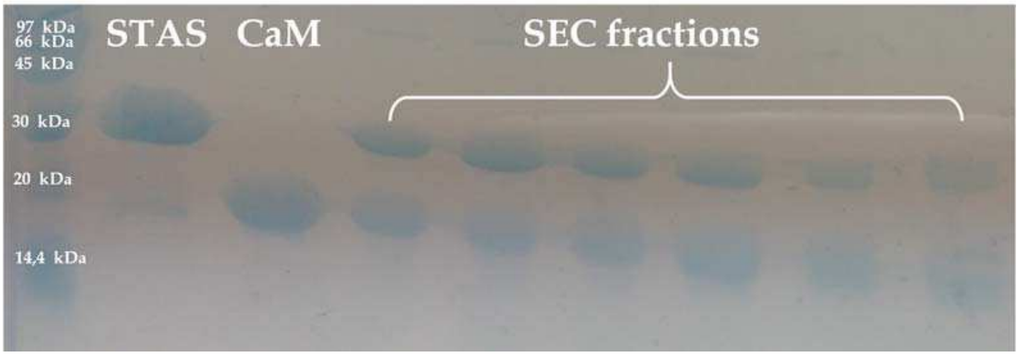


Figure 1- 32 Coomassie stained SDS-page of the fractions of the Size Exclusion chromatography.

Sec runs were also performed in presence of EGTA, as shown in figure 1-33. In this case the elution volume of calmodulin and og the STAS domain are the same. When loaded together, they elute at that same volume. The most probable explanation is that they co-elute but do not interact, since nothing is eluted at lower volumes. Therefore the complex formation is probably calcium-dependent.

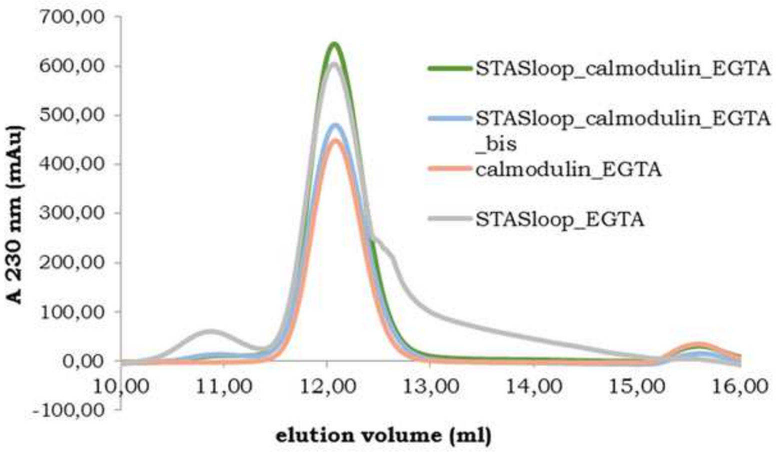


Figure 1- 33 SEC elution profiles (Abs at 230 nm)

STAS and calmodulin – Dynamic Light Scattering (DLS)

The formation of the complex in solution was also confirmed by DLS (figure 1-34).

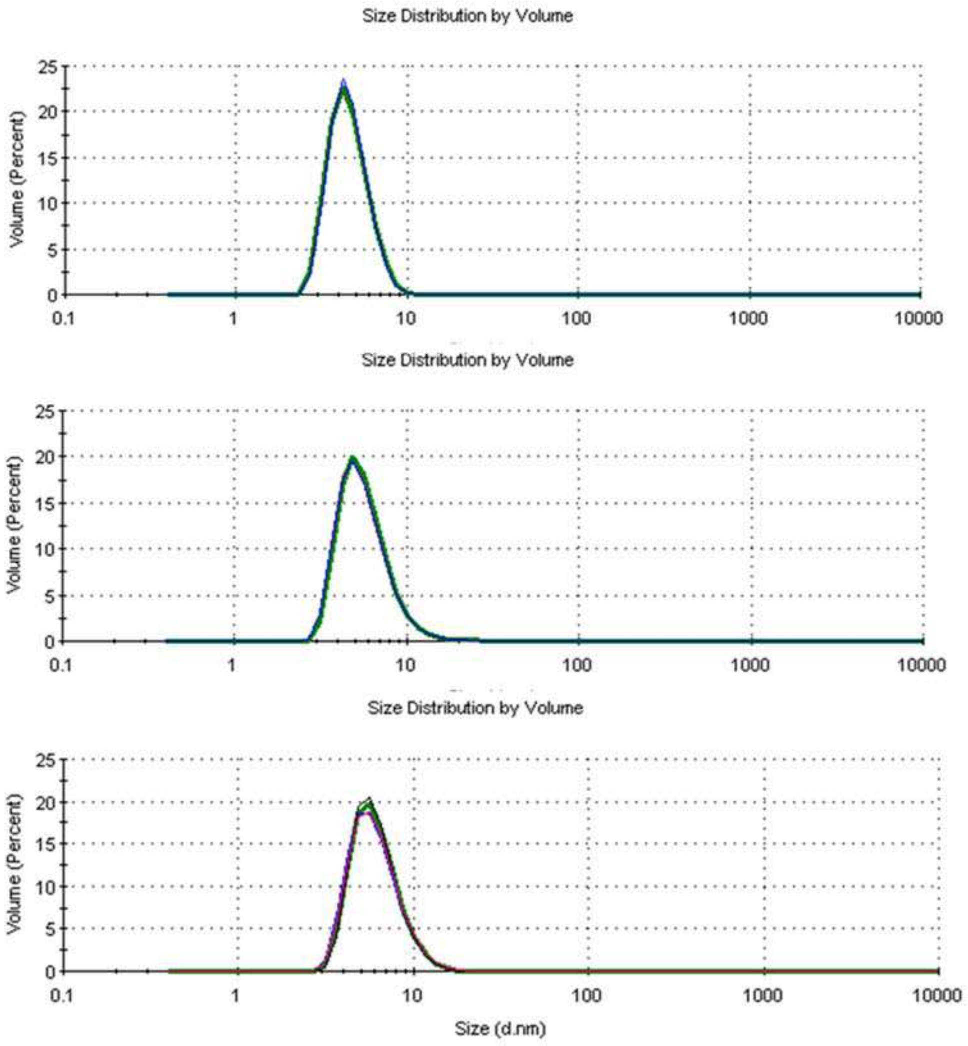


Figure 1- 34 DLS measurements on calmodulin, ratSTAS [505-718] and their complex.

The Stokes radius estimated from these measurements were 2.7 nm for calmodulin, 4 nm for ratSTAS [505-718] and 4.2 nm for their complex.

STAS and calmodulin – Small Angle X-ray Scattering (SAXS)

As regards the crystallization trials, no crystals were obtained so in the future new constructs of the STAS domain keeping the interaction region but removing the rest of the flexible variable loop, will have to be designed. I therefore performed SAXS (Small Angle X-ray Scattering) experiments on the complex in solution.

SAXS scatting profiles and primary analyses of the data collected in batch and with the SEC/SAXS mode are shown in figures 1-35, 1-36 and in table 1-3.

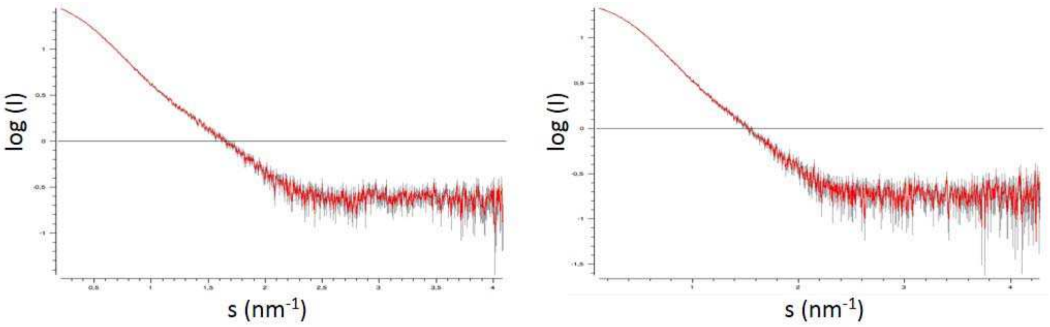


Figure 1- 35 SAXS scattering profiles of the STAS/CaM complex measured in batch (left) and in SEC/SAXS mode (right).

The radius of gyration R_G, the maximum particle size D_{max} the Porod volume and the values of the molecular weight calculated in four different modes (Table 1-3) are all compatible with the presence of a monodisperse STAS/CaM complex with a theoretical MW of 40.2 kDa.

	Primary data analysis							Modeling				
	R_G^1	R_G^2	D_{max}	Vol	MW^1	MW^2	MW^3	MW^4	NSD	$(\chi^2)^1$	$(\chi^2)^2$	$(\chi^2)^3$
Batch	2.82	2.82	9,99	71.7	42.2	37.8	40,4	40.2	0.54	0.730	0.95	1.15
SEC	2.67	2.71	9.54	62.9	37.0	34.0	31.9	36.0	0.54	0.716	-	-

Table 1-3 SAXS data analysis R_G (nm): radius of gyration derived by Guinier plot (R_G^1) and by the calculation of the particle distance distribution function P(r) by GNOM (5) (R_G^2).D_{max} (nm): maximum particle size by GNOM (5). Vol: Porod volume. MW (kDa): molecular weight derived by the Porod volume (MW^1), the invariant "volume of correlation" Vc (7) (MW^2), the SAXS MoW2 package (11) (MW^3) and the DAMMIN *ab initio* dummy atom modeling (MW^4). NSD: Normalized Spatial Discrepancy measuring the similarity between DAMMIF models (NSD < 1 implies similar models). χ^2: goodness of fit from DAMMIN $(\chi^2)^1$, SASREF $(\chi^2)^2$ and CORAL $(\chi^2)^3$.

The tail at high distances in the density distribution function (figure 1-36) is in accordance with a flexible region present in the C-terminal part of the long variable loop of the STAS domain.

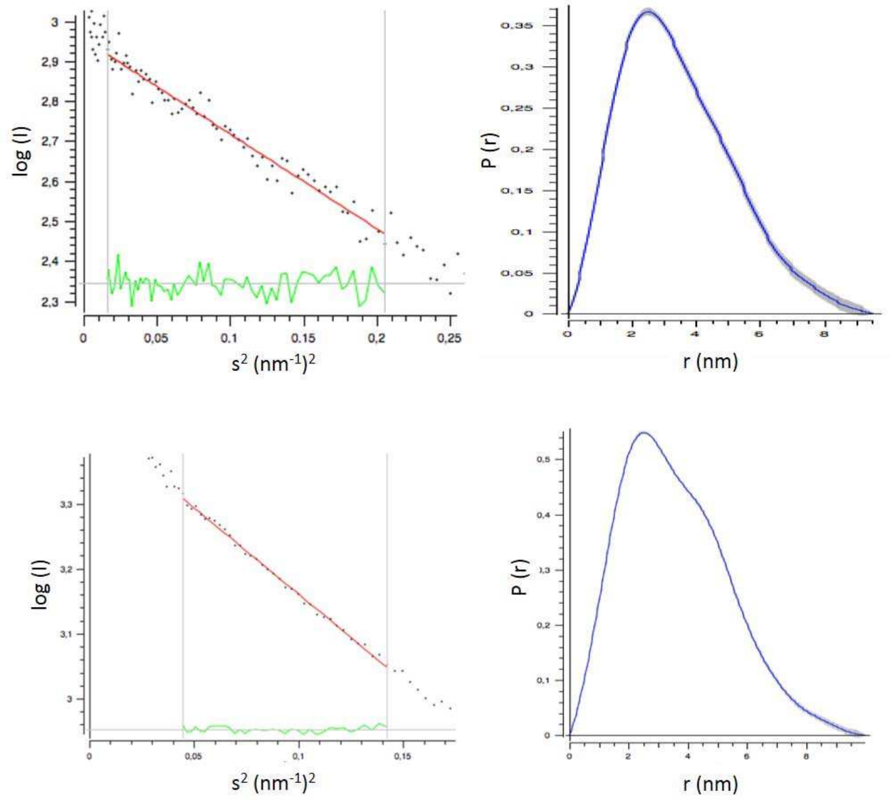


Figure 1- 36 Linear interpolations in the Guinier region for the determination of the radius of gyration (left) and density distribution functions calculated by GNOM (right). Upper panels from SEC/SAXS measurements, lower panels from batch data.

The shape of the scattering object in solution was derived by ab initio modeling with the DAMMIN/DAMMIF procedure (Franke and Svergun, 2009), revealing an L-shaped elongated form for the STAS/CaM complex, with good χ2, 0.730 and 0.716, and good fittings between the calculated and experimental scattering curves (figure 1-37).

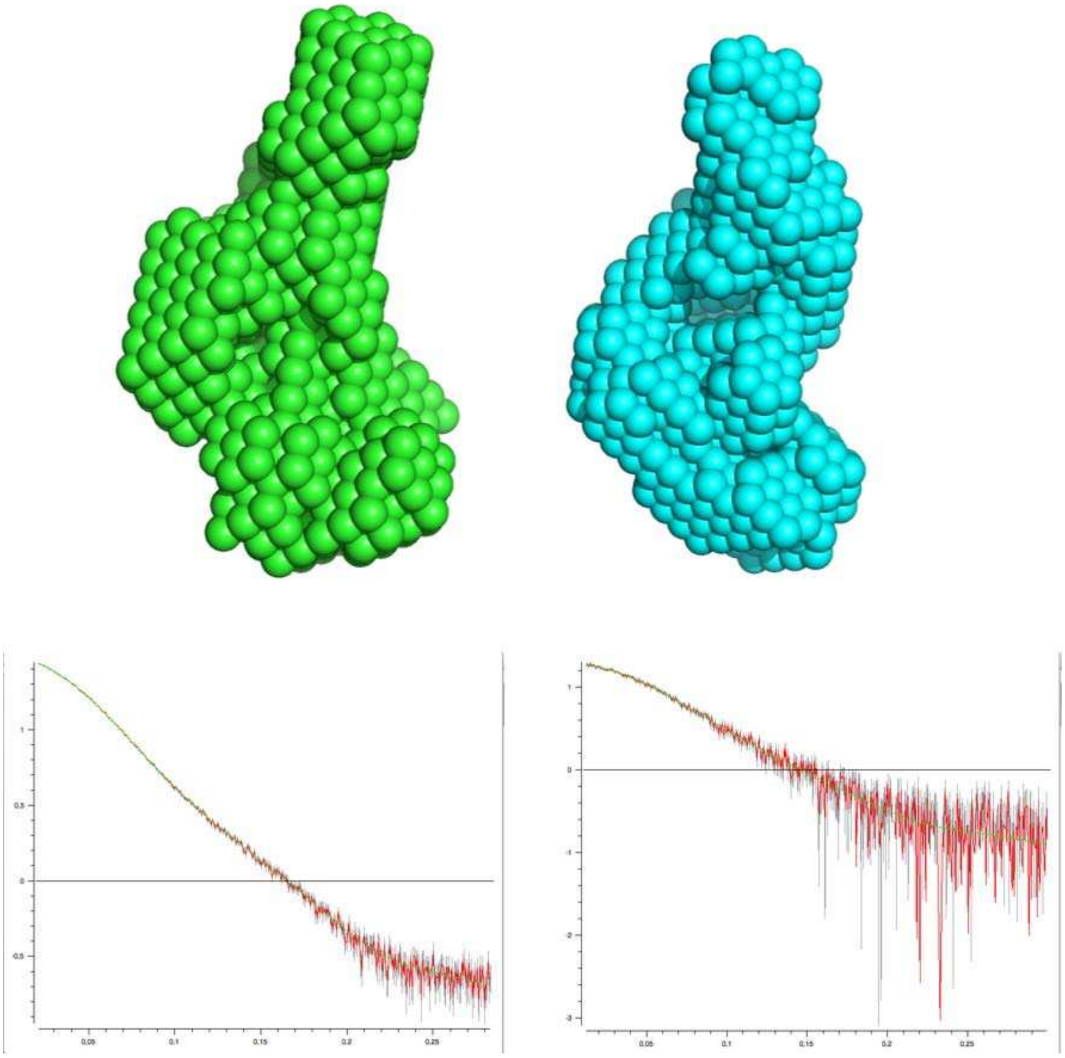


Figure 1- 37 DAMMIN *ab initio* dummy atom models derived from batch (upper left) and SEC/SAXS (upper right) SAXS data. The same orientation after superposition of the two models by SUPCOMB (5) is shown. Fittings of the calculated (green) and experimental (red) scattering profiles are show in the lower-left panel (batch, χ^2=0.730) and in the lower-right panel (SEC/SAXS, χ^2 0.716).

The Normalized Spatial Discrepancy (NSD) value of 0.54 (Table 1-3), computing the similarities between all pairs of the 10 solutions with equally good fit, confirms the good quality of ab initio models.

To obtain a three dimensional model of the STAS/CaM complex several runs of rigid body modeling with SASREF and CORAL were performed. Given the higher quality and signal-to-noise ratio of the batch scattering profile (see figures 1-35, 1-36, 1-37) this particular modeling was performed only on batch data. In this particular case, for rigid body modeling two major problems are the following: a) the presence of a long flexible stretch of amino acids in the STAS domain (around 80 residues), most of which is recognized as an intrinsically disorder region (IDR), that is missing in the crystal structure of the STAS domain; b) the very diverse conformations adopted by CaM when binding to target proteins (Tidow and Nissen, 2013) that hampers an unique obvious choice for the CaM model. A constant in the CaM-complex structures is that the partner adopts an alpha-helical conformation in the CaM-binding site (CBS), with a length of around 16-20 residues. Interestingly, the region of the STAS variable loop expected interacting with CaM (Keller et al, 2014) is predicted to adopt a helical conformation around residues 565-582. For rigid modeling the strategy was then to use different structures of CaM in complex with helical peptides the mimics the helical conformation of the STAS domain region 565-582 interacting with CaM. In particular models representative of CaM in extended, partially-compact or fully-compact conformations were selected (three of which are shown in figure 1-38).

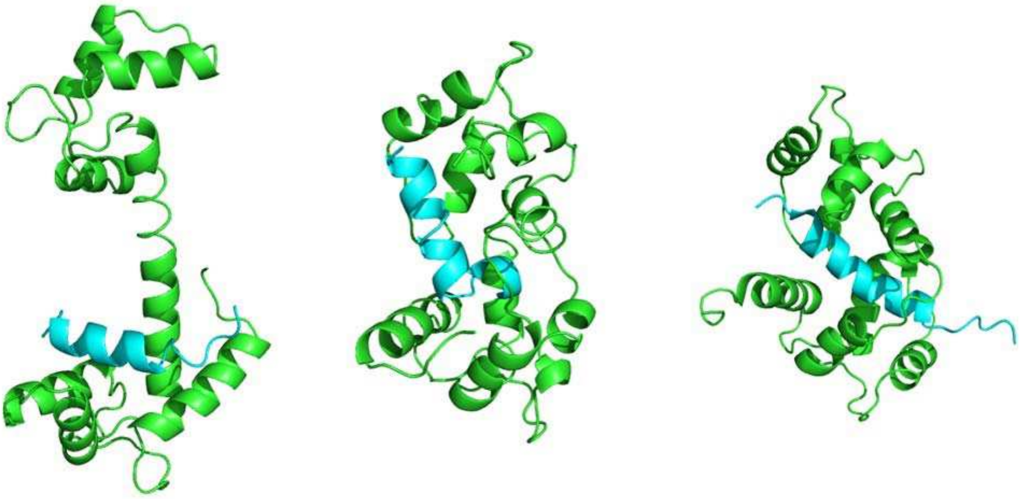


Figure 1- 38 . Models of CaM/peptide complexes representative of extended (left, PDB-code 3SJQ), partially-compact (centre, 5DOW) and fully-compact (right, 2BBM) conformations.

The best results in terms of curve fittings and χ2 values (see table 1-3) were obtained with CaM in the extended conformations, in particular using the structure of CaM in complex with a SK2 small conductance potassium channel (3SJQ, figure 1-38 left). For the STAS domain, the structure in apo form with PDB-code 3LLO was used. The best model of the STAS/CaM complex resulting from SASREF and CORAL simulations is shown in figure 1-39, with χ2 of 0.95.

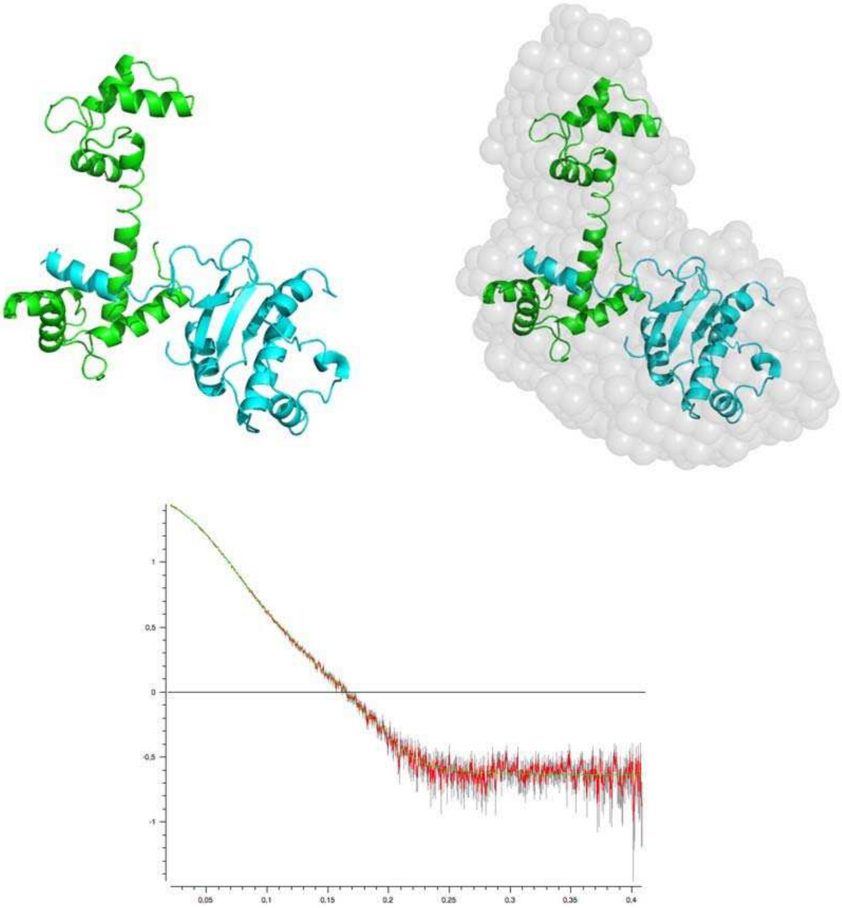


Figure 1- 39 Best outcome of the rigid body modelling (SASREF, CORAL) of the STAS/CaM complex using the extended CaM 3SJQ structure (upper left, CaM in green and STAS in cyan). The goodness of fit is shown in the lower panel, with $\chi^2 = 0.95$. Upper right panel, superposition of the STAS/CaM complex and the DAMMIN/DAMMIF *ab initio* model (grey surface).

This low resolution model, the first for CaM in complex with a STAS domain, shows that the STAS domain interacts with only one lobe of CaM and that the interaction is not limited to the helical part of the N-terminal end of the variable loop, but involves also the globular rigid core of the domain.

STAS and calmodulin – Surface Plasmon Resonance (SPR)

Having verified the existence of a close affinity between STAS and CaM, I also started doing Surface Plasmon Resonance experiments to obtain kinetic and thermodynamic parameters of the interaction, with a Biacore T100 using CM5 (GE Healthcare) chips and exploiting amine-coupling to covalently attach the STAS domain to the chip. I managed to immobilize the STAS domain reaching the target of 800 RU (Response Units), and proceeded with the experiments using calmodulin as analyte at concentrations ranging from 1 nM to 1 mM. I managed to obtain a signal in the sensorgrams, but with very low intensity, probably due to the fact that the immobilized protein might not be able to expose the interaction region to the analyte. Therefore, to be more confident with the estimation of the binding constant, it will be necessary to immobilize a higher quantity of the STAS domain but since the yield of the protein per liter of bacterial culture is quite low, several rounds of expression and purification are in progress, to obtain the necessary amount of protein. Another option would be to immobilize calmodulin to the chip, but the amino-coupling would be very tricky. In fact, calmodulin has a pI of 4 and the immobilization reaction must be carried out at a pH higher then 3, making the protein not completely charged, and therefore impairing the electrostatic interactions necessary for the approaching of the protein towards the surface of the chip. As alternative, immobilization of a Cys-labelled CaM via thiol coupling can be considered.

Conclusions

Prestin is the anion-dependent motor protein responsible for the outer hair cells electromotility, and its function depends on intracellular anions. In our group were collected diffraction data from crystals of the STAS domain of rat prestin (a mammal) in which the anions known to be necessary for the functionality of the whole protein were present, either from co-crystallization or from soaking experiments.

We therefore improved the working model explaining the function of prestin, suggesting that the first interaction site for the anions in indeed at the level of the intracellular STAS domain, and not in the TM domain as was previously thought. We also recently determined the crystal structure of the STAS domain from chicken prestin, the first one from a non-mammalian vertebrate, and in this case the binding site is absent, in accordance with the diffent function exerted by prestin in non mammals.

It is also known that the STAS domain interacts with calmodulin (CaM) with residues of the flexible variable loop, which is normally removed for crystallization purposes. I successfully expressed and purified both the STAS domain of rat prestin, including the variable loop, and calmodulin. I have characterized the STAS/CaM complex by SEC (Size Exclusion Chromatography), DLS, and SAXS obtaining the first low resolution model of the interaction.

PART TWO – protein kinases CDK2 and CK2

Protein kinases

Nearly 50 years ago, Krebs and Fisher found that glycogen phosphorylase was activated by a reversible addiction of a phosphate group by a protein kinase (phophorylase kinase) (Krebs, 1998). The second protein kinase to be discovered was the cAMP-dependent protein kinase (PKA) in the 1968 (Walsh et al., 1968). The importance of protein phosphorylation as a regulatory mechanism rapidly became clear with the advent of DNA cloning and sequencing in the mid-1970s, with the discovery of a large family of eukaryotic protein kinases. With the sequencing of the human genome almost all the human kinome was identified, being composed by 518 kinases, 478 human EPKs and 40 APK encoded by the 1.7% of the human genes (Manning et al., 2002). Emphasizing the importance of phosphorylation is the estimate that one third of cellular proteins are phosphorylated (Ahn and Resing, 2001) and often at different sites (Cohen, 2000). The major part of the protein kinases belong to the EPKs superfamily and can be classified into 9 broad groups, divided in families and subfamilies (Table 2-1).

Group	Family	Subfamily	Human kinases
AGC	14	21	63
CAMK	17	33	74
CK1	3	5	12
CMGC	8	24	61
STE	3	13	47
TK	30	30	90
TKL	7	13	43
RGC	1	1	5
Other	37	39	83
aPKs	14	22	40
Total	134	201	518

Table 2- 1 Human kinases distribution by broad groups

The classification is based primarily on kinase domain similarity, deduced from pairwise and multiple sequence alignments and phylogeny; knowledge on sequence similarity, domain structures outside the catalytic domains and known biological functions was used to refine the classification (Manning et al., 2002 b).

The kinases belonging to the **AGC** group (named after the PKA, PKC, PKG) tend to be basic amino acid-directed enzymes, phosphorylating substrates at Ser/Thr lying very near Arg or Lys. Many cytoplasmic serine/threonine kinases that are regulated by secondary messengers such as cyclic AMP (PKA) or lipids (PKC) belong to this group.

The kinases of the **CAMK** group also tend to be basic amino acid-directed; many but not all of the kinases members of this group are known to be activated by Ca^{2+}/Calmodulin binding to a small domain located near the catalytic domain (Calmodulin/Calcium regulated kinases).

The **CK1** group (Cell Kinase 1) is composed by kinases very similar to each other in sequence but very different from the structural point of view from the other ePKs.

The **CMGC** group (named after another set of families CDK, MAPK, GSK3 and CLK) is composed by Ser/Thr protein kinases which mainly phosphorylate residues lying near a Pro-rich environments and involved in cell-cycle control, stress response, splicing and metabolic control. Part of this group are the CDKs (Cycline dipendent kinases) and the CK2 family which fail to conform to the proline-directed specificity showing instead a strong preference for Ser residues located N-terminal to a cluster of acidic residues.

The **TK** group (Tyrosine Kinase) includes a large number of enzymes with quite closely related kinase domains that specifically phosphorylate on Tyr residues. This group is very young from the evolutionary point of view and it is absent from plants and unicellular organisms. Due to the fact that their function is related to the transmission of extracellular signals into the cell, half of TKs are cell receptors.

The **TKL** group (Tyrosine Kinase Like) is composed by kinases similar to members of the TK (Tyrosine Kinase) group, though they generally lack the TK-specific motifs of the TK group; their activity is generally on serine/threonine substrates.

Members of the **RGC** group (Receptor Guanylate Cyclases) are transmembrane receptors with an extracellular active guanylate cyclase domain and a cytoplasmatic catalytically inactive kinase domain. The guanylate cyclase domain makes the second messenger cGMP, and the intracellular kinase domain appears to have a regulatory function (Hanks and Hunter, 1995).

In figure 2-1 a graphic representation of the human kinome.

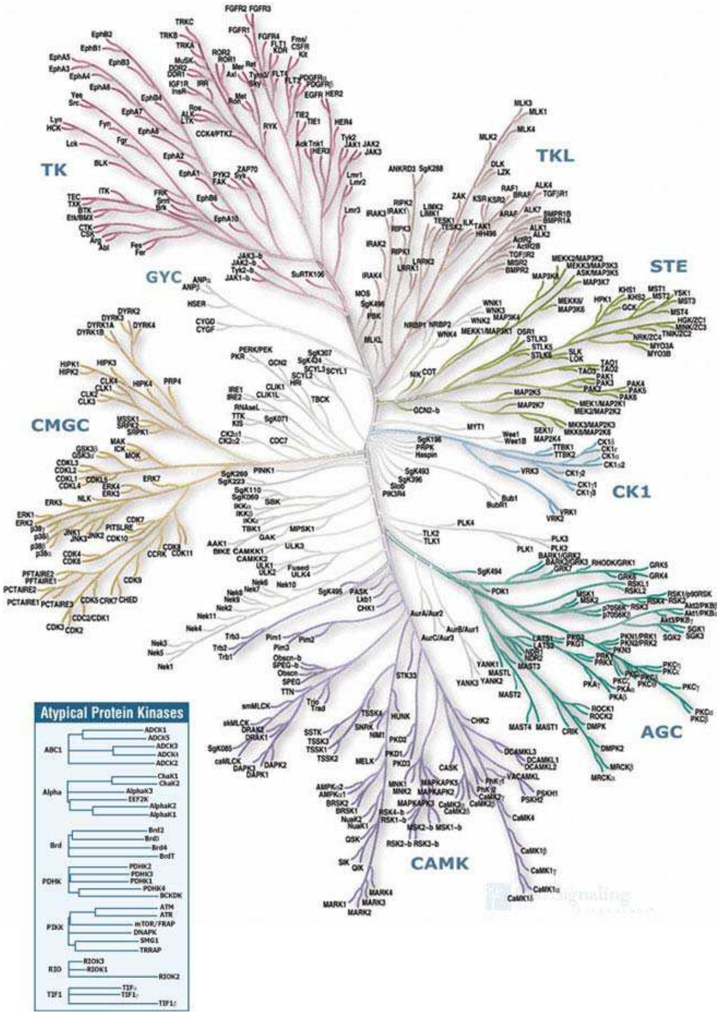


Figure 2- 1 the human kinome (Manning et al, 2002).

Structural features of kinases

A significant structural kinome is available, composed of over 150 protein kinases (Taylor and Kornev, 2011)

Focusing on the kinases of the EPKs superfamily, they are characterized by a conserved kinase core composed by a bi-lobal protein of approximately 250 amino acids. The catalytic cleft is positioned in the middle of the protein, where the two lobes convert to form a deep cleft where the adenine ring of the ATP is bound and the δ-phosphate points at the outer edge of the cleft. Catalysis is mediated by opening and closing of the active site cleft allowing for transfer of the phosphate and then release of the nucleotide (Taylor et al., 2012).

The N-terminal lobe (N-lobe) is composed of 5 anti-parallel beta sheets and one single conserved αC-helix (between β strand 3 and 4); β strand 1 and 2 are linked by a glycin-rich loop which stays on top of the adenine ring and blocks the δ-phosphate of the ATP on a correct conformation for the catalysis. β strand 3 interacts with the αC-helix by a coupling between a conserved Lys residue (Lys72 in PKA, Lys68 in CK2, Lys32 in CDK2) and a conserved Glu residue in the helix (Glu91 in PKA, Glu81 in CK2, Glu51 in CDK2) when the kinase is in an active state.

The hinge region connects the N-terminal lobe to the C-terminal lobe, mainly composed of alpha helixes with four-stranded beta sheet leaning on the top of the lobe. These beta sheets contain the other two conserved residues fundamental for the catalysis: the catalytic base of the catalytic loop (Asp164 PKA, Asp156 in CK2, Asp127 in CDK2) and the conserved motif Asp-Phe/Trp-Gly (DFG in PKA, DWG in CK2, DFG in CDK2) where a conserved Asp residue (Asp184 in PKA, Asp175 in CK2, Asp145 in CDK2) binds the catalytic magnesium ion (Figure 2-3).

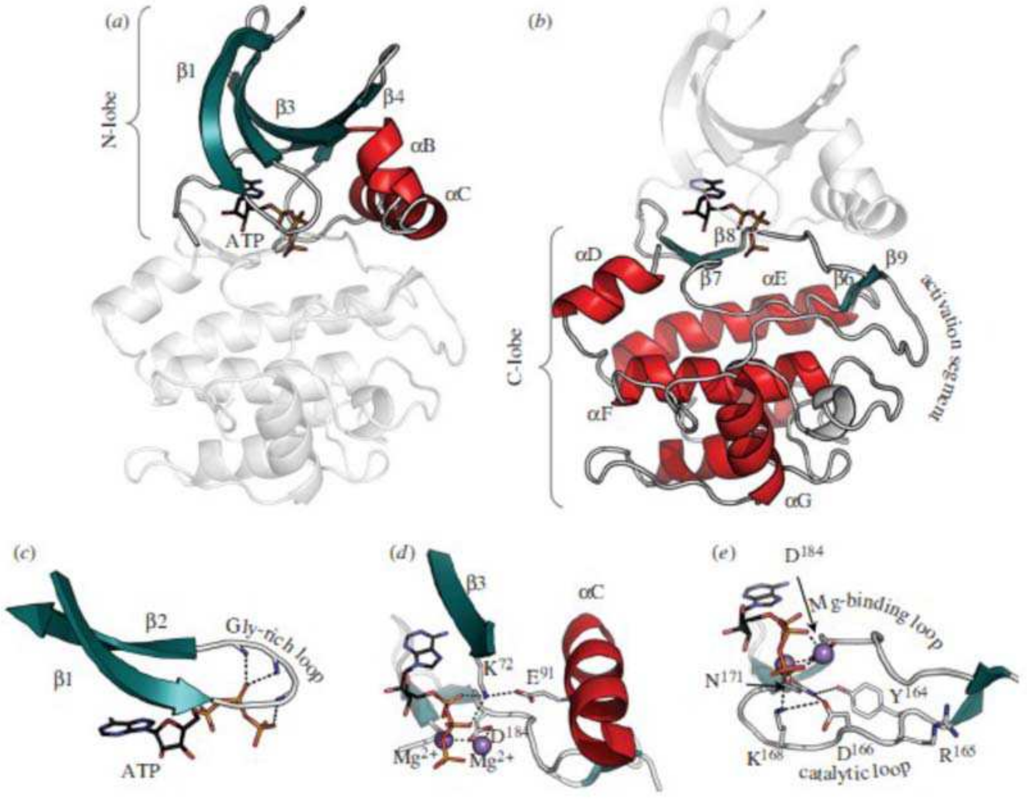


Figure 2- 2 Conserved core of the eukaryotic protein kinases. The bottom panels (c–e) highlight functional motifs in the N-terminal lobe (a) and the C-terminal lobe (b) using PKA as a prototype for the EPK family. Helices are shown in red; β-strands in teal. **(a)** The N- terminal lobe contains five β-strands and a large αC-helix. **(b)** The C-terminal lobe is mostly helical with a large activation segment. A four-stranded β-sheet rests on the helical core and forms one surface of the active site cleft. ATP is bound in the cleft between the two lobes. **(c)** The phosphates of ATP are positioned by a conserved glycine-rich loop between the β1-and β2-strands. **(d)** Conserved residues Lys72 from the β3-strand, Glu91 from the αC-helix, and Asp164 from the DFG motif in the activation segment where Mg^{2+} ions are shown as purple balls. **(e)** The catalytic loop also contains a set of catalytically important residues: Asp166, Lys168, Asn171 (Taylor et al.,2012).

Additional conserved structural elements can be identified from the hundreds of structures of EPKs representing various functional and binding states. Comparing many structures of protein kinases, it was possible to reveal that the conserved catalytic core is built around a stable yet dynamic hydrophobic core, made up of three essential elements: a single hydrophobic helix (αF-helix) and two hydrophobic spines which connect the N-teminal lobe and the C-terminal lobe. These two spines are made up of non-contiguous residues from both lobes (Taylor et al., 2012). The spine concept arises from a new approach of sequence comparison based on the novel "local spatial pattern" (LSP) alignment procedure (Kornev et al., 2006). The first application led to the detection, by the analysis of solvent accessible residues, of an hydrophobic spine composed of two residues of the N-lobe and two residues of the C-lobe (in PKA: Leu106, Leu95, Phe185 and Tyr164). This spine connects the two lobes in a non-covalent manner and, because it is broken in the inactive kinases, was later renamed "regulatory spine" (or "R-spine"). When the LSP approach was used to compare all residues of all the published kinases, another hydrophobic spine, that runs parallel to the R-spine, became visible (Kornev et al., 2008). Unlike the R-spine, this spine is always opened and only the adenine ring of the ATP completes it; this spine was therefore called "catalytic spine" (or "C-spine"). As shown in figure 2-3 both spines are anchored to the αF-helix.

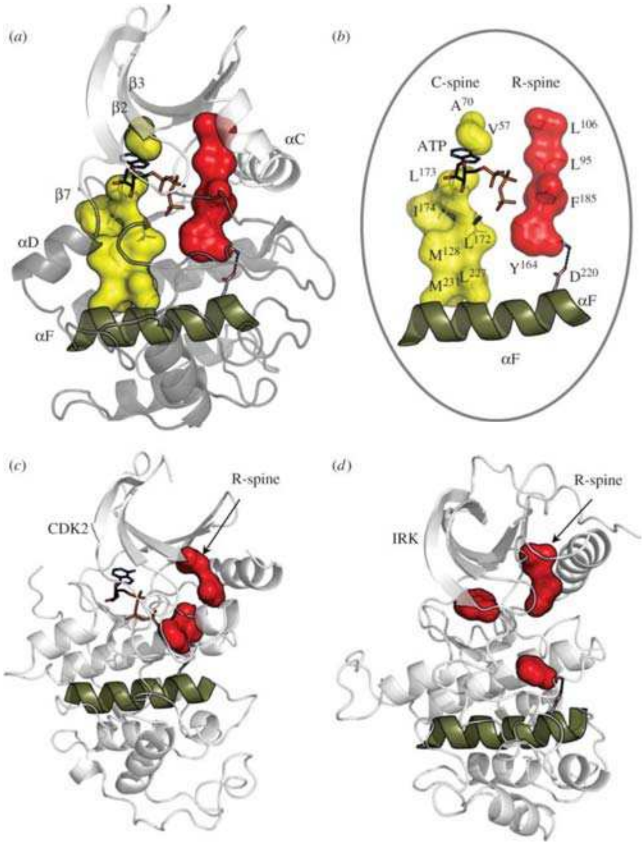


Figure 2- 3 Hydrophobic spines define the internal architecture of the EPKs. (a) Two hydrophobic spines span the two lobes of the kinase core and provide a firm but flexible connection between the N-terminal and C-terminal lobes. (b) The regulatory spine (R-spine) contains four residues from different kinase subdomains and is anchored to the αF-helix by conserved Asp220. ATP completes the catalytic spine (C-spine). (c,d) In the inactive state, the R-spine is typically disassembled. Disassembly of the R-spine can be achieved in different ways: by movement of the αC-helix like in cyclin-dependent kinase 2 (CDK2) (c) or by movement of the activation segment like in insulin receptor kinase (IRK) (d) (Taylor et al., 2012).

2.1 CDK2

Introduction

The cell cycle and CDKs

The cell cycle is an evolutionarily conserved process is from unicellular organisms, such as yeast, through to higher organisms, such as mammals.

Cell division occurs in a cycle characterized by four distinct phases, including two gap phases (G1 and G2) that prepare and assess the readiness of cells to enter the S-phase (DNA synthesis phase) or the M-phase (mitotic phase). A strict control of the S-phase is crucial to ensure that cells undergo a single round of chromosomal DNA replication, while the M-phase governs cytokinesis, where the duplicated genetic material is segregated into two identical daughter cells (Norbury and Nurse, 1992).

Cell division in eukaryotes is mediated by CDKs (cyclin-dependent kinases) as genetic and biochemical studies have demonstrated. The catalytic activity of CDKs is dependent on the association with a regulatory cyclin subunit. Cyclin–CDK dimeric kinase complexes, where the CDK is the catalytic kinase subunit and the cyclin is the activating subunit, tightly regulate progression of cells through the cell cycle.

Higher organisms possess different CDKs that associate with different cyclins and are important for progression through different cell cycle phases. It was demonstrated in mammalian cell culture studies that CDK4 and CDK6 in association with D-type cyclins are activated and important for progression during the G1 cell cycle phase (Matsushime et al, 1992; Meyerson and Harlow, 1994). During the transition from the G1- into the S phase the cyclin E–CDK2 complex is important (Endicott et al, 1999). Cyclin A associated with CDK2 is important during the S-phase and cyclin A associated with CDK1 is important the G2-phase progression (Pagano et al, 1992) The Cyclin B–CDK1 complex regulates the progression of cells through mitosis (Draetta and Beach, 1988) (Figure 2.1-1).

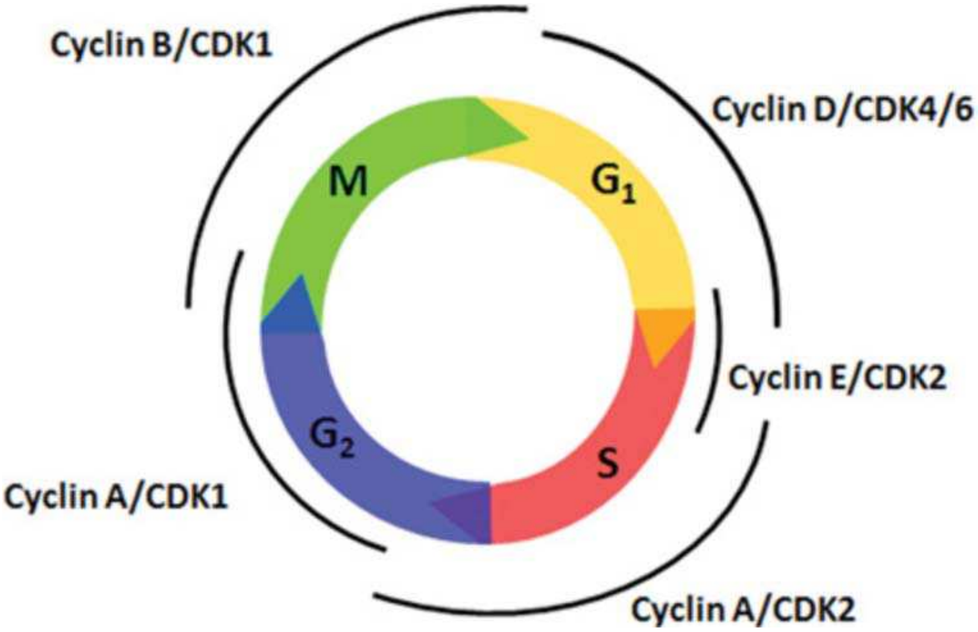


Figure 2.1- 1 The cell cycle consists of a DNA synthesis (S) phase and a mitotic (M) phase, separated by two gap (G1 and G2) phases. In mammalian cells, different cyclin–CDK complexes regulate progression of cells through the different phases of the cell cycle. (Suryadinata et al, 2010)

Given the importance of cyclin–CDKs complexes in the control of the cell cycle progression, these proteins are tightly regulated at several levels to ensure that the cell cycle progresses only under appropriate circumstances when cell division is required. Therefore the binding of inactivating proteins (cyclin-dependent kinase inhibitory proteins) and activating and inactivating phosphorylation events on critical residues of the CDK subunit, in addition to the binding of activating cyclins, regulate CDKs (Morgan, 1995).

The CDKs family

Cyclin-dependent kinases (CDKs) are Ser/Thr protein kinases that have been extensively studied due to their essential roles in the cell division cycle, transcription, differentiation, neuronal functions, as well as apoptosis (Gray et al, 1999). To date 13 CDK members of the CDK family (CDK1–CDK13) have been identified (De vivo et al, 2011). Some CDKs, such as CDK12 and CDK13, have only recently been characterized, while several other members of the family like CDK2 are well studied and plenty of structural and biochemical information is available. All the CDKs are similar in size (30–40 kDa) and share at least 40% sequence homology. It is well established that the CDK monomers are inactive in the quiescent cells, and are activated by specific cyclin partners at the proper point of time in the cell cycle (Lolli and Johnson, 2005). After the binding of cyclin subunits the complex gains basal kinase activity (Jeffrey et al, 1995). For example, cyclin A binding to CDK2 relieves the blocking of the catalytic cleft, and also exposes the phosphorylation site (Thr160) which sets the stage for the full activation of CDK2. Several other CDKs, such as CDK1, 4, 5 and 6, also need to be phosphorylated on a conserved threonine residue within the activation segment (also known as the T-loop). These activating processes can result in conformational changes in and around the catalytic domain of CDKs, which is indicative of their intrinsic conformational flexibility. This flexibility plays a central role in permitting CDKs to be regulated in several different ways, as well as in switching the states of CDKs in response to various regulatory signals involved in the growth of the eukaryotic cell (Pavletich, 1999).

CDK2

CDK2 plays a pivotal role in the promotion of the cell cycle progression. In synergy with its positive regulatory subunits of cyclins E and A, respectively, this evolutionarily conserved kinase promotes the transition of the well-established G1/S phase boundary and drives the cell cycle through S-phase. In detail, the CDK2/cyclin A complex is required for orderly S-phase progression, while the CDK2/cyclin E complex mediates the phosphorylation of retinoblastoma protein to facilitate the G1/S transition (Pines, 1991).

As a typical protein kinase, the monomeric CDK2 consists of 298 amino acids (PDB code: 1HCL, residues 1–298), and folds into a typical bilobal structure, with a smaller *N*-terminal lobe (residues 1–85) that contains an antiparallel five-strand β-sheet and a major C-helix, together with a larger *C*-terminal lobe predominantly composed of α-helix (De Bondt et al, 1993). The two terminal domains are connected through a single peptide strand (residues 81–83) which acts as a hinge linker to ensure that the two lobes can rotate with respect to each other without disruption of the secondary structure of this kinase (Cox et al, 1994). Figure 2.1-2 shows the structure of the monomeric CDK2.

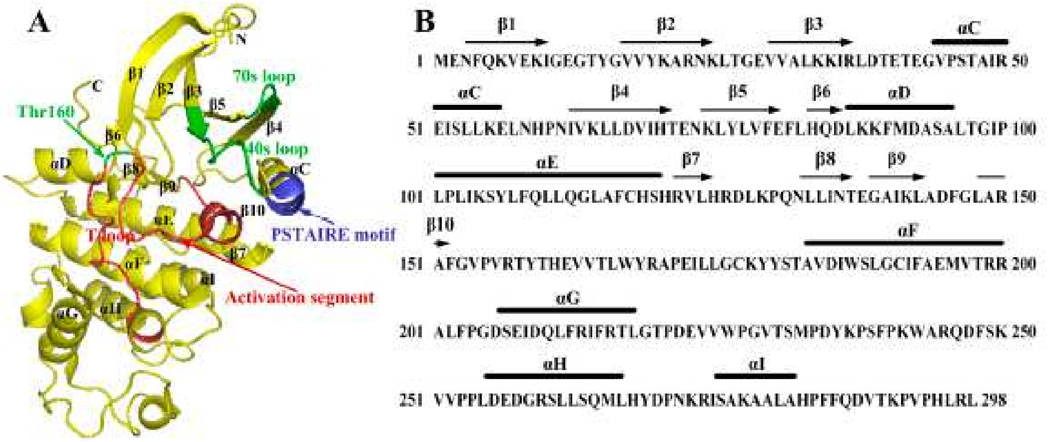


Figure 2.1- 2 The structure of monomeric CDK2. (A) Schematic diagram showing the structure of monomeric CDK2 using the coordinates from the PDB structure 4KD1. Secondary structural elements and the important fragments are labeled; (B) Sequences of CDK2 showing secondary structural elements (Li et al, 2015)

As shown, in the large *C*-terminal domain, there are two important segments: the catalytic residues that are in charge of phosphorylation promotion, and the activation segment. Among them, this activation segment spans residues between the conserved DFG (residues 145–147) and APE motifs (residues 170–172) and also includes the phosphorylation site, residue Thr160 (Brown et al, 1999). In the inactive CDK2 monomer (PDB code: 1HCK), the Thr160 among the activation segment conformation is buried away from solvent facing toward the conserved glycine-rich loop (Brown et al, 1999). The activation loop of monomeric unphosphorylated CDK2 (PDB code: 1WCC) can adopt a range of conformations which are distinct from the one of the inert CDK2 monomer (Nolen et al, 2004). The T-loop which is located at the entrance of the catalytic cleft, holds back the polypeptide substrates and prevents them from entering into the cleft to interact with ATP (Holmes and Solomon, 2001). The side chain of the catalytic site residue Glu51, which is ensconced in the PSTAIRE helix, lies toward outside the catalysis cleft (Davies et al, 2001). Also, the arginine residues (Arg50, Arg126 and Arg150) in the catalytic cleft are of great importance for the binding ability with the phosphate group of ATP (Barret and Noble, 2005). The unique sequence of the CDK family is the PSTAIRE motif (residues 45–51) embodied in the *N*-terminal C-helix, which has a key role in the interface.

CDK2 and cyclins

Three types of cyclins (A, B and E) have been crystallized successfully with CDK2. Cyclin binding at appropriate timing is one step for the activation of CDK2, and the other one is the phosphorylation by the CDK-activating kinases. Cyclins contact with one side of the catalytic cleft and interact with both the N-terminal and C-terminal lobes of the kinase as shown in 2.1-3. To accommodate cyclins well, CDK2 undergoes conformational changes, where the active site residues are restored to the correct positions for catalysis.

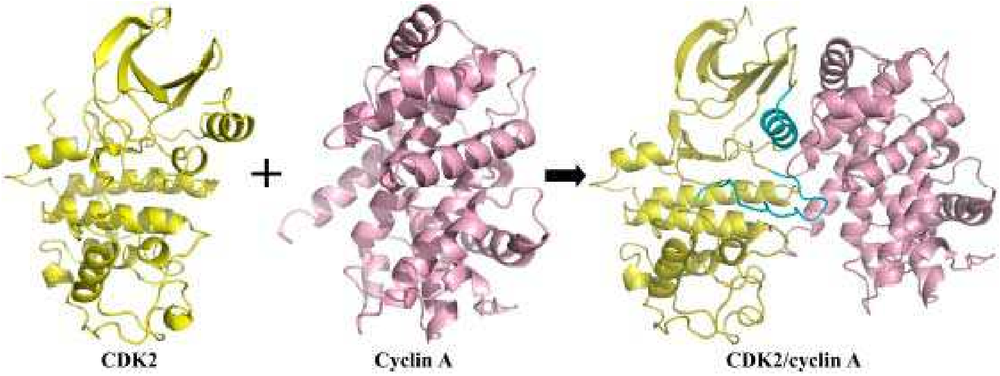


Figure 2.1- 3 Formation of the CDK2/cyclin heterodimer. The cyan fragments are regions undergoing structural changes upon association with the cyclin subunit. Coordinates used for CDK2 and CDK2/cyclin A are 1HLC and 1OL1, respectively (Li et al, 2015).

When cyclins binds to CDK2, the *N*-terminal domain, especially the αC-helix, undergoes some structural rearrangements which lead to a broadening of the active site, and then to the releasing of the catalytic cleft and the exposure of the phosphorylation site on the T-loop (Brown et al, 2007). The T-loop, which is the main obstacle for the substrate access in the monomeric apoenzyme, is displaced outside of the catalytic cleft.

CDK2 and phosphorilation

Another factor that causes conformational changes of CDK2 and completes the reorganization of the substrate binding site that is initiated by the cyclin binding is the phosphorylation on Thr160, carried out by a CDK-activating kinase (CAK) (Pines, 1999).

Upon binding of cyclins to CDK2, the T-loop, which includes the activating Thr160, is rearranged into a more active-like conformation where the activating Thr160 becomes accessible for CDK phosphorylation. The phosphorylation of Thr160 by CAK moves the T-loop further, enabling the phosphate group to interact with several amino acids, stabilizing the T-loop structure. Phosphorylation of the CDK/cyclin dimer with the phosphorylated Thr160 fully activates the kinase, and docks the T-loop onto the C lobe (Russo et al, 1996; Knighton et al, 1991).

Figure 2.1-4 shows the differences between free, partially activated and fully activated CDK2.

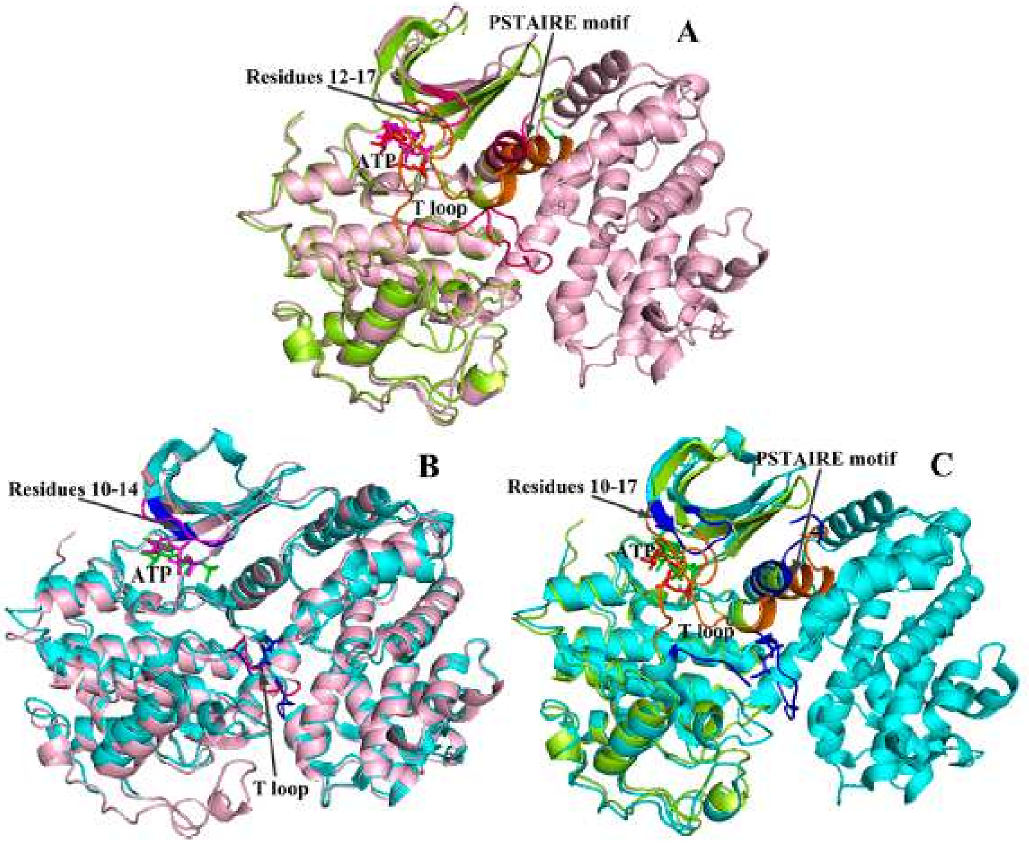


Figure 2.1- 4 Comparison of the CDK2 conformational changes in free, partially and fully activated states. (A) Structural superposition of free and partially activated CDK2 complexes; (B) The alignment of partially and fully activated CDK2 complexes; (C) Structural superposition of free and fully activated CDK2 complexes. The yellow, pink and cyan patterns indicate the monomeric CDK2 (free state, coordinates 1HCK), CDK2/cyclin A complex (partially activated form, 1FIN) and phosphorylated dimer (fully activated state, 4EOM), respectively. The changed conformations of monomeric CDK2, CDK2/cyclin A complex and phosphorylated dimer are colored in orange, magenta and blue, respectively (Li et al, 2015).

CDK2 in diseases and as a drug target

A deregulated activation of CDKs has been associated with various cancers, viral infections (Wang et al, 2003), Alzheimer (Tsai et al, 2004), Parkinson (Smith et al, 2003), renal diseases (Nelson et al, 2005) and ischemia (Wang et al, 2003). The role of CDKs in the pathogenesis of various diseases has encouraged an intensive search for potent and selective pharmacological inhibitors of CDKs.

Aberrant CDK2 activity may lead to the loss of cell control, hence inducing infinite proliferation of cell (Sherr, 1996), which is a common feature of most cancers. It has been found that CDK2 plays a role in cell proliferation of prostate cancer and non-small cell cancer (Kawana et al, 1998). If CDK2 activity is suppressed, an inhibition of the proliferation of human breast cancer cells can happen (Ali et al, 2009). Moreover, unchecked CDK2 activity constitutes a potent apoptotic signal (Lees et al, 1999). Given its critical role in cell cycle regulation, CDK2 has therefore been considered as a potential anticancer drug target (Saphiro et al, 2006). The number of CDK2-related crystal structures (including both CDK2 and CDK2/cyclin complexes with or without substrates) recorded in the Protein Data Bank database (PDB), is over 300. Plenty of CDK2 inhibitors have been studied, and they can be divided into ATP-competitive and non-ATP-competitive according to their target binding sites (Esposito et al, 2013). Among them, the ATP-competitive inhibitors are the most investigated, and some of them are evaluated by various stages of clinical trials (Malumbres et al, 2008). In general, an ATP-competitive CDK2 inhibitor binds to CDK2 via hydrogen bonding interactions (LEU 83, GLU 81), hydrophobic interactions (VAL18, ILE10, ASP146) and Pi-cation interactions (LYS33). However, due to the fact that the ATP binding site of the human kinases is well conserved, highly selective ATP-competitive inhibitors of CDK2 are quite difficult to obtain.

Cdk2, ANS and allosteric inhibitors

An allosteric binding site on CDK2 was discovered using ANS (fluorophore 8-anilino-1-naphthalene sulfonate) that binds in a site located in a region which is adjacent to the C-helix, away from the ATP site, and in fact located approximately halfway of the ATP site and the C-helix. As shown in figure 2.1-5 ANS quenches the interaction between CDK2 and cyclin, resulting in C-helix conformational fluctuations that are incompatible for cyclin A association (Betzi et al, 2011).

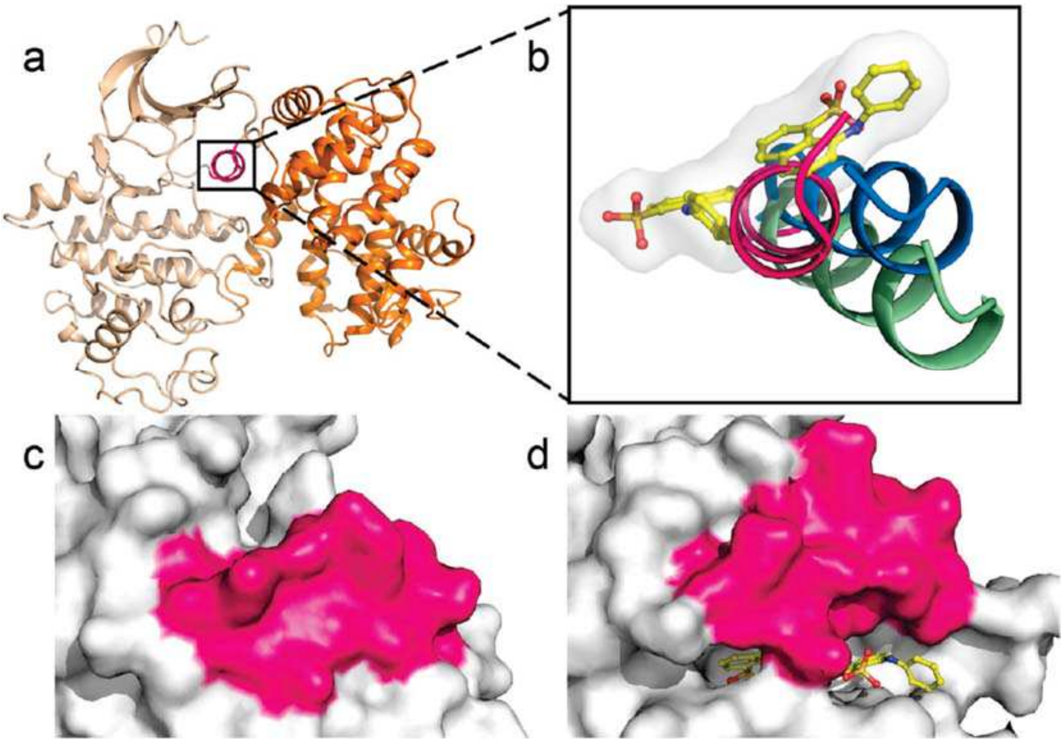
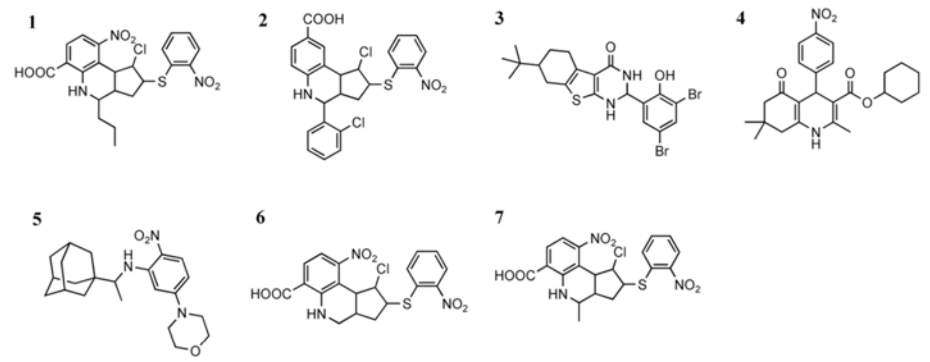


Figure 2.1- 5 Binding of ANS induces large conformational changes in the C-helix. (a) Overall view of the CDK2-cyclin A complex structure (PDB ID 2CCI). CDK2 is shown in wheat, the C-helix in magenta, and cyclin A in orange. (b) Superimposition reveals the distinct conformation of the C-helix in the CDK2-ANS complex (green) and in free CDK2 (blue) as compared to the CDK2-cyclin A complex (magenta). Surface representations of the C-helix and surrounding residues for free CDK2 (c) and the CDK2-ANS complex (d) reveal the partial opening of the ANS pocket toward solvent. (Betzi et al, 2011)

Soon after the discovery of the allosteric binding pocket, other molecules were screened out to hit this binding site with a fluorescence-based high-throughput screening assay that detects compounds able to displace ANS from the allosteric site (Martin et al, 2012). However the affinity of ANS for CDK2 (*K*d = 37 μM) (Betzi et al, 2011) is remarkably lower than that of cyclin A (*K*d = 0.6 μM) (Martin et al, 2012), therefore ANS is displaced from CDK2 upon the interaction with cyclin. In order for the tight interaction of the CDK–cyclin interface to be broken, the allosteric inhibitors need to possess extraordinarily high inhibitory affinities (Rastelli, 2014). In this context, our collaborators in the group of prof. Rastelli (University of Modena and Reggio Emilia) screened by docking procedures in silico a large library of compounds. Compounds were initially screened for their ability to bind CDK2 and displace ANS from the allosteric pocket. Selected hit compounds are shown in figure 2.1-6.



Hit expansion of compound 4

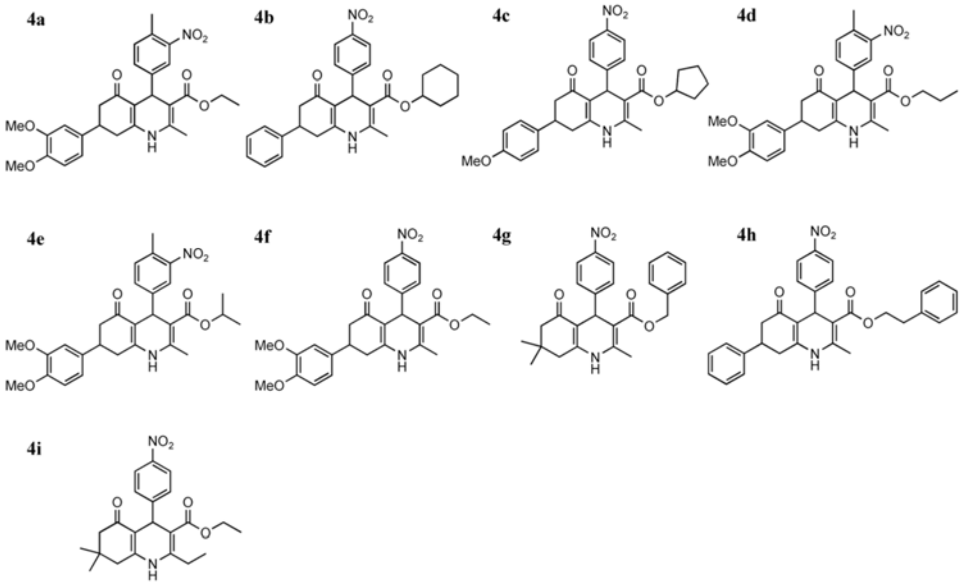


Figure 2.1- 6 potetial allosteric inhibitos of Cdk2 (Rastelli et al, 2014)

Allosteric inhibitors are in general more selective then competitive inhibitors, since the ATP binding site is common to all kinases. However, allosteric pockets near the αC helix are somewhat recurrent in protein kinases, hence allosteric modulation mechanisms could be more conserved then previously thought (Palmieri et al., 2013).

Aim of the work

Members of the Cyclin dependent kinase (CDK) family are key regulatory proteins that can become deranged in certain types of cancer. The CDK activity requires the binding of regulatory proteins known as cyclins and the CDK/cyclin complexes are mainly responsible for the transition through the cell cycle by directly promoting the cell cycle progression, regulating transcription, and exerting post-mitotic functions. CDK2 is one of the CDKs involved in the regulation of the S and G2 phases of the cell cycle and its activity is frequently altered in cancer, thus representing an interesting anti-cancer target. In a recently identified inactive conformation of CDK2 in complex with two molecules of the 8-anilino-1-naphthalene sulfonate fluorescent probe an open allosteric pocket has been detected and proposed as a site to accommodate type III protein kinase allosteric inhibitors. Allosteric modulators interfere with protein function by binding to a site distinct from the catalytic site and shifting the relative populations of the protein conformational ensemble. They are more selective than type-I inhibitors that target the conserved ATP-binding site.

The knowledge of the interactions between inhibitors and deregulated enzymes is of fundamental importance for the rational design of specific drugs targeting these proteins involved in different types of cancer. Protein kinase inhibitors are grouped into three main categories: type I ATP-competitive inhibitors bind to the active state of the enzyme in the pocket in which normally ATP (or GTP) binds, type II ATP-competitive inhibitors bind in the same place but to the inactive form of the kinase, and type III non ATP-competitive inhibitors, or allosteric inhibitors, bind to a different site of the macromolecule, causing a conformational rearrangement that makes the kinase inactive, acting via an allosteric mechanism. Type III inhibitors could be more specific than the others since they do not bind to a common region found in all kinases (the ATP-binding pocket) but to different regions of the molecule. For CDK2, we are studying both ATP-competitive inhibitors and allosteric inhibitors.

In this context, the aim of my work was to establish robust expression, purification and crystallization protocol to solve crystal structures of CDK2 in complex with potential inhibitors.

Experimental procedures

Recombinant protein expression

The coding sequence of human CDK2 was cloned in a pGEX6P1 expression vector. The construct was used to transform *E. coli* BL21(DE3) cells. Bacterial clones were then conserved at -80 °C. Cells were grown in liquid medium LB (Luria-Bertani; Sambrook and Russel, 2001): 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl. The medium was sterilized by autoclaving it for 20 minutes at 121 °C and implemented with 100 µg/ml ampicillin, to which bacterial cells transformed with the pGEX6P1 vector are resistant. *E. coli* cells were grown over-night (ON) in LB medium containing 100 µg/ml ampicillin at 37°C in a suitable shaker. LB medium supplied with 100 µg/ml ampicillin was then inoculated with this over-night culture in the ratio 1:10. The culture was kept at 37°C in a suitable shaker and the growth was followed measuring the OD_{600} (Optical Dispersion at 600 nm, directly related to the amount of cells in suspension). The temperature was slowly decrease until 16°C when an OD_{600} of 0.6 was reached, then protein expression was induced by adding IPTG to a final concentration of 0.1 mM when an OD_{600} of 0.9 was reached. After induction, bacteria were growth for 20-24h at 16 °C. Cells were then harvested by centrifugation (20 min, 4000 g, 4°C). The harvested cells were resuspended in buffer A: 50 mM Hepes, 150 mM NaCl, 10 mM MgCl2, 2 mM dithiothreitol (DTT), 1 mM ethylene glycol tetra-acetic acid (EGTA), pH 7.5 supplemented with EDTA-free protease inhibitors and lysed with a "French Press" cell disrupter at 1000 psi. Cell debris were then removed from the soluble components extracted by centrifugation of the lysate (45 min, 12000 g, 4°C). After this centrifugation step, the supernatant was filtered using 0.22 µm filters before being loaded onto the column for the first step of the purification procedure.

Purification protocol

An AKTApurifier 10 with a multiple-wavelength detector was used for the chromatographic steps of the purification procedures. The chromatographic runs were followed at 280 nm and chromatograms were recorded using the UNICORN software (GE healthcare).

The filtered supernatant obtained after the mechanical lysis of the bacterial cells was loaded onto a 1 ml GST-TRAP 4B (GE healthcare) affinity column, exploiting the presence of the GST-tag on the protein. The column was equilibrated with buffer A and the sample was loaded. The fusion protein was eluted with a step gradient procedure with buffer B: 50 mM Hepes, 150 mM NaCl, 10 mM MgCl2, 2 mM dithiothreitol (DTT), 1 mM ethylene glycol tetra-acetic acid (EGTA), 10 mM GSH(reduced), pH 7.5 using a flow rate of 1 ml/min.

To revome the GST-tag, a proteolytic cleavage using the protease 3C (also recombinantly produced and purified exploiting a GST-tag) was performed over-night at 4°C. After the cut the glutathione was removed with a desalting chromatographic step exploiting a HiPrep 26/10 desalting coloumn.

In order to separate GST and CDK2 a second GST-affinity run was performed. Then CDK2 fractions were further purified by Size Exclusion Chromatography using HiLoad 16/60 preparation grade Superdex 75 column(GE healthcare) equilibrated with 50 mM HEPES buffer (pH 7.5) containing 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 2 mM EGTA, and 0.01 mM ADP. CDK2 was concentrated to 10 mg/ml and flash-frozen in liquid nitrogen for storage at -80 °C.

SDS-PAGE

The purification procedure was followed by SDS-PAGE. Samples were heated for 5 minutes at 95 °C to denature them in the sample-loading buffer: 60 mM Tris pH 6.8, 10% glycerol, 2% SDS, 0.01% bromophenol blue, 2% β-mercaptoethanol. Then they were loaded with Low molecular weight See Blue® Plus2 Pre-Stained Standard (*Invitrogen)* markers on SDS-PAGE gels. A voltage of 180 V was used to performed the electrophoretic run with the gels mounted vertically on the electrophoretic apparatus and in contact with the running buffer: 25 mM Tris pH 8.3, 250 mM glycin, 0.1% SDS. The gels were colored for 20 minutes under soft shaking with the staining solution: 40% EtOH, 10% AcOH, 0.1% w/v Coomassie G-250 and then the destaining solution: 40% EtOH, 10% AcOH and a final wash in water were used to remove unspecific staining.

Protein crystallization

In order to perform crystallization trials CDK2 buffer was exchanged to 100 mM Na^+/K^+ phosphate pH 6.2, 2 mM DTT. Crystallization trials were performed in two different crystallization conditions: 50 mM Hepes pH 7.5, 5-35% PEG 3350 and 50 mM Hepes pH 7.5 and 5-35% Jeffamine ED2001 (Hampton Research). The first crystallization screens were performed using this wide range of precipitant concentrations, and crystals were obtained in all of the drops set up.

Single Crystal X-ray Crystallography

Crystals were cryoprotected with 20% ethylene glycol, mounted on loops, and frozen in liquid nitrogen. Diffraction data were collected at synchrotron sources (ID30B and ID29 at ESRF, Grenoble) at 100 K.

Diffraction data were initially processed (indexing, integration and scaling) with the XDS package (Kabsch, 2010). Integrated reflections were then reduced and merged with AIMLESS (Evans, 2006) (Evans, An introduction to data reduction: space group determination, scaling and intensity statistics, 2011) from the CCP4 suite (Winn, et al., 2011). The structures were solved by molecular replacement using the previously deposited model with the PDB ID 3PXR (or 3PXQ) as search model and the program PHASER (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The initial structures were then refined alternating manual steps of real space refinement using the program COOT (Emsley & Cowtan, 2004) (Emsley, Lohkamp, Scott, & Cowtan, 2010) and automatic steps of reciprocal space refinement using the suite PHENIX (Adams, et al., 2010). PYMOL was used to make all the images and the superposition analysis.

Results and Discussion

CDK2 WT and mutants for biological tests

Our group is also involved in a medicinal chemistry project involving protein kinase CDK2, in collaboration with Prof. Rastelli (Università degli Studi di Modena e Reggio Emilia), Dr. Broggini (Istituto di Ricerche Farmacologiche Mario Negri), Prof. Passarella (Università degli Studi di Milano), and Dr. Storici (Sincrotrone ELETTRA), within an AIRC project. The goal of the project is to develop and characterize allosteric inhibitors for oncogenic kinases like CDK2.

In the contest of this AIRC project I have produced the WT protein and the mutants of CDK2 used for the functional tests performed by our collaborators.

The expression was carried out as described in the experimental procedures. To produce the WT and mutated proteins for the biological tests, bacteria were suspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 1 mM DTT (buffer A) supplemented with the cOmplete™, Mini Protease Inhibitor Cocktail (Roche), and lysed using a French pressure cell press. The lysate was centrifuged for 30 minutes at 12000 rpm at 4 °C and the supernatant was applied onto a GST-affinity column (GE Healthcare) equilibrated with the buffer A. GST-CDK2 was eluted using the buffer A supplemented with 10 mM reduced GSH. The final protein was diluted at 1 mg/mL, aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C. Figure 2.1-7 shows the chromatogram of the GST-affinity run of one of the mutants.

Figure 2.1- 7 GST affinity chromatography. In blue the elution profile (Abs at 280 nm), in green the step gradient used to elute the GST-CDK2 fusion construct.

The chemical synthesis, the structure elucidation, and the obtainment of pure stereoisomers of the previously identified hits for allosteric inhibitors of CDK2 were performed. Moreover, modeling studies unveiled the details of the interaction with CDK2. Then, the biological activity of the obtained compounds was evaluated with different biological assays. Mutagenesis experiments confirmed binding to the CDK2 allosteric pocket. Finally, the allosteric ligands were shown to inhibit the growth of A549 non-small cell lung and SKOV3 ovarian cancer cell lines. All of these results ended up in an article titled “Probing an allosteric pocket of CDK2 with small-molecules” (Christodoulou et al, 2017)

CDK2 - construct 1 purification

The first CDK2 construct that I have recombinanlty expressed and purified, was already available in our group. The GST-tag was exploited for the purification procedure, which consisted of 4 steps: 1) GST-Affinity Chromatography; 2) proteolytic cleavage of the GST-tag with the specific 3C protease; 3) removal of the GST-tag from the solution by a second GST-Affinity Chromatography ; 4) SEC (Size Exclusion Cromatography). Figure 2.1-8 shows the chromatogram of the first step of the purification procedure.

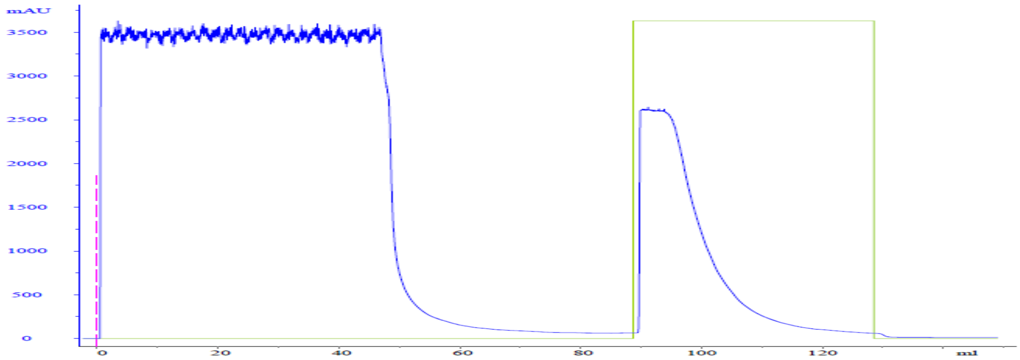


Figure 2.1- 8 GST-CDK2 GST-affinity chromatogram; in blue the elution profile (Abs at 280 nm), in green the step gradient of glutathione.

The eluted protein was incubated ON at 4 °C with a recombinantly produced GST-3C protease. After the proteolytic cleavage, the gluthatione was removed by a desalting step using a HiPREP 26/10 desalting column (GE Healthcare). This allowed to perform a second GST- affinity chromatography, to separate GST and CDK2. Figure 2.1-9 shows the chromatogram of this second purification step, where CDK2 does not interact with the stationary phase, having lost the GST-tag, while the uncleaved residues, the GST-tag, and the GST-3C protease are eluted afterwards using glutathione.

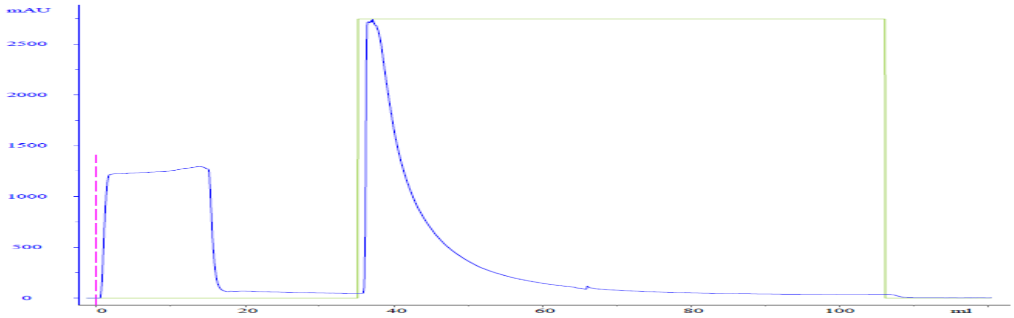


Figure 2.1- 9 Second GST-affinity chromatogram, after proteolytic removal of the GST-tag with 3C protease, in blue the elution profile (Abs at 280 nm), in green the step gradient of glutathione.

The eluate was further purified by size exclusion chromatography, using a HiLoad 16/60 preparation grade Superdex 75 column (GE Healthcare), equilibrated with the 50 mM HEPES buffer (pH 7.5) containing 150 mM NaCl, 10 mM $MgCl_2$, 2 mM DTT, 2 mM EGTA, and 0.01 mM ADP. Figure 2.1-10, shows the chromatograms for this final purification step.

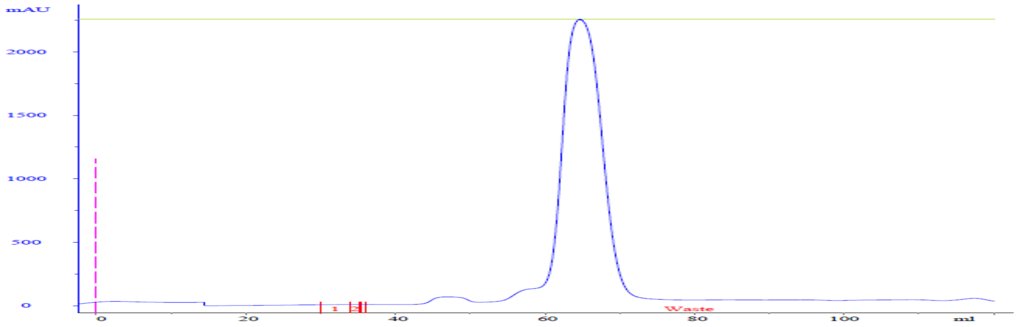


Figure 2.1- 10 CDK2 Size exclusion chromatography chromatogram.

The proceeding of the purification procedure was followed by SDS-PAGE and the final protein was pure as shown in figure 2.1-11.

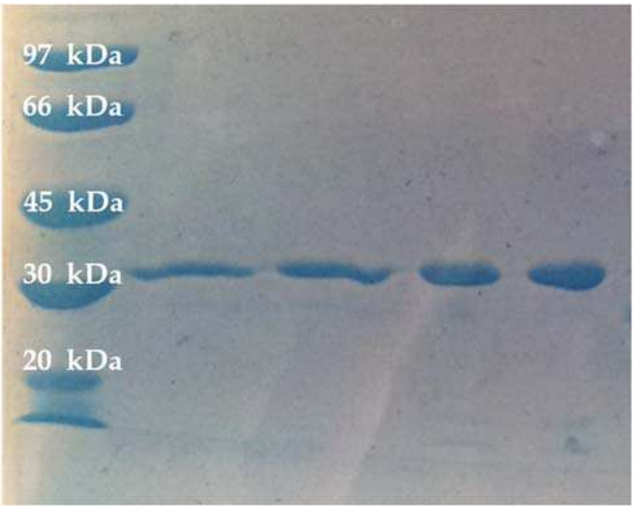


Figure 2.1- 11 Coomassie stained SDS-page of the fractions of the final size exclusion chromatography.

Several expressions and purifications of this first CDK2 construct were performed, optimizing their protocols and increasing the yields of protein per liter of bacterial culture.

CDK2 construct 1 – crystallization trials

In order to perform crystallization trials CDK2 buffer was exchanged to 100 mM Na^+/K^+ phosphate pH 6.2, 2 mM DTT. Crystallization trials were initially performed in two different crystallization conditions derived from the PDB entries 3PXQ and 3PXR: 50 mM Hepes pH 7.5, 5-35% PEG 3350 and 50 mM Hepes pH 7.5, 5-35% Jeffamine ED2001 (Hampton Research).

Several parameters were changed, including the pH of the phosphate buffer, protein concentration, precipitant composition, temperature, drop size and protein : precipitant ratio in the drop. In the end crystallization screens were also tested.

Despite having performed thousands of crystallization trials, no crystals were obtained.

CDK2 construct 2 - purification

With the first CDK2 construct, I managed to establish a reliable and reproducible purification protocol for CDK2, but despite having performed several crystallization trials, I was not able to obtain crystals of CDK2. I therefore decided to change the construct of the protein choosing one with a different N-terminal tail, which seemed to have better crystallization properties, from a search made throughout the structures deposited on the protein data bank. The N-terminal tail is left after the cleavage with the 3C protease. The old construct had a "GPGS" tail, the new one has a longer tail with the sequence "GPLGSPEF".

```
> CDK2 construct 1:

GPGSMENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIRE
ISLLKELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIKS
YLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEGAIKLADFGLARAFGVPVRTYTH
EVVTLWYRAPEILLGCKYYSTAVDIWSLGCIFAEMVTRRALFPGDSEIDQLFRIF
RTLGTPDEVVWPGVTSMPDYKPSFPKWARQDFSKVVPPLDEDGRSLLSQMLHYDP
NKRISAKAALAHPFFQDVTKPVPHLRL
```

```
> CDK2 construct 2:

GPLGSPEFMENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPST
AIREISLLKELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLP
LIKSYLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEGAIKLADFGLARAFGVPVR
TYTHEVVTLWYRAPEILLGCKYYSTAVDIWSLGCIFAEMVTRRALFPGDSEIDQL
FRIFRTLGTPDEVVWPGVTSMPDYKPSFPKWARQDFSKVVPPLDEDGRSLLSQML
HYDPNKRISAKAALAHPFFQDVTKPVPHLRL
```

After cloning, bacteria were transformed with the new plasmid. Bacterial growth, expression and purification were carried out as described for the previous construct. Figure 2.1-12 and 2.1-13 show the chromatograms of the first two steps of the purification procedure.

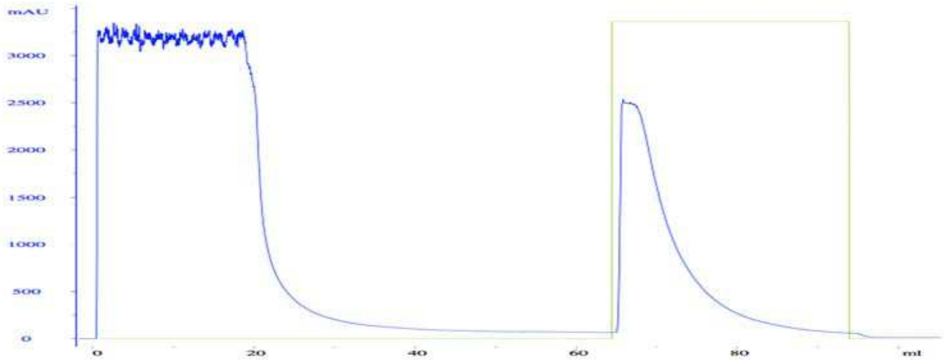


Figure 2.1- 12 2 GST affinity chromatography. In blue the elution profile (Abs at 280 nm), in green the step gradient used to elute the GST-CDK2 fusion construct.

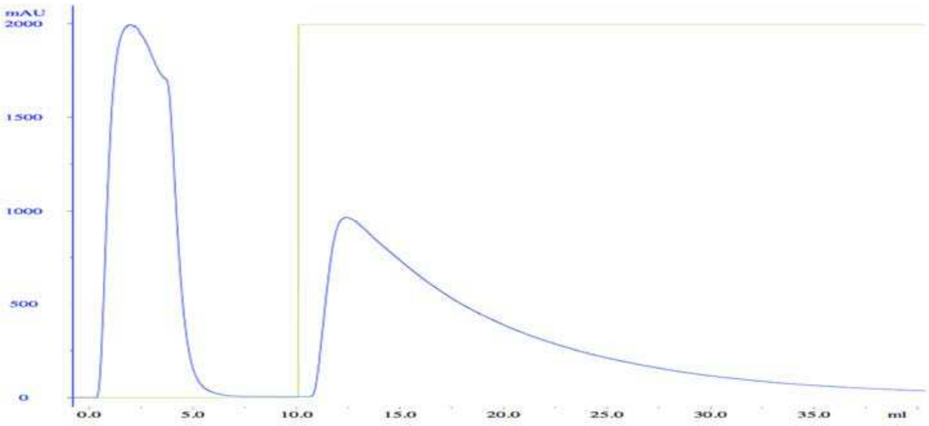


Figure 2.1- 13 GST affinity chromatography. In blue the elution profile (Abs at 280 nm), in green the step gradient used to elute GST and residues of the uncut GST-CDK2 fusion construct.

The purification procedure was followed by SDS-PAGE. As can be seen in figure 2.1-14 the fusion GST-CDK2 construct was recovered by elution with glutathion in the first GST-affinity chromatography (lane: el.GSH1), then it was successfully cut by 3C protease (lane: postcut), buffer exchanged and CDK2 and GST were succesfully separated by the second GST-affinity chromatography (lanes: FT2 and el.GSH2).

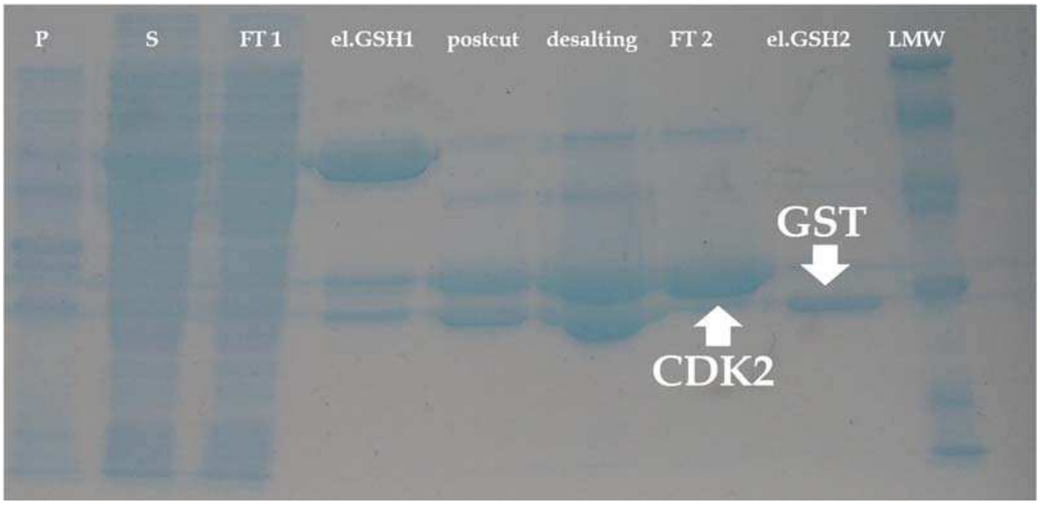


Figure 2.1- 14 Coomassie stained SDS-page of the fractions of the purification procedure for CDK2 construct 2 LMW: markers of known MW; P: lysate pellet (insoluble fraction); S: lysate supernatant (soluble fraction); FT1: “flow through” of the first GST-affinity chromatography; el.GSH1: glutathione elution of the first affinity chromatography; postcut: proteins after the proteolytic cleavage; desalting: proteins after the desalting run; FT2: “flow through” of the second GST-affinity chromatography; el.GSH2: glutathione elution of the second affinity chromatography.

The last step of the purification procedure was a size exclusion chromatography (figure 2.1-15).

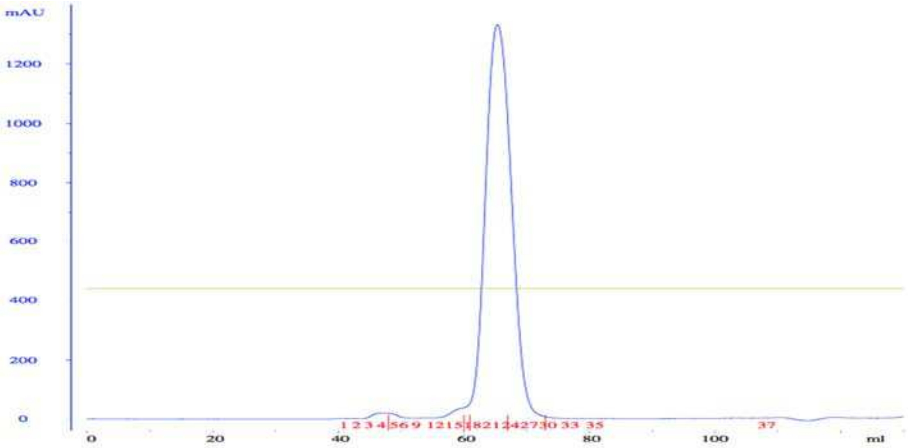


Figure 2.1- 15 Size Exclusion Chromatography. In blue the elution profile (Abs at 280 nm).

The fractions of the last run were also analyzed by SDS-PAGE, showing only one band corresponding to the purified CDK2 (figure 2.1-16) as happened for the first construct.

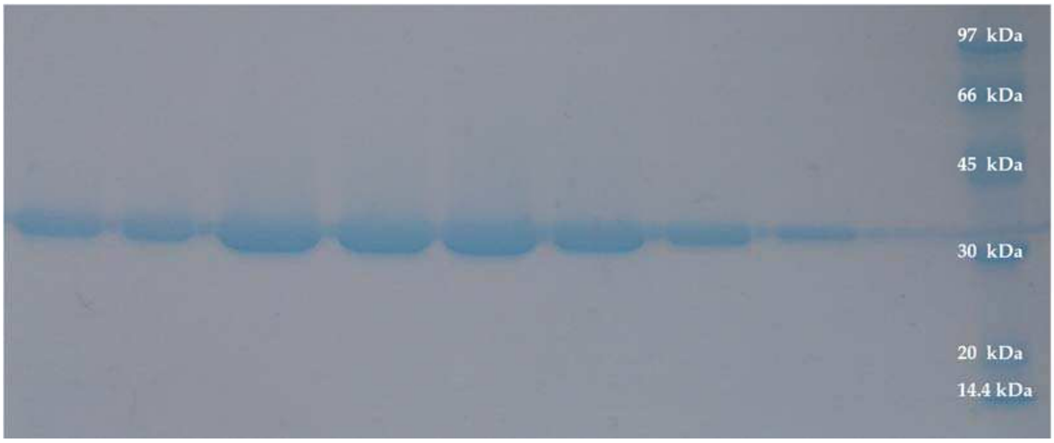


Figure 2.1- 16 Coomassie stained SDS-page of the fractions of the final size exclusion chromatography.

CDK2 – contruct 2 crystallization trials

In order to perform crystallization trials CDK2 buffer was exchanged to 100 mM Na$^+$/K$^+$ phosphate pH 6.2, 2 mM DTT. Crystallization trials were performed in two different crystallization conditions: 50 mM Hepes pH 7.5, 5-35% PEG 3350 and 50 mM Hepes pH 7.5, 5-35% Jeffamine ED2001 (Hampton Research). With the new construct, I was able to obtain large well-diffracting crystals (figure 2.1-17).

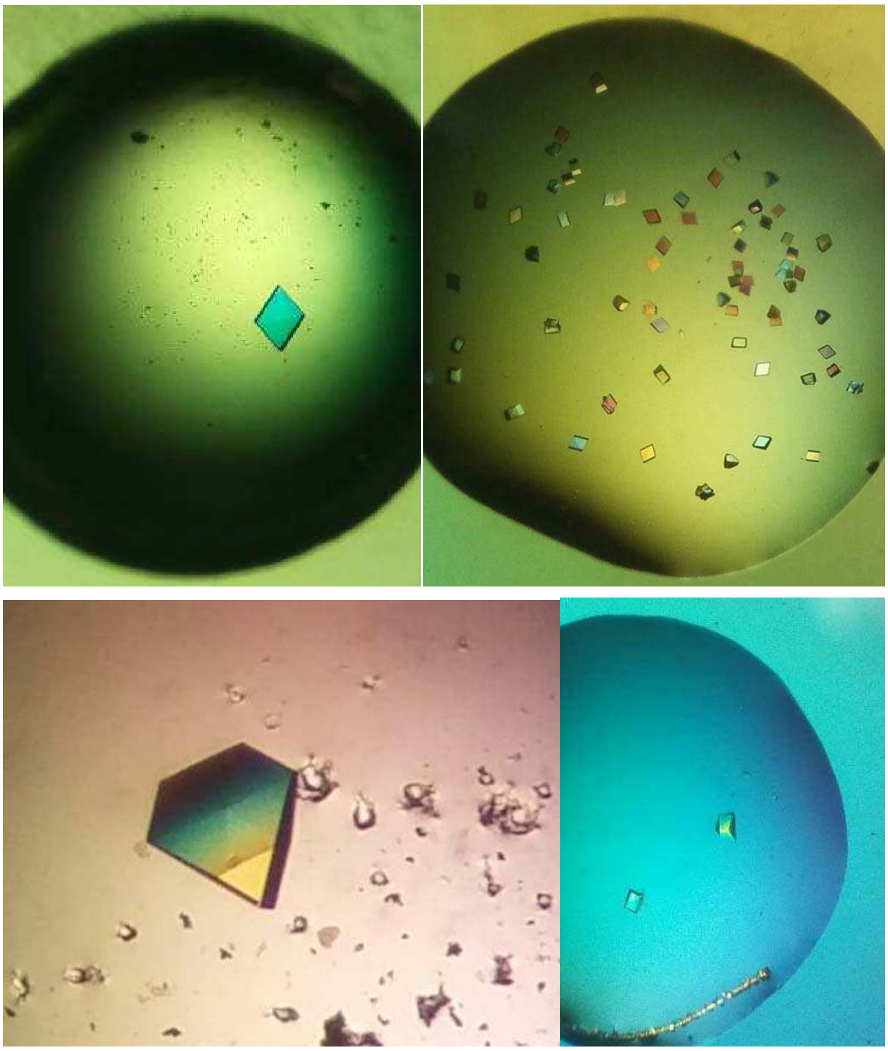


Figure 2.1- 17 CDK2 crsytals

CDK2 – construct 2: Single Crystal X-ray Crystallography

CDK2 and ATP-Competitive inhibitors

As regards the ATP-competitive inhibitors, I performed several crystallization trials, both in co-crystallization and soaking experiments. Crystals were cryoprotected with 20% ethylene glycol, mounted on loops, and frozen in liquid nitrogen. Diffraction data were collected at synchrotron sources (ID30B and ID29 at ESRF, Grenoble) at 100 K. Figure 2.1-18 shows a CDK2 crystal mounted on a loop.

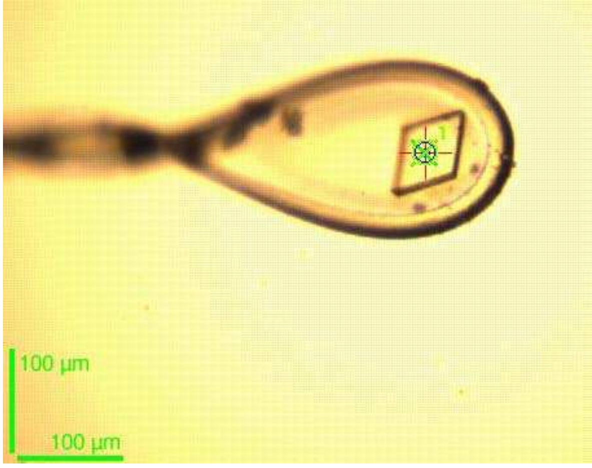


Figure 2.1- 18 A CDK2 crystal mounted on a loop

Diffraction data were collected and the structures were solved by molecular replacement. I was able to solve the structures of four complexes between CDK2 and competitive inhibitors: CX4945, CX5279, K137 and CX5011. Table 2.1-1 shows the data collections and refinements statistics for these four datasets.

	CX4945	**CX5279**	**CX5011**	**K137**
Data collection				
Wavelenght (Å)	0,96862	0,96862	1,072	0,96862
Space group	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)
Cell dimensions				
a b c (Å)	53,8 72,3 72,7	53,6 72 72,4	53,69 72,08 72,42	53,8 72 72,7
α β γ (°)	90 90 90	90 90 90	90 90 90	90 90 90
Resolution (Å)	43,28-1,27 (1,31-1,27)	43,11-1,42 (1,47-1,42)	43,13-1,25 (1,27-1,25)	43,3-1,3 (1,35-1,3)
R_{sym}	0,078 (9,579)	0,095 (3,324)	0,083 (3,8)	0,075 (2,02)
R_{meas}	0,085 (10,4)	0,103 (3,66)	0,089 (4,19)	0,082 (2,20)
$CC_{(1/2)}$	1,000 (0,475)	0,998 (0,562)	0,999 (0,430)	0,999 (0,655)
< I/σI >	14,1 (0,3)	10,9 (0,6)	11,5 (0,5)	14,2 (1,0)
Completeness (%)	99,9 (99,4)	99,9 (99,7)	97,1 (84,3)	99,7 (99,2)
Multiplicity	12,4 (12,4)	12,6 (11,0)	13,1 (10,9)	12,5 (12,3)
Refinement				
Resolution (Å)	43,28-1,27	43,1-1,42	43,13-1,25	43,29-1,3
Number of reflections	73314	53531	75751	69975
R_{work}/R_{free}	21,01/22,22	21,55/24,49	20,55/22,33	20,98/22,13
B-factors (Å^2)				
protein	32	35,11	31,41	33,11
ligands	21,47	42,62	35,38	59,61
r.m.s. deviations				
bond lenght (Å)	0,004	0,009	0,012	0,006
bond angles (°)	0,793	1,196	1,42	0,948
Ramachandran plot				
favoured (%)	98,56	97,23	96,3	97,41
outliers (%)	0	0,79	0	0,74

Table 2.1- 1 Data collection and refinement statistics

As shown in figure 2.1-19 all the inhibitors bind in the same position, the ATP-binding site.

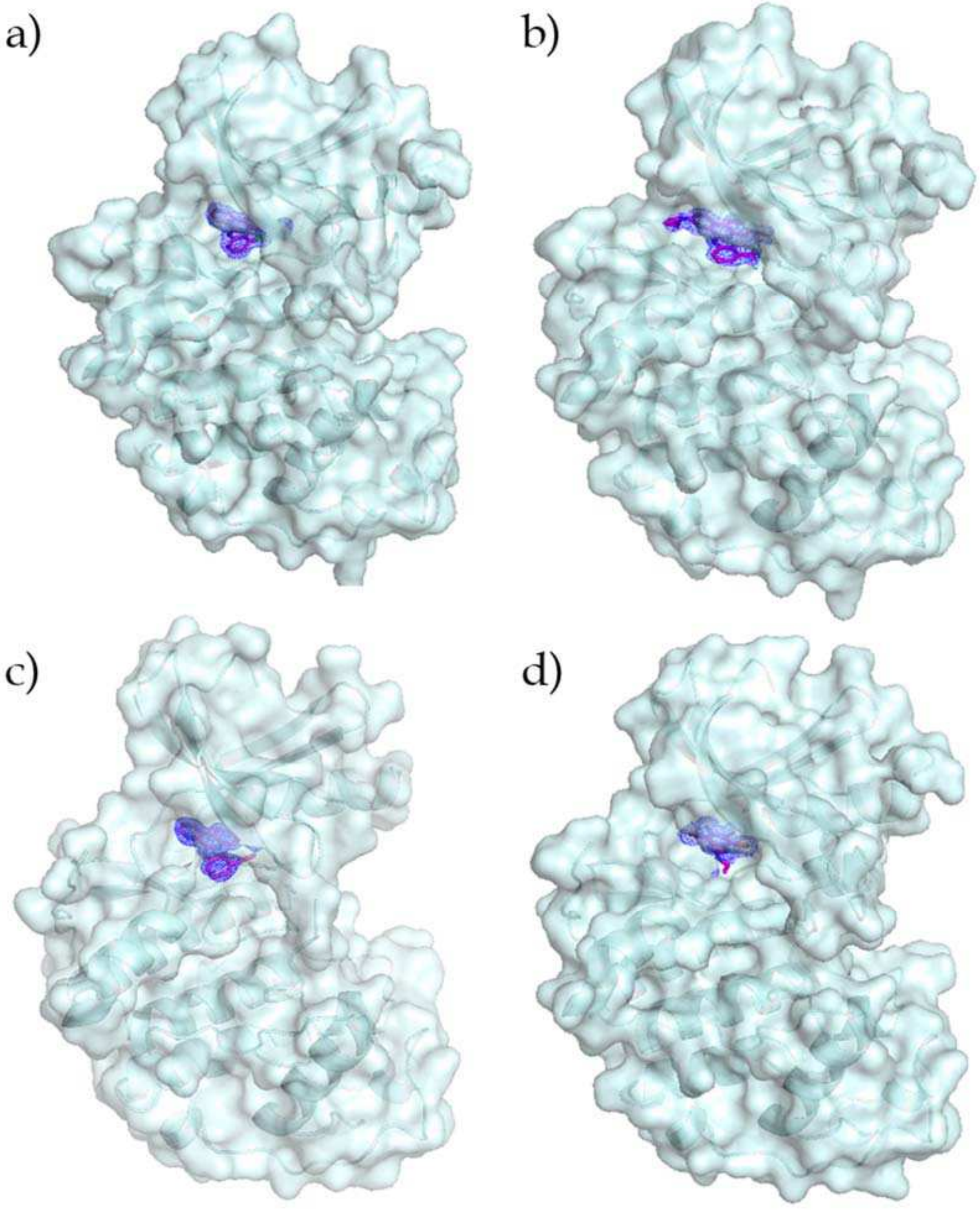


Figure 2.1- 19 CDK2 in complex with the ATP-competitive inhibitors a) CX4945, b) CX5279 c) K137 and d) CX5011

Figure 2.1-20 shows the $2F_O$-F_C maps for the three inhibitors contoured at 1 σ.

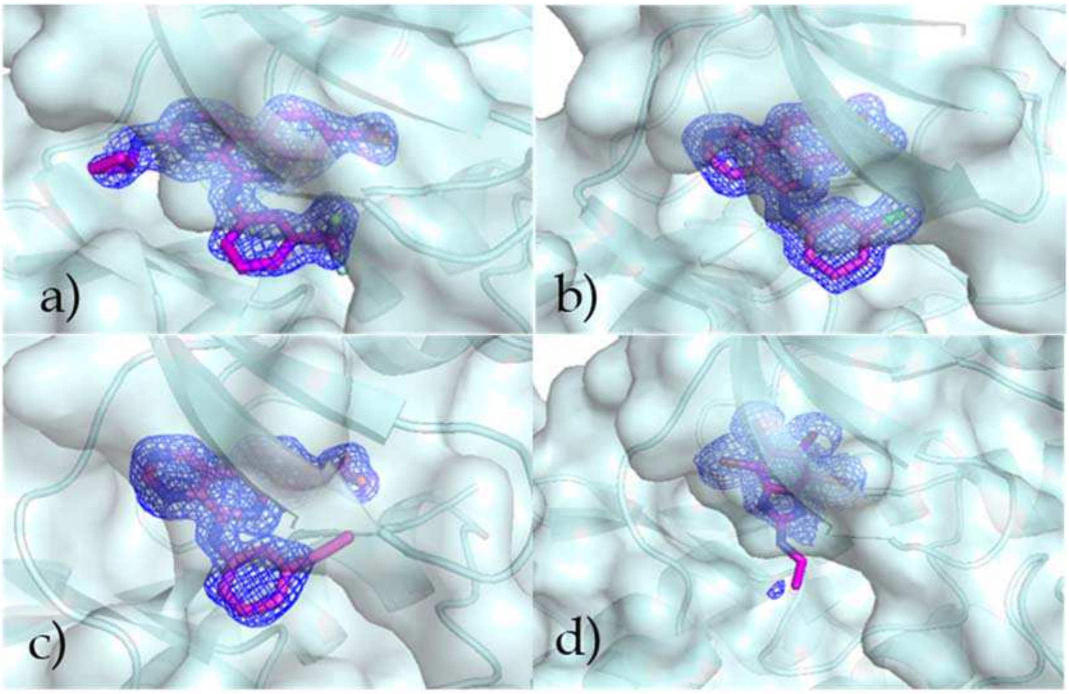


Figure 2.1- 20 1 $2F_O$-F_C maps contoured at 1 σ of a) CX4945, b) CX5279 c) K137 and d) CX5011

As regards the interactions, the three CX inhibitors establish hydrophobic interactions in the hydrophobic ATP-binding pocket, and hydrogen bonds with the side chains and main chains of the surrounding amino acids (figure 2.1-21).

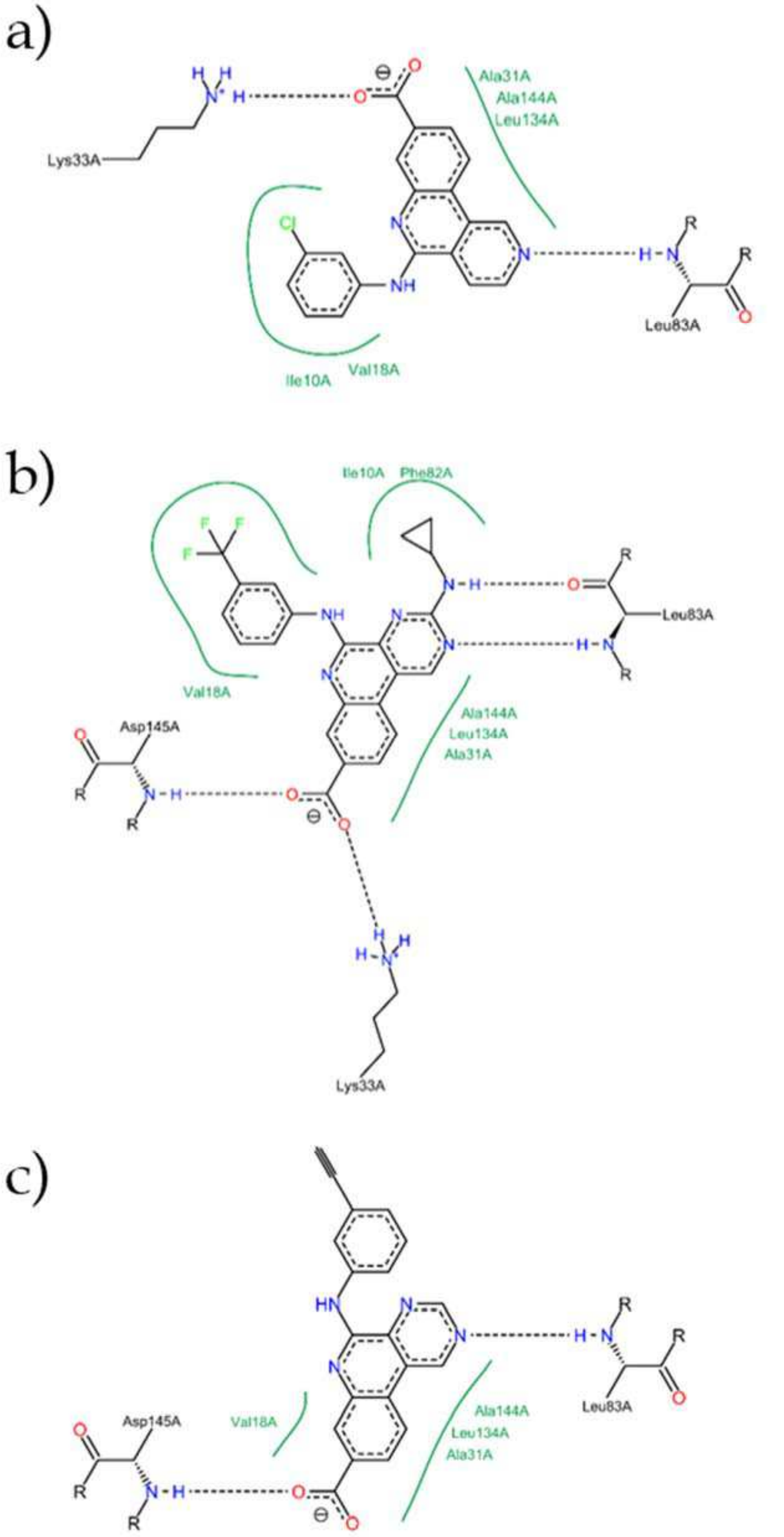


Figure 2.1- 21 Black dashed lines indicate hydrogen bonds, green solid line show hydrophobic interactions of a) CX4945, b) CX5279 c) CX5011

As regards K137, two of the bromine atoms make halogen bonds with the COs of I83 and E81 as shown in figure 2.1-22. A very weak anomalous signal was exploited to correctly orient the bromine atoms, and therefore the whole ligand.

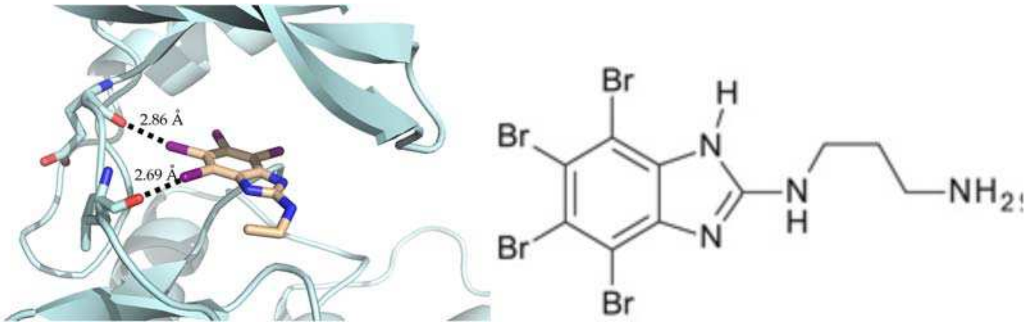


Figure 2.1- 22 On the left, Halogen bonds established by two of the four bromide atoms (purple) of inhibitor K137 are shown as dashed lines. On the right, the structure of K137.

All of these molecules are known inhibitors of CK2α, and the relative structures were previously determined in this group. From the comparison of the binding modes of the same ligand to the two kinases important clues regarding their selectivity characteristics can be derived. Activity was tested on both kinases (Battistutta et al, 2011;Cozza et al, 2015) but the active complex between CDK2 and cyclin A had to be used, therefore testing the inhibitors as type I ATP-competitive. Table 2.1-2 summarizes the residual activity of the two kinases in presence of the various inhibitors. The structures solved are instead on the inactive form of the kinase. Depending on their affinity, these inhibitors could act as type II ATP-competitive, avoiding the binding of cyclins and hence the activation of CDK2.

	CX4945 0,5 µM	**CX5279 0,5 µM**	**CX5011 0,5 µM**	**K137 1 µM**
CK2α	0	0	0	0
CDK2/Cyclin A	40	90	63	41

Table 2.1- 2 residual activity on kinases in presence of the various competitive inhibitors.

Further, having one molecule targeting two different kinases is interesting for treating diseases in which both proteins are involved, using a single drug.

CDK2 and allosteric inhibitors

As regards the project involving the allosteric inhibitors, in order to be sure that the purification and crystallization protocols of CDK2 were suitable to obtain crystals of the protein with the allosteric pocket in the open conformation, I have started by performing crystallization trials in complex with ANS (8-anilino-1-naphthalene sulfonate) fluorescent probe, the molecule that led to the discovery of the pocket. As shown in figure 2.1-23 I was able to obtain crystals of CDK2 in co-crystallization with ANS.

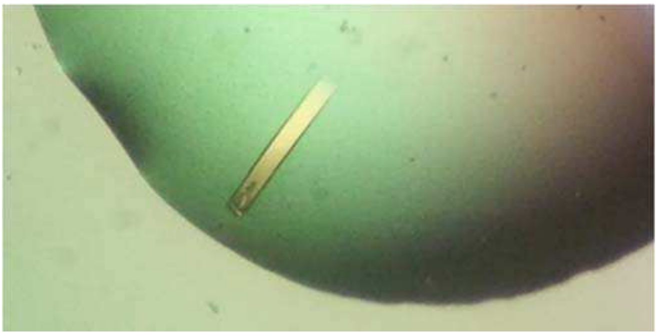


Figure 2.1- 23 CDK2-ANS crystal

Diffraction data were collected and I solved the structure of CDK2 in complex with three molecules of ANS, one in the ATP-binding site, and two in the allosteric pocket as shown in figure 2.1-24.

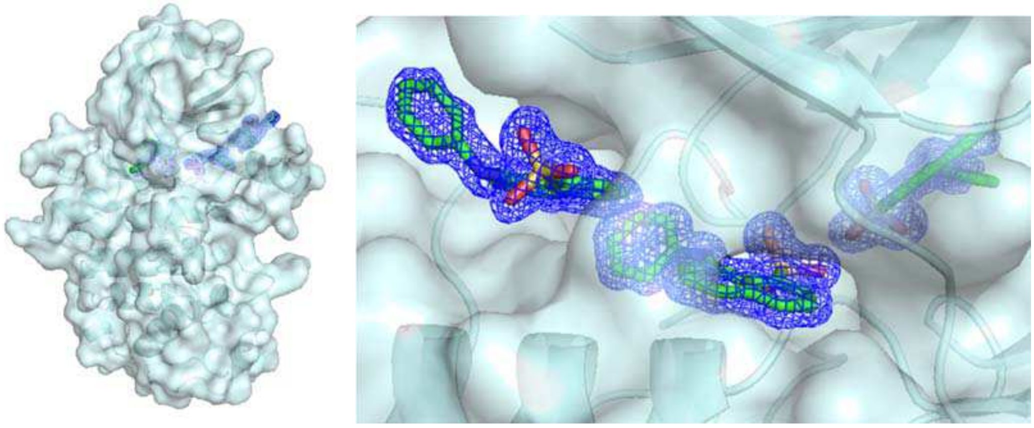
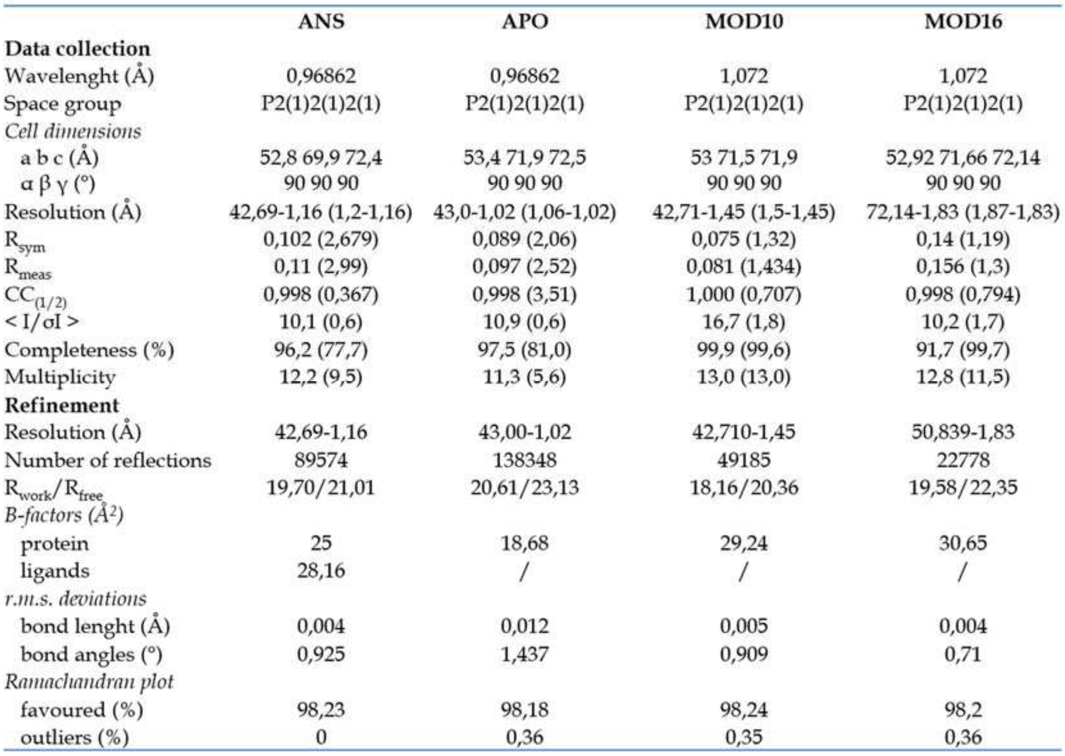


Figure 2.1- 24 On the left, CDK2 in complex with three molecules of ANS, on the right the $2F_O$-F_C map contoured at 1 σ

As regards the allosteric inhibitors, I have performed several crystallization trials both in co-crystallization and soaking experiments with 6 different inhibitors and solved more then 30 structures of CDK2 but so far I have never found electron density that could be attributed to the inhibitors probably due to the low affinity and/or solubility of the molecules.

I also tried to displace ANS from ANS containing crystals, with the allosteric inhibitors, first reducing the ANS concentration by dilution of the drop containing the crystals, and then moving the crystal in a drop containing the allosteric inhibitors. I collected data of the APO form, the ANS form, and the washed crystals with the two most promising allosteric inhibitors (MOD10 and MOD16). Data collection and refinement statistics are shown in table 2.1-3. The cell of the washed crystals has the same dimensions of the APO form, in particular the length of b, that changes in the ANS containing crystals.



	ANS	**APO**	**MOD10**	**MOD16**
Data collection				
Wavelenght (Å)	0,96862	0,96862	1,072	1,072
Space group	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)
Cell dimensions				
a b c (Å)	52,8 69,9 72,4	53,4 71,9 72,5	53 71,5 71,9	52,92 71,66 72,14
α β γ (°)	90 90 90	90 90 90	90 90 90	90 90 90
Resolution (Å)	42,69-1,16 (1,2-1,16)	43,0-1,02 (1,06-1,02)	42,71-1,45 (1,5-1,45)	72,14-1,83 (1,87-1,83)
R_{sym}	0,102 (2,679)	0,089 (2,06)	0,075 (1,32)	0,14 (1,19)
R_{meas}	0,11 (2,99)	0,097 (2,52)	0,081 (1,434)	0,156 (1,3)
$CC_{(1/2)}$	0,998 (0,367)	0,998 (3,51)	1,000 (0,707)	0,998 (0,794)
$< I/\sigma I >$	10,1 (0,6)	10,9 (0,6)	16,7 (1,8)	10,2 (1,7)
Completeness (%)	96,2 (77,7)	97,5 (81,0)	99,9 (99,6)	91,7 (99,7)
Multiplicity	12,2 (9,5)	11,3 (5,6)	13,0 (13,0)	12,8 (11,5)
Refinement				
Resolution (Å)	42,69-1,16	43,00-1,02	42,710-1,45	50,839-1,83
Number of reflections	89574	138348	49185	22778
R_{work}/R_{free}	19,70/21,01	20,61/23,13	18,16/20,36	19,58/22,35
B-factors (Å²)				
protein	25	18,68	29,24	30,65
ligands	28,16	/	/	/
r.m.s. deviations				
bond lenght (Å)	0,004	0,012	0,005	0,004
bond angles (°)	0,925	1,437	0,909	0,71
Ramachandran plot				
favoured (%)	98,23	98,18	98,24	98,2
outliers (%)	0	0,36	0,35	0,36

Table 2.1- 3 data collection and refinement statistics

Indeed, the structures solved from the two washed crystals, reverts to the apo form, as shown in figure 2.1-25.

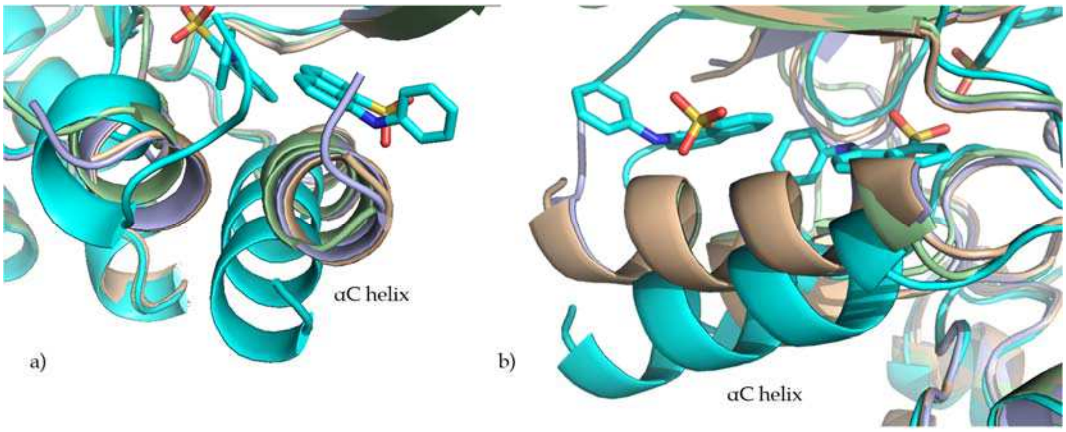


Figure 2.1- 25 superposition of the apo form (wheat), ANS bound (cyan), MOD10 washed (green) and MOD16 washed (purple) structures. The washed structures, and in particular the αC helix, revert to the apo form conformation, incompatible with the binding of molecules in the allosteric pocket.

Conclusions

CDK2 plays a pivotal role in the promotion of the cell cycle progression and being involved in several pathologies, among which cancers, is an interesting anti-cancer drug target.

Two constructs of CDK2 were expressed and purified. I managed to establish a reliable and reproducible purification protocol for the first construct of CDK2, but despite having performed several crystallization trials, I was not able to obtain crystals of CDK2. I therefore decided to change the construct of the protein choosing one with a different N-terminal tail, which seemed to have better crystallization properties, from a search made throughout the structures deposited on the protein data bank. With the new construct, I was able to obtain large well-diffracting crystals.

Several crystallization trials were set up, using both ATP-competitive and allosteric inhibitors, and several structures were solved. I managed to solve the structure of CDK2 in complex with four ATP-competitive inhibitors.

As regards the allosteric inhibitors, I was able to reproduce the structure of CDK2 in complex with ANS, that was used to discover the allosteric pocket, and several datasets were collected on crystals obtained in either co-crystallization or soaking experiments. However, I never found electron density that could be attributed to the inhibitors probably due to the low affinity and/or solubility of the molecules.

2.2 CK2

Introduction

CK2

Originally discovered in 1954 (Burnett and Kennedy, 1954), CK2 is a family of enzymes that in humans consists of two catalytic subunits, termed CK2α and CK2α′, and one regulatory subunit, CK2β (St-Denis and Litchfield, 2009). CK2 is a Ser/Thr protein kinase, highly pleiotropic and conserved and it is involved in many cellular processes such as, cell cycle progression, gene expression, embryogenesis, cell growth and differentiation, circadian rhythms and apoptosis (Pinna, 2013).

CK2 is a member of the superfamily of eukaryotic protein kinases (EPKs); however, it shows some unique features like its high constitutive activity and the lack of an acute mode of regulation (Pinna, 2002). The regulation of protein phosphorylation is fundamental for cellular signaling pathways, as loss of regulation in these pathways underlies many human diseases, including cancer (Hanahan and Weinberg, 2000). Indeed, deregulation of the CK2 catalytic activity has been linked to several pathologies, mainly cancers, but also neurodegenerative disorders. The oncogenic potential of CK2 is mainly due to its ability to act as an anti-apoptotic agent (Ruzzene and Pinna, 2010) and has been linked to abnormally high levels of activity found in a large variety of tumors (Guerra and Issinger, 2008). CK2 is a validated drug target, with two inhibitors in ongoing clinical trials as antitumor agents.

Structural studies on CK2

The CK2 catalytic subunit, CK2α, has the canonical bilobal shape of an EPK, with the N-lobe composed of β-sheets and only one conserved α helix (the αC helix), and the C-lobe predominantly composed by α helixes.

An important region in the C-lobe is the “activation segment”, formed by the magnesium binding loop (DFG/DWG motif), the “activation loop” and the P+1 loop. In many EPKs the activation loop can assume two different conformations, one fully active and the other inactive, by large rearrangement of the structure, usually due to a phosphorylation of one or more residues of the loop. In CK2, no phosphorylation site is present on the loop and no conformational change was observed within the several structures of CK2α published (Niefind et al., 2009). The activation loop of CK2α is always in a conformation that is very similar to the fully active state of the nearest kinases in the EPKs (CDKs and MAP kinases), due to intense contacts between the N-terminal domain and the activation loop. In CK2, the αC-helix, the activation segment and the N-terminal domain are therefore rigid, and lack the plasticity that characterizes other EPKs. The active state of the activation segment is stabilized by an internal constraint in addition to the contact with the N-terminal region, an hydrogen bond across the β8–β9 joining loop between Trp176-NE1 and Leu173-O (Niefind et al., 1998). Interestingly CK2α and CK2α' are the only kinases of the EPKs superfamily with a Trp instead of a Phe in the DFG/DWG triplet and an analysis on the active state of the EPKs found that this phenylalanine is part of the R-spine, that is completed and fully active only with the phenylalanine in the DFG-in conformation. In the CK2α (or α') this conformation is the only possible, due to structural contacts of Trp176, strengthening the constitutively active character of the protein (Figure 2.2-1).

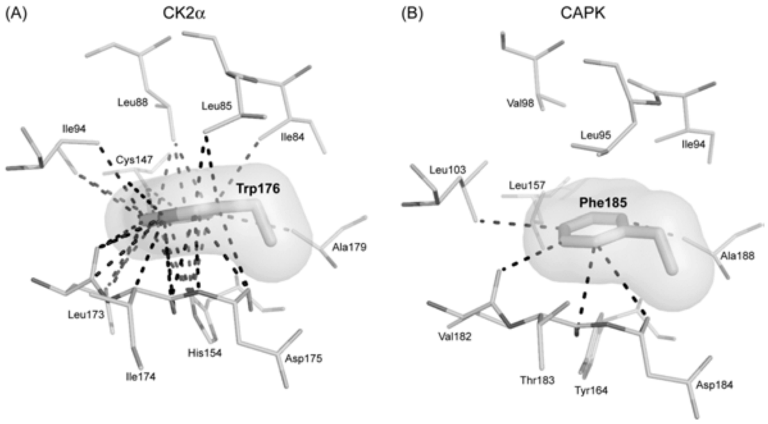


Figure 2.2- 1 The central side chain of the DFG-motif –a member of the regulatory spine and its atomic contacts to neighboring residues. The atomic contacts are drawn by dashed lines with a cutoff of 4 Å. (A) CK2α (from maize) with the unique mutation of the DFG-motif to DW176G. More than 30 atomic contacts are indicated, among them a 3-Å-hydrogen bond from the Trp176 side chain to the peptide carbonyl oxygen of Leu173. (B)PKA with a canonical DFG-motif in the “DFG-in” conformation. A comparison of the two panels reveals the intensive embedding of the Trp176 side chain into its hydrophobic environment (Niefind et al., 2012).

Another important region is the so called hinge/αD helix that connects the two lobes of the protein and supplies the hydrogen bonding counterparts for the binding of the adenine and guanine base of ATP or GTP. This region is flexible and can adopt two different conformations, an open and a close one. The open conformation is unique in respect to all the EPKs structures published to date and causes the conserved R-spine to not be correctly assembled, while in this close conformation the R-spine is completely and correctly assembled in a EPK-canonical way (Battistutta and Lolli, 2011).

The regulatory CK2β subunit is an obligate dimer (Canton et al., 2001). An important region is the "acidic groove" composed by elements from αC (Asp51), αD (Asp70, Glu73, Glu77) and the loop connecting the two helix named "acidic loop" (Asp55, Glu57, Asp59, Glu60, Glu61, Glu63, Asp64) (Chantalat et al.,1999). This acidic cluster has been reported to be able to bind polyamines like spermine (Leroy et al., 1995), polycationic molecules that are able to stimulate the CK2 activity in vitro. CK2 can undergo an autoinhibitory polymerization mediated by the acidic loop (Lolli et al., 2012).

In vivo the fully functional form of CK2 is considered the α2β2 holoenzyme, a heterocomplex of 140 kDa composed of two catalytic α-subunits and two regulatory β-subunits. Besides those of the isolated α- and β-subunits, several crystal structures of the holoenzyme are now available (Niefind et al, 2001; Lolli et al, 2012; Lolli et al, 2014; Schnitzler et al, 2014); the constitutive β2 dimer recruits two α-subunits on opposite sides, originating a 'butterfly'-shaped, prolate heterocomplex in which the two catalytic subunits do not interact with each other. The formation of the holoenzyme causes only minor conformational changes between CK2β-bound and unbound CK2α, leaving unaltered the structural determinants for an effective catalysis, already present in the isolated α-subunit (Lolli et al., 2012). Therefore, the structure of the holoenzyme does not explain how

CK2β is able to modulate the activity of CK2α and if there is a regulation mechanism hidden behind this interaction.

CK2 regulation

The regulatory mechanism of the catalytic activity of CK2 has not been fully understood yet. CK2 is not regulated by 'conventional' mechanisms common to other eukaryotic protein kinases, such as phosphorylation or dephosphorylation events, second messenger binding or reversible association with regulatory subunits. Therefore, the catalytic domain is not shifting between inactive and active conformations. Indeed CK2 is often referred to as 'constitutively active', as its catalytic α subunit has been always found in an active state in the many crystal structures determined so far, with all the structural elements necessary for catalysis locked in the proper position by structural features peculiar to CK2 (Niefind and Battistutta, 2013; Battistutta and Lolli, 2011; Papinutto et al, 2012).

The binding of the so-called 'regulatory' β subunit to form the tetrameric $\alpha_2\beta_2$ holoenzyme leaves the structure of the catalytic subunit substantially unaffected in its main features, so that both the holoenzyme and the isolated α-subunit are catalytically active.

Based on self-aggregation properties leading to filaments of Drosophila CK2, in 1986 Claiborne V. C. Glover proposed that this enzyme could be unconventionally regulated by an oligomerization process, with the formation of inactive regular aggregates that could constitute the functional state of this kinase at rest (Glover, 1986). Glover found that aggregation of Drosophila CK2 at low ionic strength results from a polymerization of the enzyme to form linear filaments. The forces that stabilize the polymer were in contrast with the forces that stabilize the tetramer (stable up to 1M NaCl). The results were confirmed in 1995 by Valero and co-workers (Valero et al., 1995) and they were able to distinguish three different forms of CK2 composed by the single protomer, a ring-like structure and a filamentous form. Currently this mechanism is the most accepted working model for CK2 regulation and it has been supported over time by in vitro and in vivo studies (Valero et al, 1995; Filhol et al, 2003; Thels-Febvre et al, 2005; Hubner et al, 2014). In this model, activity is restored by still unknown events that trigger the dissociation of oligomers into the monomeric, active form of the tetrameric holoenzyme.

Crystal structures of the holoenzyme have supported this regulation model suggesting how CK2 can oligomerize via the formation of circular trimers (PDB IDs 1JWH and 4DGL) and linear oligomers (PDB IDs 4MD7, 4MD8, 4MD9 and 4NH1) (Figure 2.2-2).

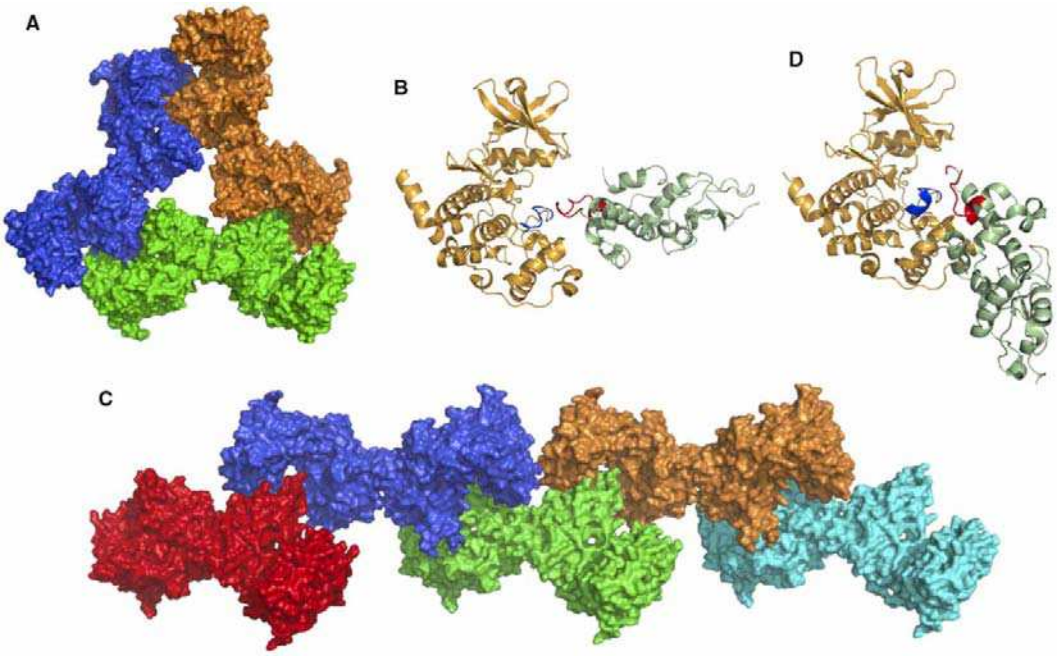


Figure 2.2- 2 Oligomeric forms of CK2. Architecture of the circular trimer (A, viewed along the three-fold symmetry axis) and linear oligomers (C, a linear pentamer in this case) of the CK2 holoenzyme, based on the PDB IDs 4DGL (hexagonal form with asymmetric α2β2) and 4MD7 (monoclinic form with symmetric α2β2), respectively. Single α2β2 monomeric assemblies have different colours. Different inter-molecular orientation between the α-subunit of one monomer (orange) and the β-subunit of the neighbouring monomer (green) in the circular trimer and in the linear oligomers are shown in panels B and D, respectively. In both cases, electrostatic interactions between the basic P + 1 loop of the α-subunit (blue) and the 'acidic loop' of the β-subunit (red) are involved, even though with a different arrangement.

In both cases aggregation is driven by intermolecular electrostatic interactions involving the basic P + 1 loop of the α-subunit and the so-called 'acidic loop' of the β-subunit of another holoenzyme. These two regions were previously found to be implicated in the regulation of the catalytic activity by means of mutagenesis studies [23,24], but, due to the lack of the crystal structure of the holoenzyme at that time, these inhibitory 'secondary interactions' were wrongly interpreted as intra-tetrameric.

Another interesting feature of CK2 that could be involved in the regulation mechanism is that the enzyme is extensively autophosphorylated in vivo, when obtained from cells (Litchfield et al., 1991). The autophosphorylation sites are βSer2 and βSer3; when acidic residues of the acidic loop are mutated to alanines, autophosphorylation is inhibited and the enzyme is hyperactivated (Pagano et al., 2005). Autophosphorylation and the tendency to form aggregates are in accordance with the trimeric organization of the butterfly shaped tetramers found in the crystal packing of the 4DGL structure. The autophosphorylation sites are indeed close to the catalitic cleft of CK2α coming from neighboring tetramers. Moreover, the surface of interaction within the three tetramers is very extended and composed mainly of electrostatic interaction, in accordance with the low ionic strength-dependence of polymerization. In the polymeric organization, CK2 is supposed to be completely inactive due to steric hindrance of the catalytic cleft of CK2α, by different elements coming from neighboring CK2β.

CK2 in vivo might be a "constitutive inactive" kinase that is stimulated by alterations of the ionic status and by polycationic activators.

The significance of the oligomeric forms deduced from crystal packings of the CK2 holoenzyme is substantiated by a recent structural characterization by SAXS that provides evidences of their existence and stability in solution (Lolli et al, 2017)

The idea of a "constitutive active" kinase is therefore recently changing with the data supporting a mechanism of auto-inhibitory polymerization (Figure 2.2-3).

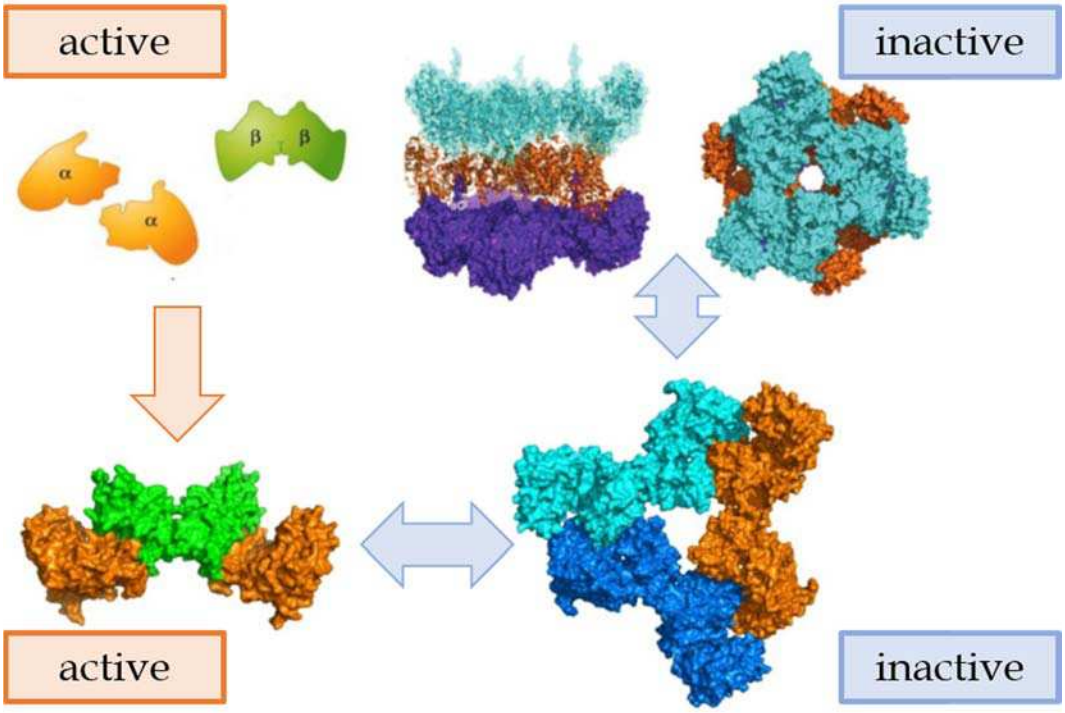


Figure 2.2- 3 Schematic representation of the regulation mechanism proposed for CK2. Once the CK2 tetramer is formed by the stable interaction by regulatory β2-dimer (green) and αsubunits (orange) it is in the fully active form; than it can undergo to auto-inhibitory polymerization mechanism first by forming trimer (inactive) and second by pilling organization of the trimers (inactive). This process of polymerization is reversible and the activity of CK2 can be restored upon necessity via depolymerisation mechanism.

Aim of the work

Recently our collaborators Dr. Stefania Sarno (University of Padua, Department of Biomedical Sciences) and Dr. Giorgio Cozza (University of Padua, Department of Biomedical Sciences) identified a mutant of CK2α with altered activity, hence the interest in knowing if it differs from the WT protein, form a structural point of view. The point mutation is in the N-terminal lobe, and is not specified in this chapter for scientific restrains.

Experimental procedures

Recombinant protein expression

The coding sequence of a mutated human CK2α was cloned in a pT7 expression vector. The construct was used to transform *E. coli* BL21(DE3) cells. Bacterial clones were then conserved at -80 °C. Cells were grown in liquid medium LB (Luria-Bertani; Sambrook and Russel, 2001): 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl. The medium was sterilized by autoclaving it for 20 minutes at 121 °C and implemented with 100 µg/ml ampicillin, to which bacterial cells transformed with the pT7 vector are resistant. *E. coli* cells were grown over-night (ON) in LB medium containing 100 µg/ml ampicillin at 37°C in a suitable shaker. LB medium supplied with 100 µg/ml ampicillin was then inoculated with this over-night culture in the ratio 1:10. The culture was kept at 37°C in a suitable shaker and the growth was followed measuring the OD_{600} (Optical Dispersion at 600 nm, directly related to the amount of cells in suspension). The temperature was slowly decrease until 30°C when an OD_{600} of 0.4 was reached, then protein expression was induced by adding IPTG to a final concentration of 0.1 mM when an OD_{600} of 0.6 was reached. After induction, bacteria were growth for 4-5h at 30 °C. Cells were then harvested by centrifugation (20 min, 4000 g, 4°C). The harvested cells were resuspended in buffer A: 25 mM Tris-HCl, 350 mM NaCl, 1 mM dithiothreitol (DTT), pH 8.0 supplemented with EDTA-free protease inhibitors and lysed with a "French Press" cell disrupter at 1000 psi. Cell debris were then removed from the soluble components extracted by centrifugation of the lysate (45 min, 12000 g, 4°C). After this centrifugation step, the supernatant was filtered using 0.22 µm filters before being loaded onto the column for the first step of the purification procedure.

Purification protocol

An AKTApurifier 10 with a multiple-wavelength detector was used for the chromatographic steps of the purification procedures. The chromatographic runs were followed at 280 nm and chromatograms were recorded using the UNICORN software (GE healthcare).

The filtered supernatant obtained after the mechanical lysis of the bacterial cells was loaded onto a 5 ml HiTrap Heparin HP (GE healthcare) affinity column. The column was equilibrated with buffer A and the sample was loaded. The protein was eluted with a linear gradient procedure with buffer B: 25 mM Tris-HCl, 1 M NaCl, 1 mM dithiothreitol (DTT), pH 8.0.

In order to perform the next step of the purification procedure, the buffer was changed in desalting chromatographic step exploiting a HiPrep 26/10 desalting coloumn equilibrated with buffer C: 25 mM Tris-HCl, 100 mM NaCl, 1 mM dithiothreitol (DTT), pH 8.5.

An anion exchange chromatography was then performed using a MonoQ column (GE Healthcare) equilibrated with buffer C. Elution was performed with a linear gradient procedure with buffer D: 25 mM Tris-HCl, 1 M NaCl, 1 mM dithiothreitol (DTT), pH 8.5.

Then CK2 fractions were further purified by Size Exclusion Chromatography using HiLoad 16/60 preparation grade Superdex 75 column(GE healthcare) equilibrated with 25 mM Tris-HCl, 500 mM NaCl, 1 mM dithiothreitol (DTT), pH 8.5. CK2 was concentrated to 10 mg/ml and flash-frozen in liquid nitrogen for storage at -80 °C.

SDS-PAGE

The purification procedure was followed by SDS-PAGE. Samples were heated for 5 minutes at 95 °C to denature them in the sample-loading buffer: 60 mM Tris pH 6.8, 10% glycerol, 2% SDS, 0.01% bromophenol blue, 2% β-mercaptoethanol. Then they were loaded with Low molecular weight See Blue® Plus2 Pre-Stained Standard (*Invitrogen)* markers on SDS-PAGE gels. A voltage of 180 V was used to performed the electrophoretic run with the gels mounted vertically on the electrophoretic apparatus and in contact with the running buffer: 25 mM Tris pH 8.3, 250 mM glycin, 0.1% SDS. The gels were colored for 20 minutes under soft shaking with the staining solution: 40% EtOH, 10% AcOH, 0.1% w/v Coomassie G-250 and then the destaining solution: 40% EtOH, 10% AcOH and a final wash in water were used to remove unspecific staining.

Protein crystallization

Crystallization trials were performed in the known crystallization conditions for the WT protein 0.1 M MES pH 6.5, 0.2 Ammonium Sulphate, 20-24% w/v PEG 5000 MME/PEG 2000 MME.

Single Crystal X-ray Crystallography

Crystals were cryoprotected with 20% ethylene glycol, mounted on loops, and frozen in liquid nitrogen. Diffraction data were collected at synchrotron sources (ID29 at ESRF, Grenoble) at 100 K.

Diffraction data were initially processed (indexing, integration and scaling) with the XDS package (Kabsch, 2010). Integrated reflections were then reduced and merged with AIMLESS (Evans, 2006) (Evans, An introduction to data reduction: space group determination, scaling and intensity statistics, 2011) from the CCP4 suite (Winn, et al., 2011). The structures were solved by molecular replacement using the previously deposited model with the PDB ID 3QA0 as search model and the program PHASER (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The initial structures were then refined alternating manual steps of real space refinement using the program COOT (Emsley & Cowtan, 2004) (Emsley, Lohkamp, Scott, & Cowtan, 2010) and automatic steps of reciprocal space refinement using the suite PHENIX (Adams, et al., 2010). PYMOL was used to make all the images and the superposition analysis.

Results and Discussion

Recombinant protein expression and purification

The expression and the purification of the mutant of CK2α were carried out as described in the experimental procedures. The first step exploited an intrinsic affinity of the protein for a heparin column. Figure 2.2-4 shows the chromatogram of the first step of the purification procedure.

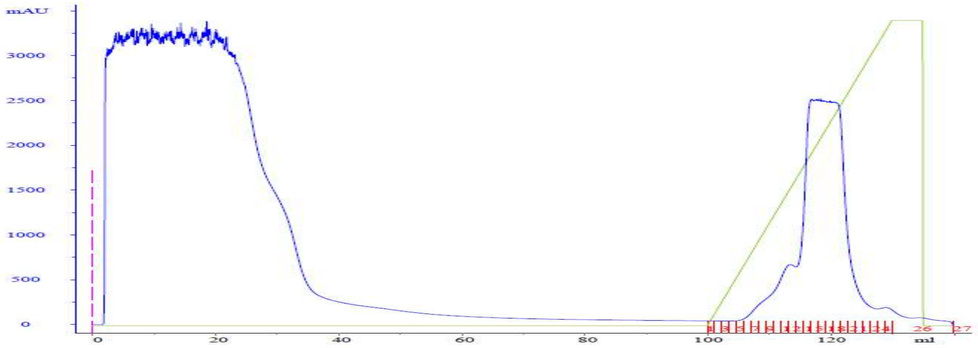


Figure 2.2- 4 CK2α heparin affinity chromatogram; in blue the elution profile (Abs at 280 nm), in green the step gradient of NaCl, in red the collected fractions.

The buffer was exchanged in a desalting step, and then an anion exchange chromatography was performed (figure 2.2-5).

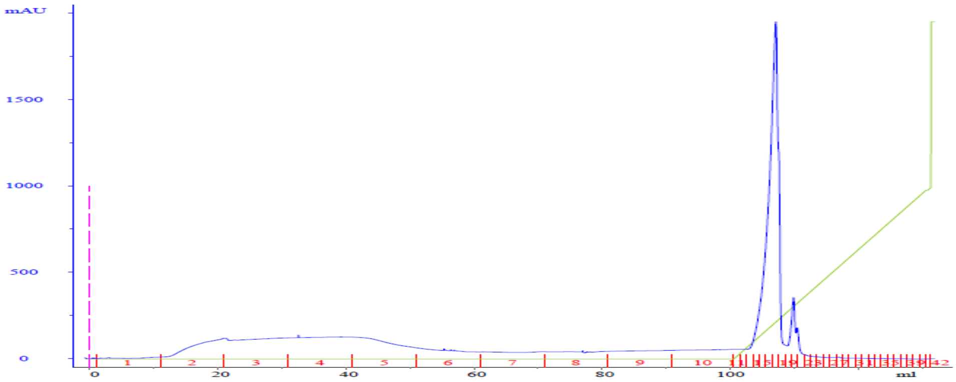


Figure 2.2- 5 CK2α anion exchange chromatogram; in blue the elution profile (Abs at 280 nm), in green the step gradient of NaCl, in red the collected fractions.

The eluate was further purified by size exclusion chromatography, using a HiLoad 16/60 preparation grade Superdex 75 column (GE Healthcare). Figure 2.2-6, shows the chromatograms for this final purification step.

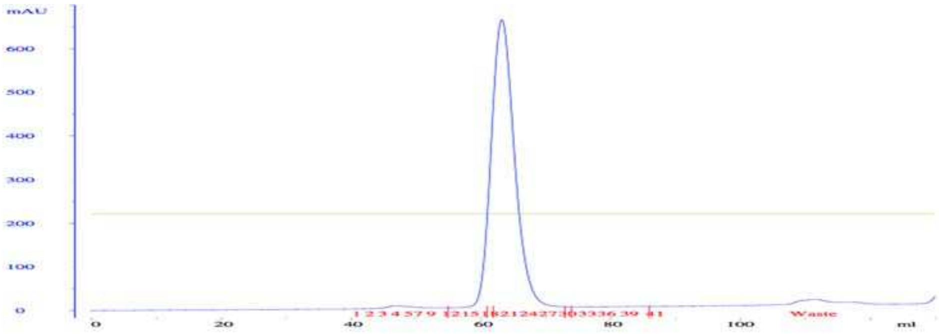


Figure 2.2- 6 CK2α Size exclusion chromatography chromatogram.

The proceeding of the purification procedure was followed by SDS-PAGE, figure 2.2-7 shows the final product of the purification. A second band is present in the fractions, however the purity of the protein was high enough to obtain crystals.

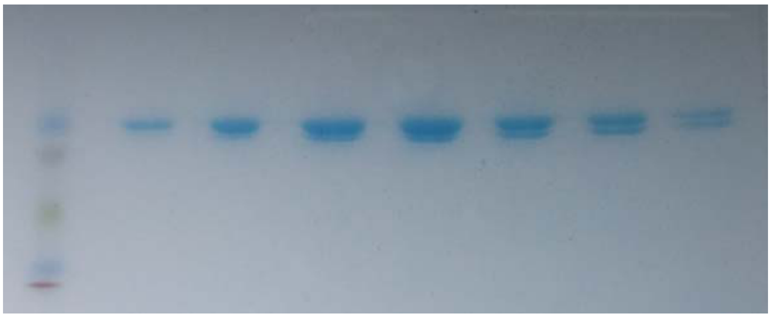


Figure 2.2- 7 Coomassie stained SDS-page of the fractions of the Size exclusion chromatography.

Protein crystallization and single crystal x-ray crystallography

In order to perform crystallization trials the mutant of CK2α was either diluted in water in presence of 1.5-3% DMSO before setting up the crystallization drops, or the drop was made in a 1 + 1 + 1 fashion, with 1 µl protein, 1 µl precipitant, 1 µl water. Crystallization trials were performed in the known crystallization conditions for the WT protein. I managed to obtain crystals as shown in figure 2.2-8.

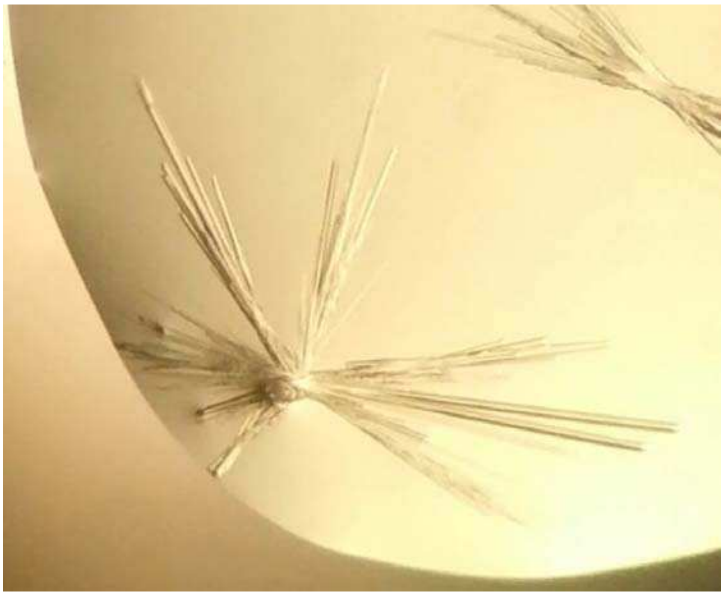
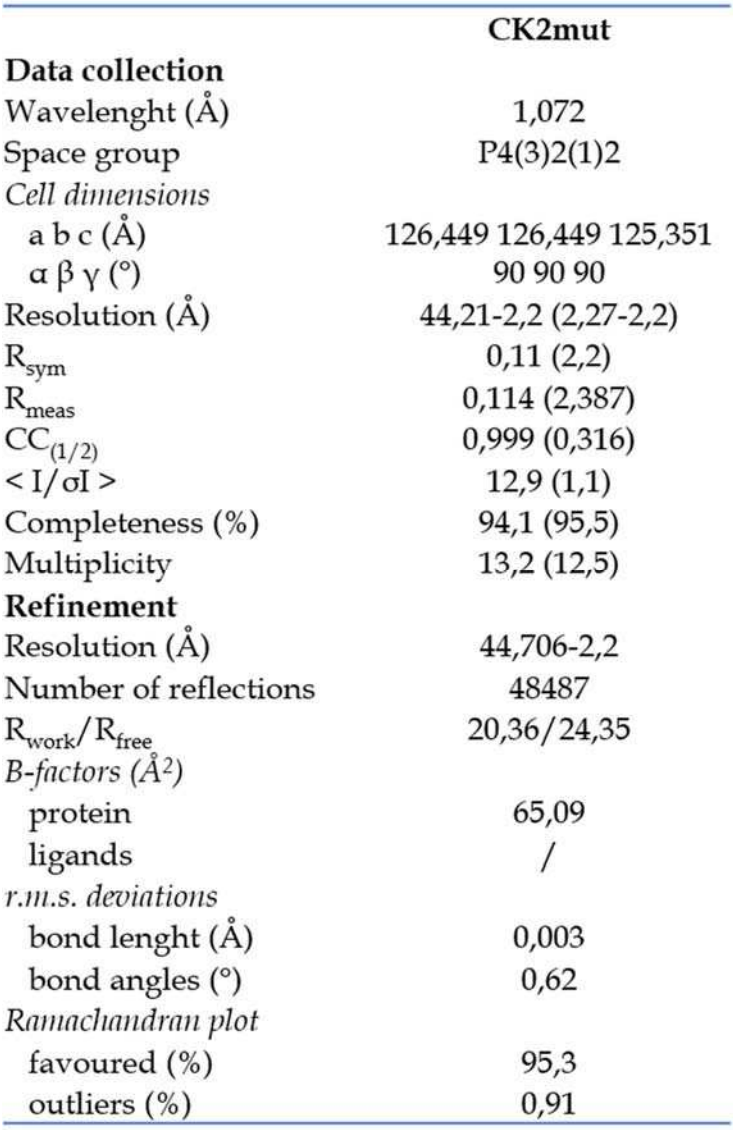


Figure 2.2- 8 crsytals of the mutant of CK2α

I collected more then 20 datasets at ESRF (European Synchrotron Radiation Facility), both on APO crystals and crystals obtained in presence of Mg^{2+} and AMP-PNP, at beamline ID29. No differences were observed in the two types of structures, hence I am here reporting only one of them, solved from the crystal that gave the best data collection statistics, shown in table 2.2-1.



	CK2mut
Data collection	
Wavelenght (Å)	1,072
Space group	P4(3)2(1)2
Cell dimensions	
a b c (Å)	126,449 126,449 125,351
α β γ (°)	90 90 90
Resolution (Å)	44,21-2,2 (2,27-2,2)
R_{sym}	0,11 (2,2)
R_{meas}	0,114 (2,387)
$CC_{(1/2)}$	0,999 (0,316)
$< I/\sigma I >$	12,9 (1,1)
Completeness (%)	94,1 (95,5)
Multiplicity	13,2 (12,5)
Refinement	
Resolution (Å)	44,706-2,2
Number of reflections	48487
R_{work}/R_{free}	20,36/24,35
B-factors (Å²)	
protein	65,09
ligands	/
r.m.s. deviations	
bond lenght (Å)	0,003
bond angles (°)	0,62
Ramachandran plot	
favoured (%)	95,3
outliers (%)	0,91

Table 2.2- 1 Data collection and refinement statistics

In the asymmetric unit, as in the WT structure (PDB ID: 3QA0) there are two molecules. The ATP-binding region is flexible, and there are two different conformations in the two molecules of the asymmetric unit. One of these conformations is identical to the WT protein, as shown in figure 2.2-9a. In the other one the side chain of Arg 47, is inside the cavity, as shown in figure 2.2-9b.

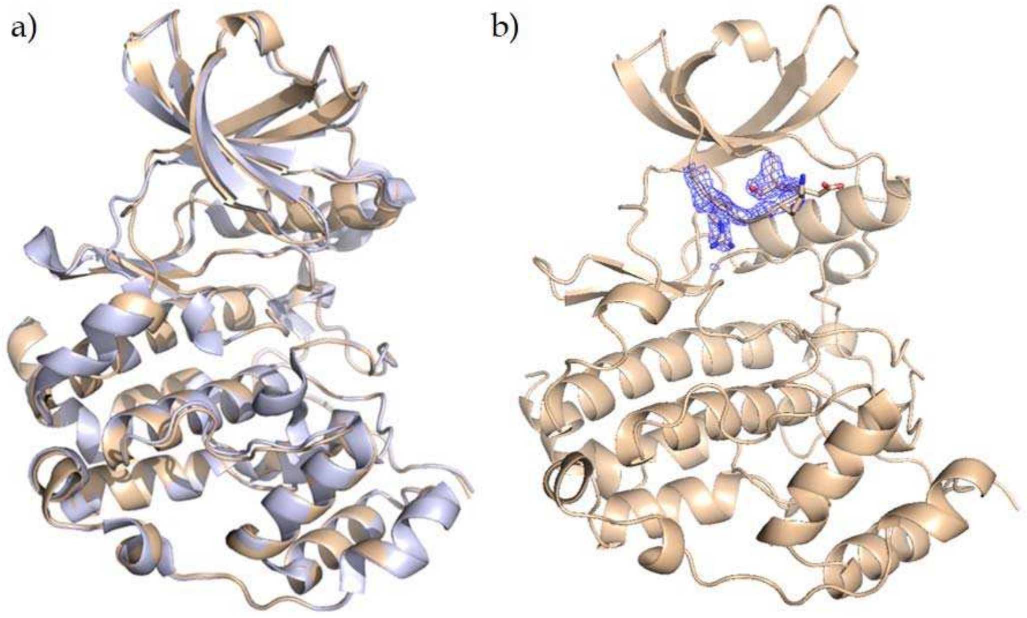


Figure 2.2- 9 a) superposition of molecule A (wheat) with the apo structure (purple) b) CK2α mutant in a conformation with Arg47 in the ATP-binding site. The $2F_O$-F_C map is shown contoured at 1 σ.

Also the conformation with Arg47 in the binding pocket is found in a WT structures (PDB ID: 4IB5), as shown in figure 2.2-10.

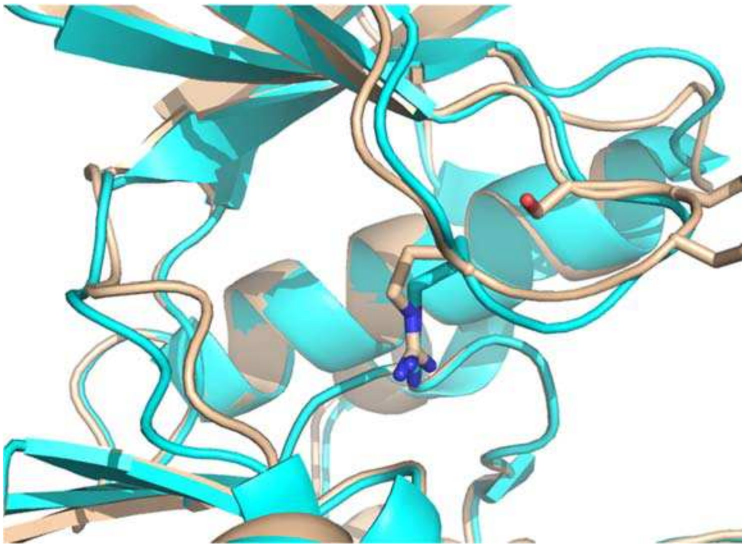


Figure 2.2- 10 superposition of the structure of the CK2α mutant with Arg47 in the ATP-binding pocket (wheat) with the same conformation found in the WT protein (cyan, PDB ID: 4IB5).

Given that from a structural point of view, there is not a stabilization of a particular conformation differentiating this mutant in respect to the WT protein, the explanation of its altered activity has to be found in a different behavior in its interaction with the substrate.

Conclusions

A mutant of CK2α with altered activity was succesfully expressed, purified, crystallized and several diffraction datasets were collected at synchrotron sources. The structure was solved and no significant differences were found in respect to the WT protein. This suggests that the altered activity is not caused by a structural conformational change or a stabilization of a particular conformation of flexible regions, but the explanation has to be found in a altered interaction with the substate.

PART THREE – other proteins: BCL-XL and CASQ

3.1 BCL-XL

Introduction

Programmed Cell Death: an essential fate

Programmed cell death, is an essential cell fate of multicellular organisms. Cell death is a necessary part of the life cycle of living organisms, and it is involved in various physiological processes such as differentiation, homeostasis and regulation of the immune system (Lodish, et al., 2013).

Several different stimuli can trigger this process, called "apoptosis" (Kerr, Wyllie, & Currie, 1972) which is a result of a complex interplay between factors that promote it or suppress it.

Being a programmed process, apoptosis is different from other types of cell death, such as "necrosis" which involves a loss of the membrane integrity with the consequent release of cellular contents and a strong inflammatory response in tissue. In apoptosis, on the other hand, membrane enclosed structures, called apoptotic bodies, form and are then incorporated by macrophages without causing an inflammatory response (Gewies, 2003).

Over the last decades, a number of biological and genetic studies have been performed, showing that the Bcl-2 family of proteins plays a key role in controlling the process of programmed cell death. The first mammalian apoptotic gene to be cloned was *bcl-2* (B-cell CLL/lymphoma 2), isolated from human follicular lymphomas (Tsujimoto & Croce, 1986).

Apoptosis can be triggered in two main ways called the extrinsic and the intrinsic pathway.

In the intrinsic pathway (figure 3.1-1), the mitochondrion has a fundamental role. The first step is the expression of "BH3-only" proteins like Puma, Noxa and Bim, that be controlled by transcriptional and post-transcriptional factors, or the activation of Bad, otherwise inactive in the cytosol, by dephosphorylation. Different stimuli result in the initiation of the process by different proteins; growth factor deprivation leads to the activation of Bad, unrepaired DNA damage induces the transcription of Puma and Noxa, or Bim is released from the cytoskeleton as a result of loss of extracellular signalling. As a second step, the BH3-only proteins induce the mitochondrial outer membrane permeabilization (MOMP) through highly specific interactions with other proteins members of the same Bcl-2 family that are present on this membrane.

The low permeabily of the membrane is normally maintained by the binding of pro-apoptotic Bcl-2 family members (e.g. Bax or Bak) and anti-apoptotic Bcl-2 family members (e.g. Bcl-2 or Bcl-xL) on the MOM. During the apoptotic process, BH3-only proteins induce the formation of Bax and Bak oligomers that generate pores in the MOM, necessary for the release in the cytosol of the cytochrome *c*. After the binding of cytochrome c to the protein Apaf-1, the apoptosome, an heptameric wheel shaped protein complex, forms. The so formed apoptosome then activates caspase-9 which initiates a caspase cascade, a key step in the cell death by apoptosis (Youle & Strasser, 2008) (Lodish, et al., 2013).

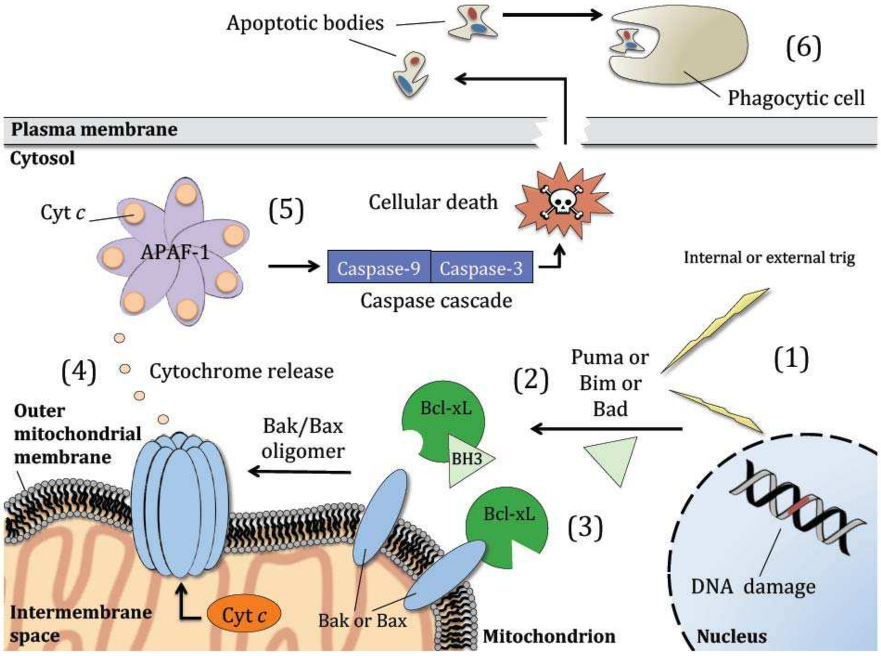


Figure 3.1 - 1 The intrinsic pathway of apoptosis. (1) Trigger for the intrinsic pathway like hormone signal deprivation, oxidative stress, UV radiation, DNA damage, and cytotoxic drugs. (2) Activation of the BH3-only proteins. (3) Induction of the Bak/Bax oligomerization by direct activation of Bax and Bad or via inhibition of one of the anti-apoptotic partner (indirectly). (4) Cytochrome c release after MOMP and consequent binding to Apaf-1 in order to form the apoptosome. (5) Caspase activation and caspase cascade leading to cell death. (6) Formation of the apoptotic bodies and engulfment by macrophages. Adapted from (Lodish, et al., 2013).

In the extrinsic pathway (figure 3.1-2), apoptosis is stimulated by death signals like the Fas ligand, a surface protein situated on a natural killer cell or cytotoxic T lymphocytes, that binds to the Fas receptor of the cell that has to be killed. The Fas receptor is thus activated and binds to FADD (Fas-associated death domain) that activates caspase-8, which then initiates the caspase cascade. Caspase-8 cleaves the BH3-only protein Bid, then the pathway continues as in the intrinsic way (Youle & Strasser, 2008). The key point in the apoptotic process is the permeabilization of the mitochondrial outer membrane (MOMP) as a result of a complex interplay between proteins of the Bcl-2 family.

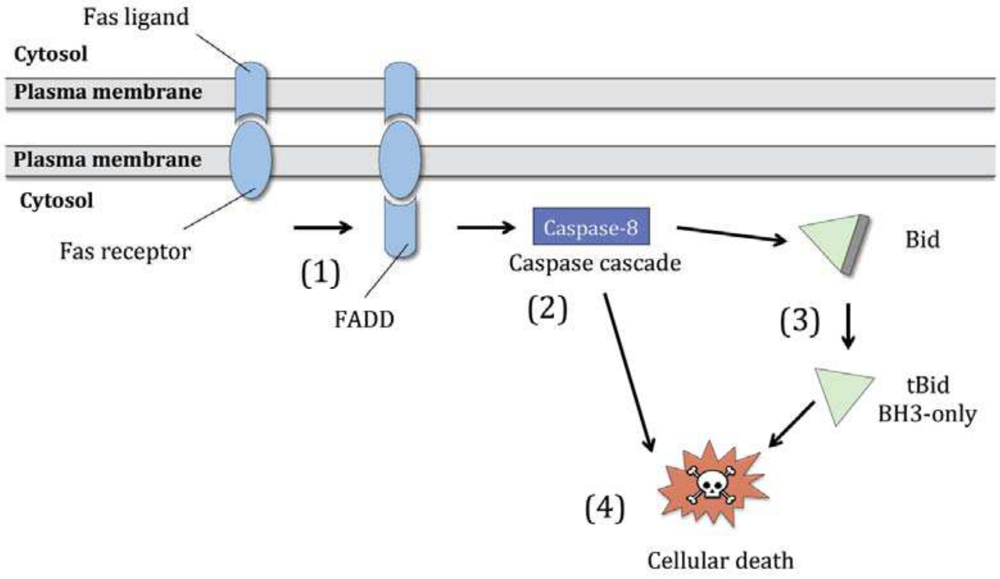


Figure 3.1 - 2 The extrinsic pathway of apoptosis. (1) Activation of the cytosolic protein FADD after interaction between Fas ligand and Fas receptor. (2) Activation of caspase-8 that leads to cellular death via caspase cascade or (3) via C-terminal truncation of BID. The truncated form of Bid, tBid, translocates to mitochondria and promotes the intrinsic pathway leading to cellular death (4). Adapted from (Lodish, et al., 2013).

Malfunctions in the apoptotic process play an important role in several human diseases. A defective apoptosis can result in cancer, autoimmune diseases and spreading of viral infection, while an excessive apoptosis causes or enhances neurodegenerative disorders, AIDS and ischaemic disease (Gewies, 2003).

The anti-apoptotic members are often altered in tumour cells, over-antagonizing the apoptosis (Vela & Isabel, 2015). An overexpression of anti-apoptotic members, like Bcl-2 and Bcl-xL, has been found in several cancers, among which: follicular lymphomas (Gaulard, 1992), small cell lung carcinomas (Higashiyama, Doi, Kodoama, Yokouchi, & Tateishi, 1995), breast cancer (Lipponen, 1995), chronic lymphocytic leukaemias (Harada , 1998), prostate (Keshgegian, Johnston, & Cnaan, 1998) and melanoma (Leiter, Schmid, Kaskel, Peter, & Krahn, 2000).

Bcl-2 family members have therefore been identified as drug targets with the goal to develop new anticancer therapies (Min & Reynods, 2009).

The Bcl-2 family

The Bcl-2 protein family is a family of proteins consists of more than 25 member that can be classified according to their function or structural features (figure 3.1-3).

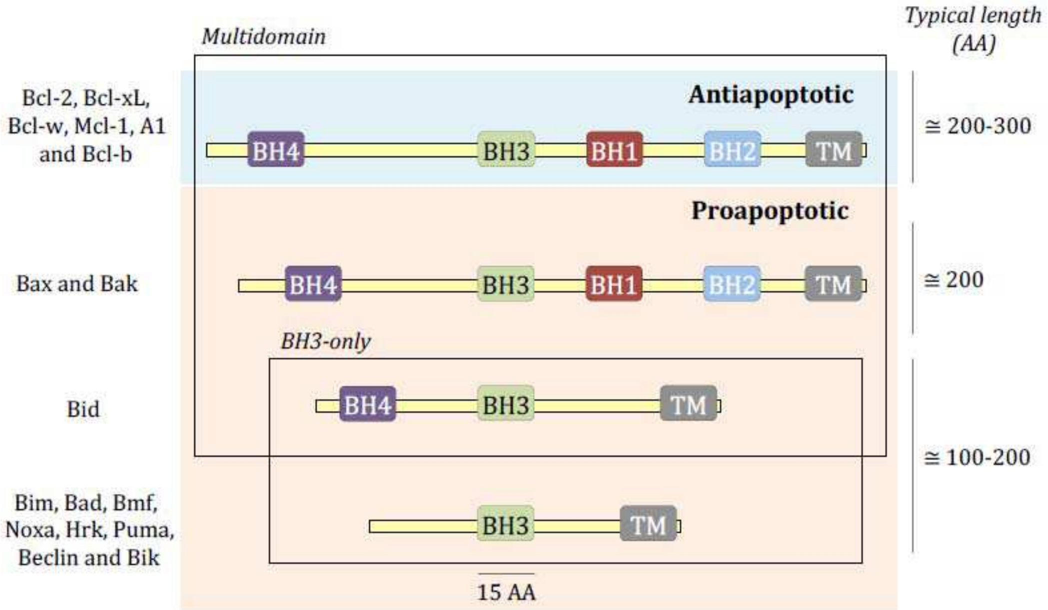


Figure 3.1 - 3 Schematic overview of the most relevant members of the Bcl-2 family, divided functionally in pro-apoptotic and anti-apoptotic members. The second classification regards structural features and is based on sequence homology to the founding member, Bcl-2, in one of the four Bcl-2 homology domains (BH domains). Multi-domain members have multiple BH domains. BH3-only proteins on the other side are intrinsically unstructured and contain only the BH3 domain. Furthermore, most members contain a C-teminal transmembrane domain (TM). The member Bid is an exception, being classified as a pro-apoptotic BH3-only protein but sharing structural and functional features with multi-domain members. Adapted from (Shamas-Din, Kale, Leber, & Andrews, 2013).

From the fuctional point, in the family we find the pro-apoptotic members (Bax, Bak, Bim, Bad, Bmf, Noxa, Hrk, Puma, Beclin, and Bik) and the anti-apoptotic members (Bcl-2, Bcl-xL, Bcl-w, A1, Mcl-1 and Bcl-b).

From the structural point, the proteins can by classified on the basis of the presence of one or more of the four "Bcl-2 homology" domains (BH domains), dividing the multi-domain members from the BH3-only proteins. The BH3 domain is common to all family members and is responsible for mediating the interaction with the anti-apoptotic members (Petros, Olejniczak, & Fesik, Structural biology of the Bcl-2 family of proteins, 2004). Within the class of multi-domain Bcl-2 proteins pro-apoptotic members and anti-apoptotic members show remarkable structure similarity (Youle & Strasser, 2008).

Many of the Bcl-2 family members possess a TM C-terminal hydrophobic domain. These proteins can however also be found in the cytosol, and be translocated in the membrane only after being activated (Petros, Olejniczak, & Fesik, Structural biology of the Bcl-2 family of proteins, 2004).

Bcl-xL: structural features

Structural studies on the proteins belonging to the Bcl-2 family are crucial for understanding their molecular mechanism of action at atomic level and how they interact with one another.

Bcl-xL belongs to the Bcl-2 family, and is an anti-apoptotic protein. After finding that the recombinant expression and the purification of Bcl-xL were practically easier to perform in respect to those of other anti-apoptotic proteins, like Bcl-2, Bcl-xL was chosen as a model (Shamas-Din, Kale, Leber, & Andrews, 2013). Exploiting a combination of X-ray crystallography and NMR spectroscopy the first structure of Bcl-xL was solved in 1996 (Muchmore, et al., 1996).

Bcl-xL is a multi-domain member of the Bcl-2 family (Figure 3.1-4), 233 amino acids long and with a MW of 26 kDa.

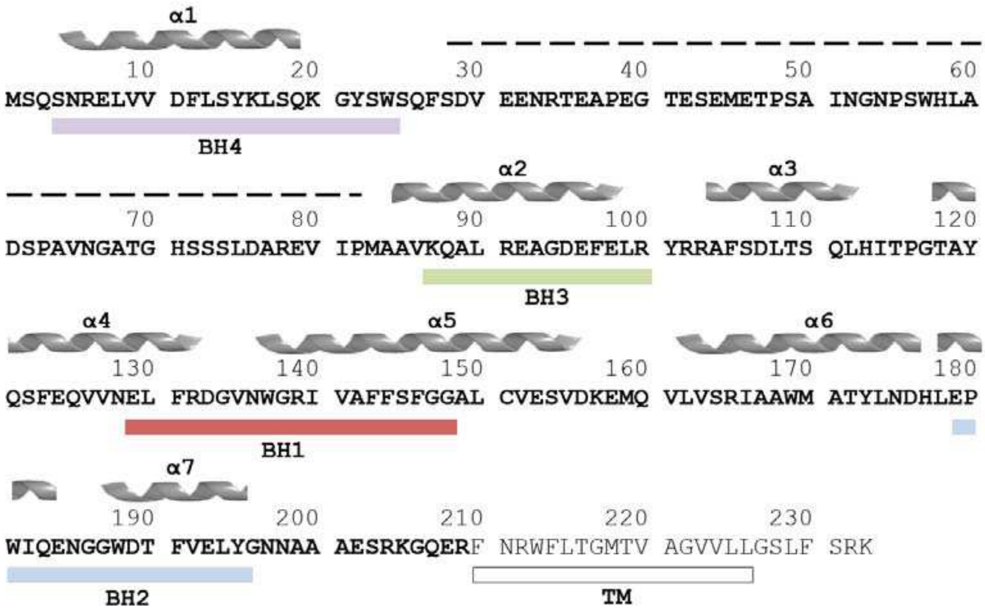


Figure 3.1 - 4 Sequence of Bcl-xL. The colored bars indicate the four domains of the protein (BH1, BH2, BH3, BH4) that are also shared by the other Bcl-2 multi-domain family members; the white one the transmembrane domain. The dotted line represents the flexible loop from residue 27 to 82. α-Helices secondary structures are shown above the sequence in grey cartoon representations. Adapted from (Muchmore, et al., 1996).

The structure of Bcl-xL consists of two central “core” α-helices, $\alpha5$ and $\alpha6$, flanked by the amphipathic helices $\alpha1$, $\alpha2$, $\alpha3$, $\alpha4$ and $\alpha7$. The two central helices contain predominantly hydrophobic residues and form a helix-turn-helix hairpin (Figure 3.1-5a). Between $\alpha1$ and $\alpha2$, there is a long flexible loop, as evident by the lack of electron density from residue 27 to residue 82 and by NMR data in solution. The amino acid sequence of this region is highly variable among Bcl-2 family members (Petros, Olejniczak, & Fesik, Structural biology of the Bcl-2 family of proteins, 2004).

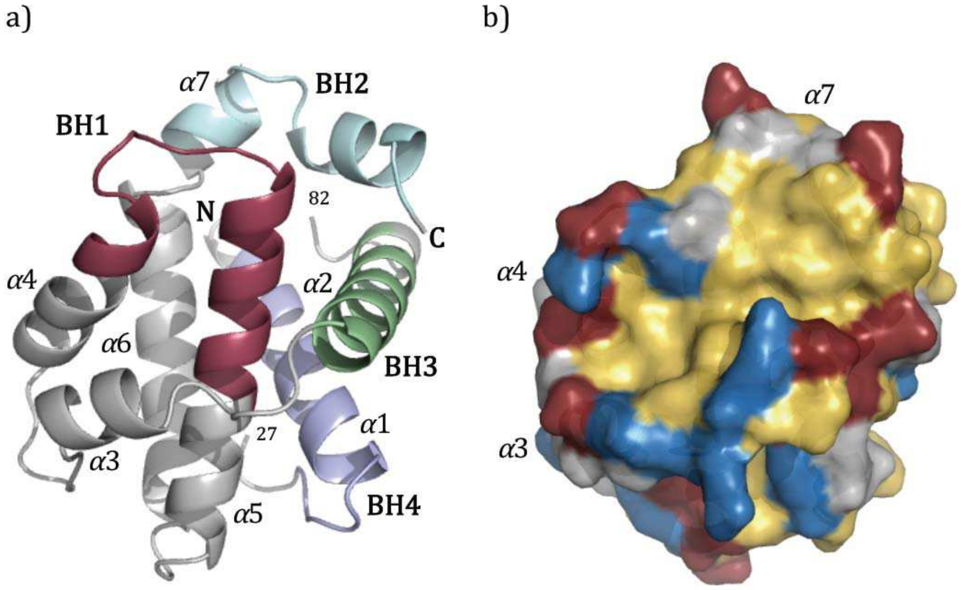


Figure 3.1 - 5 Bcl-xL 3D structure in cartoon and surface representation. a) 3D cartoon representation of Bcl-xL from PDB ID 1MAZ. In the picture are represented BH1 red, BH2 cyan, BH3 green and BH4 purple. The unstructured loop is here omitted. b) Bcl-xL accessible surface area (ASA) with a probe radius of 1,4 Å from PDB ID 1MAZ. The yellow surface represent the hydrophobic groove composed by Leu, Val, Ile, Phe, Tyr, Gly, Met, Trp and Ala residues. The red one represents the Asp and the Glu. The blue one represents Arg, Lys and His.

The protein has a large hydrophobic groove (Figure 3.1-5b), that was confirmed to be the interaction site for the BH3 region pro-apoptotic members of the Bcl-2 family (Sattler, et al., 1997).

After the first apo structure in 1996, several structures of Bcl-Xl in complex with peptides corresponding to the BH3 regions of pro-apoptotic members have been solved as summarized in table 3.1.1

protein	PDB ID	reference
Bak	1BXL	(Sattler, et al., 1997)
Bad	1G5J	(Petros, et al., 2000)
Beclin	2P1L	(Obersatain, Jeffrey, & Shi, 2007)
Bax	3PL7	(Czabotar, Lee, Thompson, Wardak, Fairlie, & Colman, 2011)
Puma	2M04	(Follis, et al., 2013)
Bid	4QVE	(Rajan, Choi, Baek, & Yoon, 2015)
Bim	4QVF	(Rajan, Choi, Baek, & Yoon, 2015)

Table 3.1 - 1 summary of bcl-xl structures in complex with BH3 domains of pro-apoptotic proteins

With a combination of structural an mutational studies, four hydrophobic residues positioned at regular intervals on the BH3 domain were found to be essential for the binding to Bcl-xL, with the establishment of hydrophobic interactions with residues present in the hydrophobic groove of Bcl-xL, like Phe 97, Tyr 101, Phe 105, Leu 130, Val 141 and Tyr 195. Minor structural details based on sequence variations among the BH3-only proteins as well as the binding partner have thus to be identified in order to better understand the mode of interactions and specificity (Rajan, Choi, Baek, & Yoon, 2015).

In figure 3.1-6 an example of interaction of a BH3 peptide (from Bid) with the four hydrophobic residues highlighted.

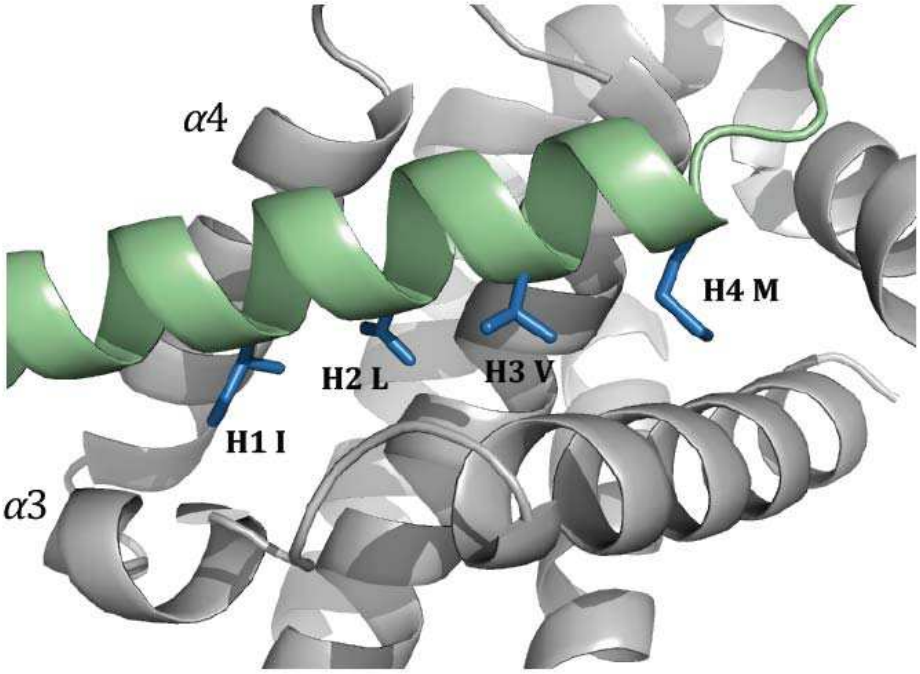


Figure 3.1 - 6 Cartoon representation of Bcl-xL in complex with a BH3 peptide from Bid. Structure of Bcl-xL/Bid complex from PDB ID 4QVE. The picture shows the α-helix conformation of Bid (green) bound to Bcl-xL (grey). The four hydrophobic residues for Bid are show in blue.

These structural studies led to the design of BH-mimetic compounds with the identification of modified peptides, peptidomimetics and small-molecules, some of which are now in clinical trials (Vela & Isabel, 2015). Since modified peptides and peptidomimetics rarely progress to late-stage clinical trials because of their poor pharmacological properties (Lessene, Czabotar, & Colman, 2008), small organic molecules are more promising.

Bcl-xL inhibitors

A crucial point in the development of Bcl-Xl inhibitors was the discovery in 2005 of ABT-737 (Figure 3.1-7c) a molecule that showed an affinity for Bcl-xL a few orders of magnitude higher than previously reported compounds (Oltersdorf, et al., 2005). The crystal structure of Bcl-xL with ABT-737 was then solved in 2007 (Lee, et al., 2007).

ABT-737 interacts with Bcl-xL in a way that resembles that of the BH3 peptides. In particular the chlorobiphenyl group of the molecule mimics the second hydrophobic residue (H2) of the BH3 peptide, engaging with the hydrophobic pocket p2 on the surface of Bcl-xL. In a similar way, the thiophenyl group mimics the H4 residue engaging thus the p4 hydrophobic pocket (Oltersdorf, et al., 2005)(Figure 3.1-7a,b).

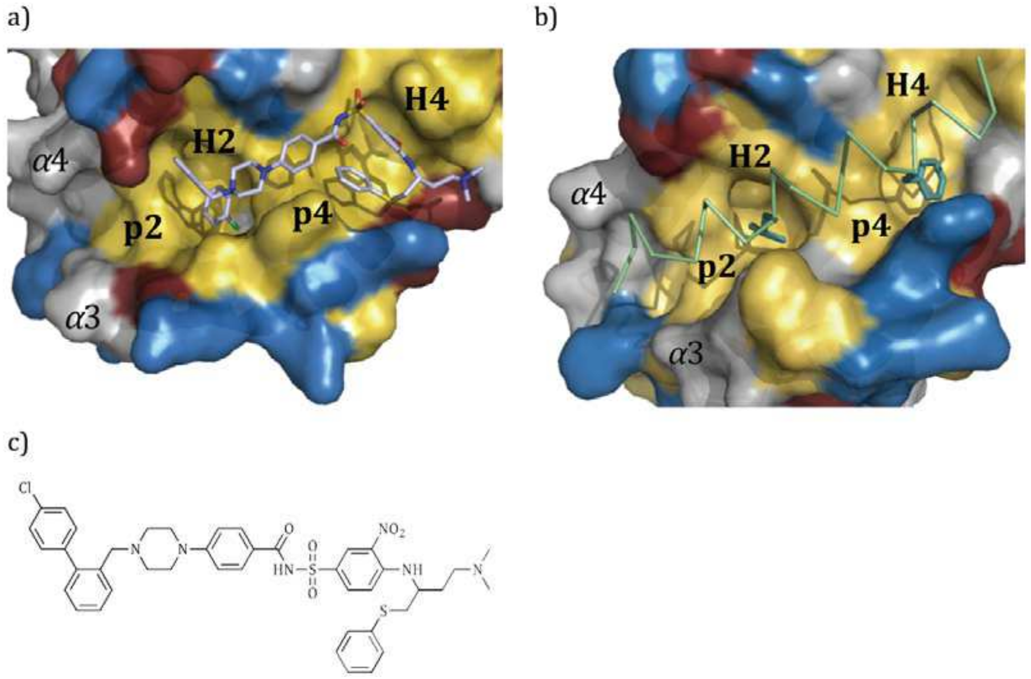


Figure 3.1 - 7 Bcl-xL surface representation in complex with ABT-737 and a BH3 peptide from Bim. a) Bcl-xL/ABT-737 structure from PDB ID 2YXJ. ABT-737 engages the hydrophobic pocket p2 and p4 on the Bcl-xL surface. Hydrophobic residues Phe, Leu, Ile, Pro, Ala, Val, Tyr and Trp are colored in yellow. His, Lys and Arg are colored in blue while Asp and Glu are colored in red. b) Bcl-xL/Bim complex from PDB ID 4QVF.A Leu and a Phe corresponding to H2 and H4, from Bim interacts with the p2 and p4 pockets respectively c) Molecular structure of ABT-737.

ABT-737 fails the classical rules of drug likeness like molecular mass, number of heteroatoms and functional groups (Lessene, Czabotar, & Colman, 2008). However, the knowledge that the interactions with the p2 and p4 pockets are crucial led to modifications of ABT-737 designing other compounds that entered clinical trials for cancers treatment (Table 3.1-2) (Besbes, Mirshahi, Pocard, & Billard, 2015).

Name	Target	Affinity	Stage
GX15-070	Bcl-2, Bcl-xL, Bcl-w, Mcl-1	μM	Phase II
AT-101	Bcl-xL, Bcl-2, Mcl-1	sub-nM	Phase II
ABT-263	Bcl-xL, Bcl-2, Bcl-w	nM	Phase II
ABT-199	Bcl-2	sub-nM	Phase III

Table 3.1 - 2 | Most advanced small molecules targeting Bcl-2 family proteins. GX15-070 (Nguyen, et al., 2007), AT-101 (Wang, et al., 2006), ABT-263 (Tse, et al., 2008), ABT-199 (Touzeau, et al., 2014).

Around fifteen crystal structures of Bcl-xL in complex with small-organic molecules are deposited on the PDB, showing that the interaction with the p2 and p4 pockets are crucial .

The Bcl-xL binding site is highly flexible, as shown by a comparison of the different structures of Bcl-xL in complex with BH3 peptides, peptidomimetics and small molecules and this should be a critical consideration in rational drug design approaches and optimization efforts (Liu, Beugelsdijk, & Chen, 2015).

This high flexibility of the binding site probably has a functional role, as it could facilitate interaction of Bcl-xL with the BH3 domain of various proteins during the apoptosis (Liu, Beugelsdijk, & Chen, 2015). These observations are coherent with a recent work that analyses all existing PDB structures of monomeric Bcl-xL, both in the apo form and in complex with various small molecule and peptide ligands. The results show that the binding interface is much more dynamic than previously recognized (Liu, Beugelsdijk, & Chen, 2015) (Figure 3.1-8).

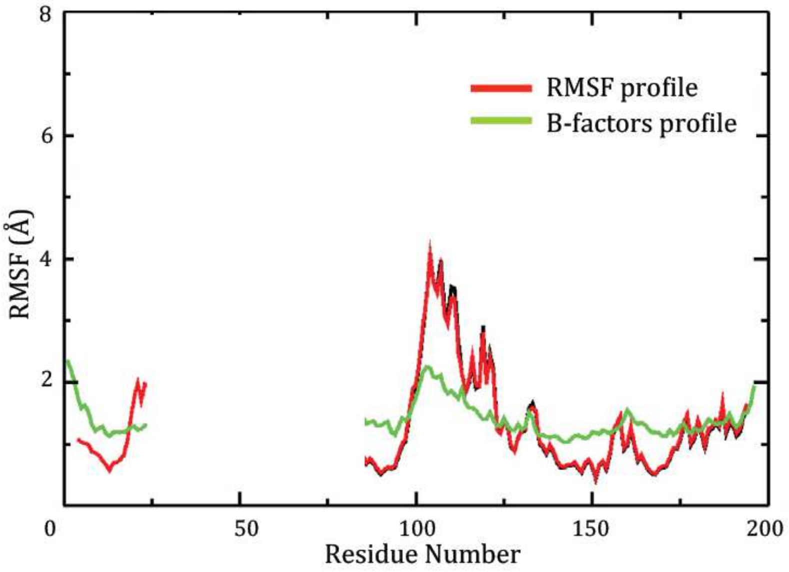


Figure 3.1 - 8 The graphic represents the root mean-square fluctuation profiles (RMSF). RMSF quantifies the magnitude of atomic positional fluctuation around the mean, in order to quantify the variations in the Bcl-xL PDB structures. The most dynamic region overlaps the α2- α4 helices. The use of B-factors to calculate the RMSF profiles is consistent with that derived from atomic positions.

Aim of the work

Apoptosis, also known as programmed cell death, is an essential cell fate in multicellular organisms. Its malfunction plays an important role in a lot of human disorders like cancer and autoimmune diseases. In the last decades, the biological process was extensively studied and the attention focused on the Bcl-2 protein family. These proteins have been found to play a key role in apoptosis influencing the cell fate through a complex network of protein-protein interactions. Bcl-2-like proteins were divided into anti-apoptotic (such as Bcl-xL) and pro-apoptotic members. It has been reported that anti-apoptotic members are often overexpressed in tumor cells, protecting them from programmed cell death stimuli, and conferring resistance to conventional anticancer therapies. This has burst investigations on members of the Bcl-2 family, identifying them as suitable drug targets for the development of new anticancer therapies. Structural studies on Bcl-2 proteins are relevant for understanding their molecular mechanism of action at atomic level and to reveal how they can interact with each other.

In this context, I was involved in a medicinal chemistry project in collaboration with Dr. Massimo Bellanda (Dept. of Chemical Sciences), Dr. Mattia Sturlese (Dept. of Farmaceutical Sciences) and prof. Giuseppe Zagotto (Dept of Farmaceutical Sciences), on a pro-survival and anti-apoptotic protein factor, Bcl-Xl, with the aim of developing inhibitory compounds with anti cancer potential. The structural characterization is fundamental for the rational drug design of ligands, therefore some small molecules (Table 3.1-3) were identified by in silico docking procedures by our collaborators, to be studied and further functionalized as possible drugs targeting Bcl-xL.

To study their binding modes and their interactions with Bcl-xL by single crystal x-ray crystallography, first I had to produce in bacteria two constructs of the protein, then I had to establish a solid protocol for purification and crystallization and to perform crystallization trials with the molecules identified by our collaborators.

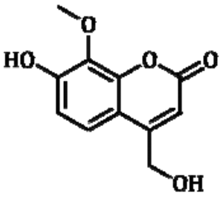
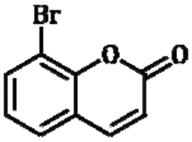
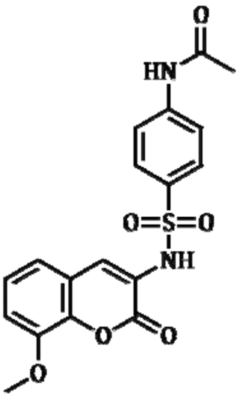
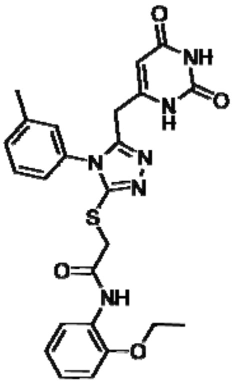
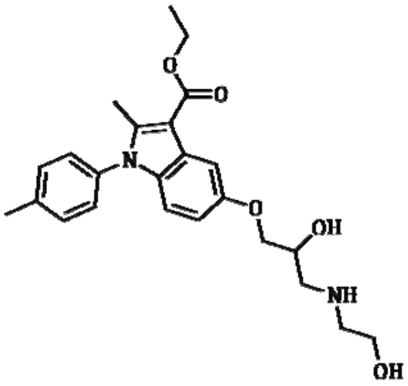
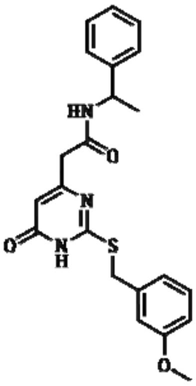
Comp. 413	Comp. 622	Comp. 177
		
Comp. 0160	Comp. 0200	Comp. 0390
		

Table 3.1 - 3 Ligands identified by in silico analysis.

Experimental procedures

Recombinant protein expression

Two different coding sequences of human Bcl-xL were previously selected and cloned in pET15b expression vectors. These two constructs were previously used to transform *E. coli* BL21(DE3) cells. Clones were then conserved at -80 °C. In both construct the TM domain was deleted (Bcl-xL [ΔTM209-233]). In addition to this deletion, in the second construct was removed also the region from AA27 and AA82, corresponding to an unstructured flexible loop (Bcl-xL [ΔTM209-233, Δ27-82]). Cells were grown in liquid medium LB (Luria-Bertani; Sambrook and Russel, 2001): 10 g/l tryptone, 5 g/l yeast extract, 10 g/l NaCl. The medium was sterilized by autoclaving it for 20 minutes at 121 °C and implemented with 100 µg/ml ampicillin, to which bacterial cells transformed with the pET15b vector are resistant. *E. coli* cells were grown over-night (ON) in LB medium containing 100 µg/ml ampicillin at 37°C in a suitable shaker. LB medium supplied with 100 µg/ml ampicillin was then inoculated with this over-night culture in the ratio 1:50. The culture was kept at 37°C in a suitable shaker and the growth was followed measuring the OD_{600} (Optical Dispersion at 600 nm, directly related to the amount of cells in suspension). The temperature was slowly decrease until 20°C when an OD_{600} of 0.6 was reached, then protein expression was induced by adding IPTG to a final concentration of 1 mM when an OD_{600} of 0.8 was reached. After induction, bacteria were growth over-night at 20 °C. Cells were then harvested by centrifugation (20 min, 4000 g, 4°C). The harvested cells were resuspended in buffer A: 20 mM Na_2HPO_4, 500 mM NaCl, 1mM β-mercaptoethanol, pH 8.0 supplemented with 3% of buffer B: 20 mM Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, 1mM β-mercaptoethanol, pH 8.0 and EDTA-free protease inhibitors and lysed with a “French Press” cell disrupter at 1000 psi. Cell debris were then removed from the soluble components extracted by centrifugation of the lysate (45 min, 12000 g, 4°C). After this centrifugation step, the supernatant was filtered using 0.22 µm filters before being loaded onto the column for the first step of the purification procedure.

Purification protocol

An AKTApurifier 10 with a multiple-wavelength detector was used for the chromatographic steps of the purification procedures. The chromatographic runs were followed at 280 nm and chromatograms were recorded using the UNICORN software (GE healthcare).

The filtered supernatant obtained after the mechanical lysis of the bacterial cells was loaded onto a 1 ml *HisTrap HP* (GE healthcare) affinity column, exploiting the presence of the $(His)_6$-tag on the protein. The column was equilibrated in 3% buffer B and the sample was loaded. The fusion protein was eluted with a linear gradient procedure increasing the concentrations of imidazole from 3% B to 100% B in 20 minutes.

The $(His)_6$-Bcl-xL protein was eluted with 500 mM imidazole, and since the protease thrombin is sensitive to the presence of imidazole, it was necessary to remove it with a desalting chromatographic step exploiting a HiPrep 26/10 desalting column. The proteolytic cleavage was then performed over-night at 4°C.

The sample was then concentrated by ultrafiltration, using Vivaspin concentrators. Then it was loaded onto a HiLoad 16/60 preparation grade Superdex 75 column equilibrated with 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 7.4, to eliminate possible aggregate forms and place the protein in the most appropriate buffer system for the following studies. The fractions corresponding to Bcl-xL were collected and stocked, without concentrate them, in aliquots of 1 ml. The aliquots were then flash-frozen using liquid nitrogen and conserved at -80 °C.

The purification procedure was followed by SDS-PAGE. Samples were heated for 5 minutes at 95 °C to denature them in the sample-loading buffer: 60 mM Tris pH 6.8, 10% glycerol, 2% SDS, 0.01% bromophenol blue, 2% β-mercaptoethanol. Then they were loaded with Low molecular weight See Blue® Plus2 Pre-Stained Standard (*Invitrogen)* markers on SDS-PAGE gels. A voltage of 180 V was used to performed the electrophoretic run with the gels mounted vertically on the electrophoretic apparatus and in contact with the running buffer: 25 mM Tris pH 8.3, 250 mM glycin, 0.1% SDS. The gels were colored for 20 minutes under soft shaking with the staining solution: 40% EtOH, 10% AcOH, 0.1% w/v Coomassie G-250 and then the destaining solution: 40% EtOH, 10% AcOH and a final wash in water were used to remove unspecific staining.

Protein crystallization

The vapour diffusion technique was exploited to perfrom crystallization trials manually in both sitting drop and hanging drop setups. The crystallization conditions for the protein in the apo form and in complex with ligands were inspired by the most frequent conditions found in the PDB. The protein was used in a concentration range of 10-20 mg/ml. Crystallization drops for the protein in the apo form were placed at 277 K, all the other drops were placed at 298 K. Protein and precipitants were mixed in equal ratio within the drop either 1 + 1 μl or 0.5 + 0.5 μl.

Crystallization trials on the apo form of the protein Bcl-xL [ΔTM209-233] were performed screening around the conditions of deposited structures, using the following precipitant solutions: Ammonium sulphate 1.4-2.2 M, Sodium citrate pH 5.0 or 5.5/Mes pH 6.0 or 6.5. The crystals obtained were also used for soaking experiments. In parallel, Co-crystallization trials were performed with the ligands 413, 622, 1777, 0160, 0200 and 0390 in various crystallization conditions. Co-crystallization trials were performed with ABT-737 adding it directly to the concentrated protein and also concentrating the protein in presence of the inhibitor. The protein:ligand ratio used was ABT-737 : protein = 1 : 1. Supplementary DMSO was add to the solution until the cut off of 4% in order to minimize the precipitation of the ligand in the protein buffer.

As regards the second construct (Bcl-xL [ΔTM209-233, Δ27-82]) co-crystallization trials were performed with ABT-737, and all the ligands 413, 622, 1777, 0160, 0200 and 0390 concentrating the protein in presence of the inhibitor or directly mixing the concentrated protein with the inhibitor. The protein:ligand ratio used was ABT-737 : protein = 1 : 1. Supplementary DMSO was add to the solution until the cut off of 4% in order to minimize the precipitation of the ligand in the protein buffer. Various crystallization conditions were tested.

Single Crystal X-ray Crystallography

After being cryoprotected with a solution composed by the reservoir solution supplemented with 25% ethylene glycol or 25% glycerol, crystals were flash-frozen in liquid nitrogen.

X-ray diffraction datasets were collected at the XRD1 beamline of the ELETTRA Synchrotron (Basovizza, TS), at the PX beamline at the Swiss Light Source Synchrotron (Villigen, CH) and at the ID30A-3 beamline at the ESRF European Synchrotron (Grenoble, F) at 100 K.

Diffraction data were initially processed (indexing, integration and scaling) with the XDS package (Kabsch, 2010). Integrated reflections were then reduced and merged with AIMLESS (Evans, 2006) (Evans, An introduction to data reduction: space group determination, scaling and intensity statistics, 2011) from the CCP4 suite (Winn, et al., 2011). The structures were solved by molecular replacement using the previously deposited models with the PDB ID 1R2D or 2YXJ as search model and the program PHASER (McCoy, Grosse-Kunstleve, Adams, Winn, Storoni, & Read, 2007). The initial structures were then refined alternating manual steps of real space refinement using the program COOT (Emsley & Cowtan, 2004) (Emsley, Lohkamp, Scott, & Cowtan, 2010) and automatic steps of reciprocal space refinement using the suite PHENIX (Adams, et al., 2010). PYMOL was used to make all the images and the superposition analysis.

Results and Discussion

Protein expression and purification

As described in the experimental procedures, two forms of Bcl-xL, two constructs of different length, were recombinantly produced in bacteria: Bcl-xL [ΔTM209-233] and Bcl-xL [ΔTM209-233, Δ27-82]. Both the constructs are lacking the trans membrane (TM) domain, and the short one also lack a disordered loop. After expression, both constructs were purified using chromatographic methods with a protocol consisting four steps: 1) Immobilized Metal-ion Affinity Chromatography – IMAC; 2) desalting; 3) proteolytic cleavage of the $(His)_6$-tag; 4) SEC (Size Exclusion Cromatography).

Bcl-xL [ΔTM209-233]

The first construct, Bcl-xL [ΔTM209-233] lacks only the C-terminal transmembrane domain. Figure 3.1-9 shows the chromatogram of the first chromatographic step.

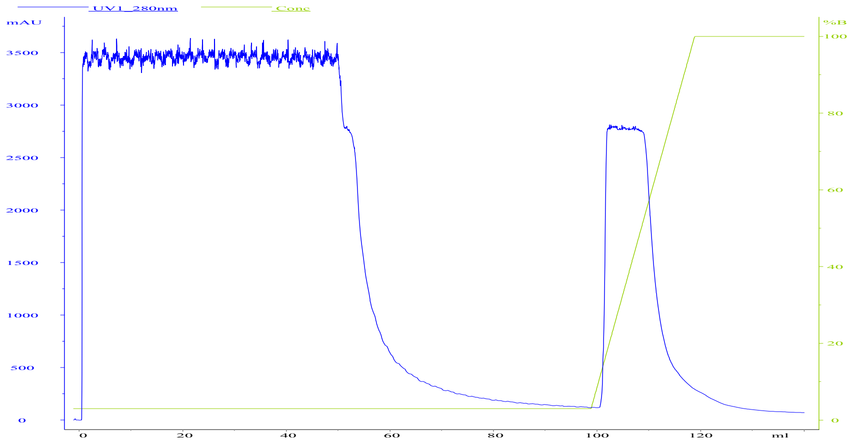


Figure 3.1 - 9 Bcl-xL [ΔTM209-233] IMAC chromatogram; in blue the elution profile (Abs at 280 nm), in green the linear gradient of imidazole. The large, out of scale, front band corresponds to the unbound bacterial lysate proteins ("flow-through").

Before the ON cleavage of the (His)$_6$-tag with thrombin, a desalting chromatographic step (Figure 3.1-10) was performed in order to place the protein in an appropriate buffer for the optimal activity of the protease.

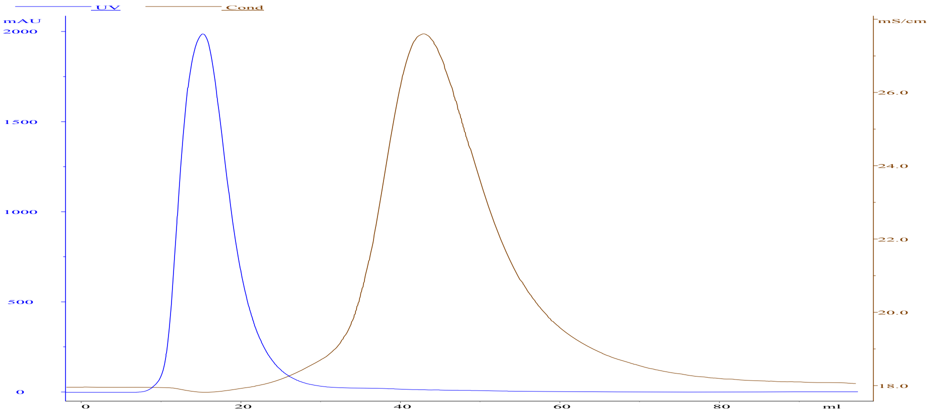


Figure 3.1 - 10 Bcl-xL [ΔTM209-233] desalting chromatogram; in blue the elution profile (Abs at 280 nm), in brown the conductivity profile.

The protein was further purified by size exclusion chromatography (Figure 3.1-11) step in order to remove aggregates and place the purified protein (fractions 23 to 29) in the appropriate buffer for storage and crystallization trials.

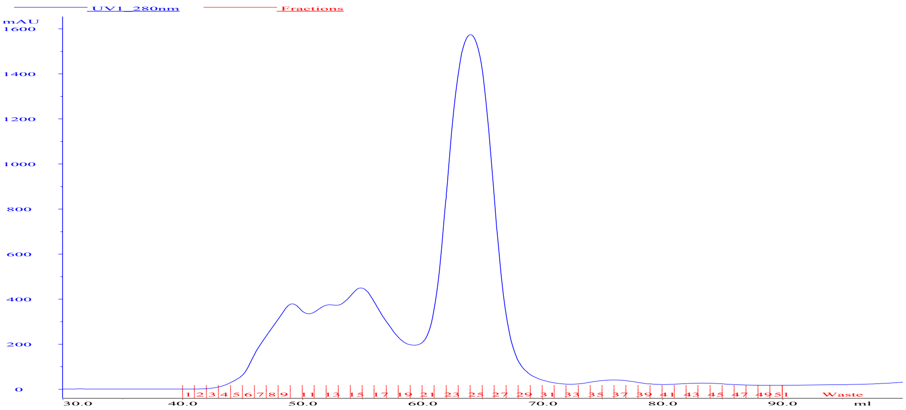


Figure 3.1 - 11 Bcl-xL [ΔTM209-233] SEC chromatogram.

Bcl-xL [ΔTM209-233, Δ27-82]

The second construct, Bcl-xL [ΔTM209-233, Δ27-82], lacks both the transmembrane domain ([ΔTM209-233]) and the flexible loop from residue 27 to residue 82 ([Δ27-82]). The same purification protocol was followed also for this second construct.

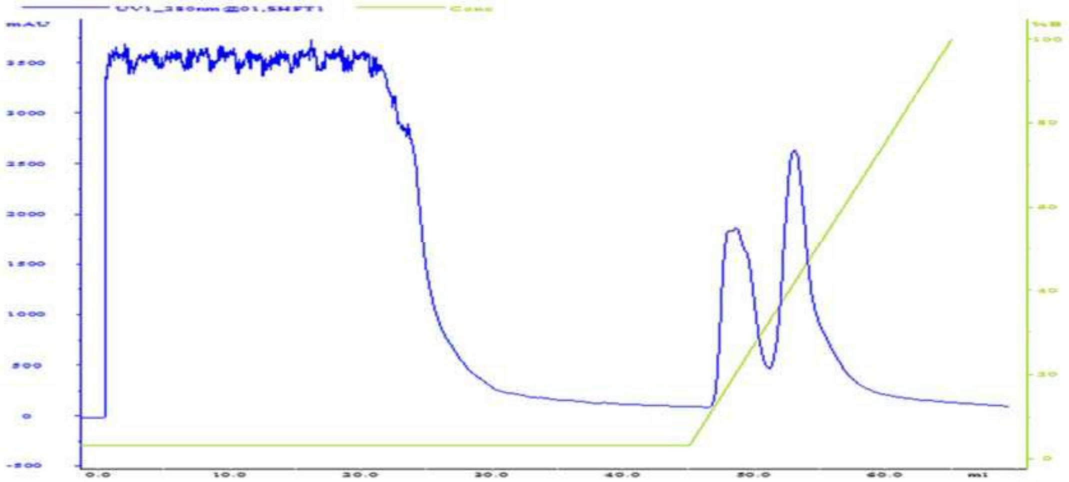


Figure 3.1 - 12 Bcl-xL [ΔTM209-233, Δ27-82] IMAC chromatogram; in blue the elution profile (Abs at 280 nm), in green the linear gradient of imidazole. The large, out of scale, front band corresponds to the unbound bacterial lysate proteins ("flow-through").

The two peaks in the IMAC step (Figure 3.1-12) were analyzed by SDS-PAGE before proceeding with the purification procedure, which continued on the second peak. The first peak corresponded to proteins with low affinity for the Ni^{2+} resin, for which the presence of 3% buffer B was not enough to avoid the binding. The second peak was therefore collected, buffered exchanged (Figure 3.1-13), cut ON with thrombin and further purified by SEC (Figure 3.1-14), where only fractions from 23 to 29 were collected for the crystallization trials.

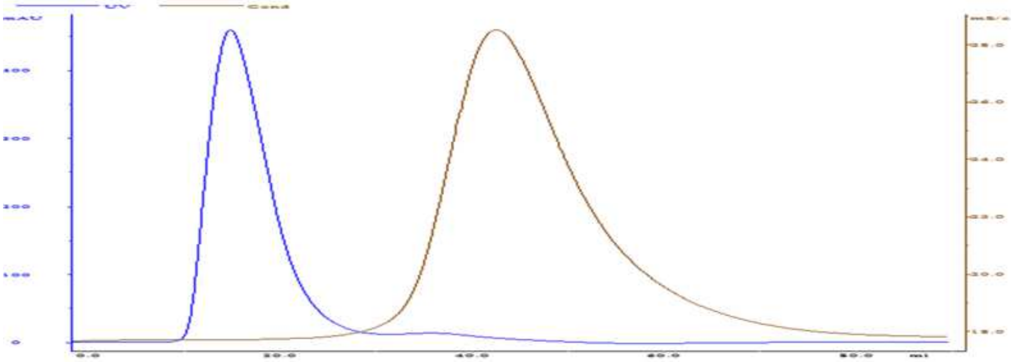


Figure 3.1 - 13 Bcl-xL [ΔTM209-233, Δ27-82] desalting chromatogram; in blue the elution profile (Abs at 280 nm), in brown the conductivity profile.

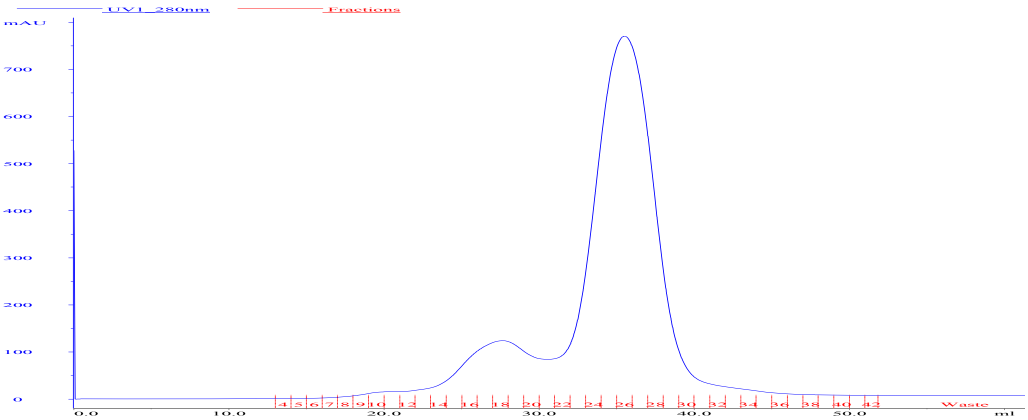


Figure 3.1 - 14 Bcl-xL [ΔTM209-233, Δ27-82] SEC chromatogram.

The whole purification was followed by SDS-PAGE. In figure 3.1-15 the two purified proteins are shown.

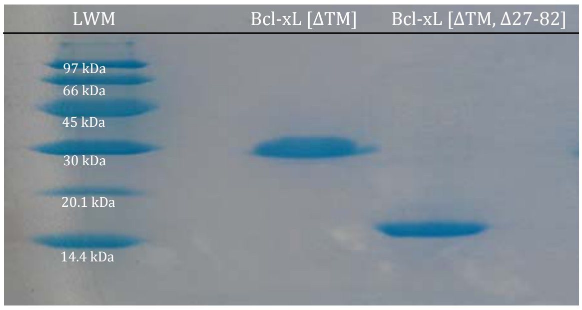


Figure 3.1 - 15 Coomassie stained SDS-page of the purified proteins at the end of the purification procedure.

Protein crystallization and single crystal x-ray crystallography

Bcl-xL [ΔTM209-233]

Already prom the purification procedure, the longer version of the protein seemed to be easier to manage, showing higher yields of expression and less tendency to aggregate upon concentration. This was confirmed also in the crystallization trials, where after an optimization phase of the crystallization conditions (figure 3.1-16), huge well-diffracting crystals were obtained.

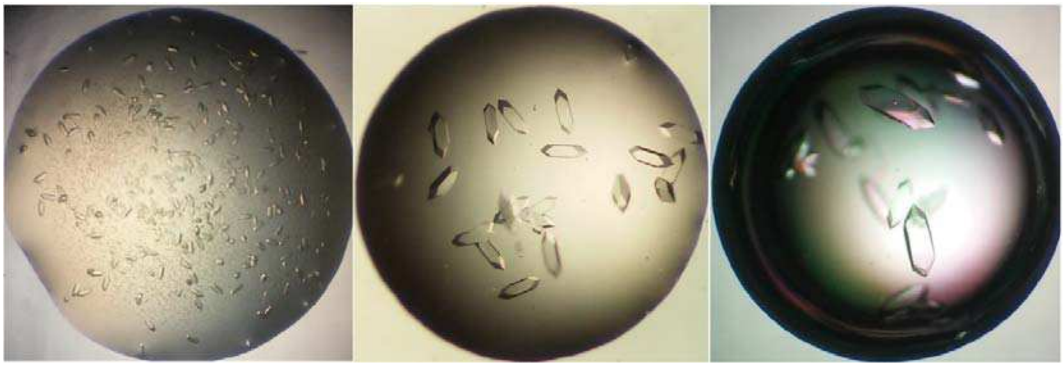


Figure 3.1 - 16 | optimization of Bcl-xL [ΔTM209-233] crystals.

After having optimized the purification and crystallization protocol, several crystals were obtained both in the apo-form and in the presence of the putative ligands provided by our collaborators (figure 3-1). Soaking experiments were also performed on the apo crystals. More than 30 diffraction datasets were collected on these crystals and several structures were solved but unfortunately, no electronic density imputable to the ligands was detected.

Data collection and refinement statistics for three representative structures (apo, one soaking and one co-crystallization) are reported in table 3.1-4.

	APO	Co-xtal_0160	soaking_1777
Data collection			
Wavelenght (Å)	0,979	0,9677	1
Space group	P4(1)2(1)2	P4(1)2(1)2	P4(1)2(1)2
Cell dimensions			
a b c (Å)	63,33 63,33 110,92	63,13 63,13 110,88	63,13 63,13 110,95
α β γ (°)	90 90 90	90 90 90	90 90 90
Resolution (Å)	44,78-1,96	44,64-1,75	44,64-1,45
R_{sym}	0,060 (0,487)	0,180 (2,086)	0,051 (1,294)
R_{meas}	0,060 (0,504)	0,180 (2,153)	0,053 (1,383)
$< I/\sigma I >$	30,8 (4,7)	12,9 (1,5)	25,7 (1,5)
Completeness (%)	99,5 (95,7)	99,2 (95,4)	99,9 (98,1)
Multiplicity	15,5 (15,0)	16,8 (15,8)	12,4 (8,0)
Refinement			
Resolution (Å)	44,78-1,96	44,64-1,75	44,63-1,45
Number of reflections	16543	23134	40436
R_{work}/R_{free}	20,1/22,3	19,4/19,8	19,6/20,3
B-factors (Å^2)			
protein	39,0	26,8	26,9
ligands	/	/	/
r.m.s. deviations			
bond lenght (Å)	0,003	0,007	0,006
bond angles (°)	0,667	0,983	1,021
Ramachandran plot			
favoured (%)	97,84	97,14	98,58
outliers (%)	0	0	0

Table 3.1 - 4 Data collection and refinement statistics on the three dataset presented. The apo form and the soaking 1777 were collected at the PX beamline, Swiss Light Source Syncrotron, CH. Co-xtal 0160 were collected at the ID30A-3, ESRF European Syncrotron, F.

No electronic density for the ligands was detected and only small conformational changes in the binding site region were observed. In figre 3.1-17 the superposition of the three structures is reported.

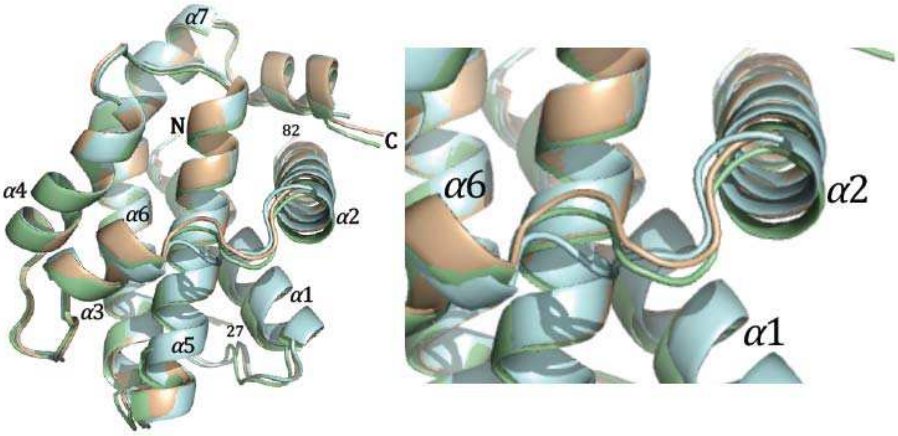


Figure 3.1 - 17 Superposition between the Bcl-xL [ΔTM209-233] apo form (cyan), soaking 1777 (green) and co-xtal 0160 structures (orange).

Bcl-xL [ΔTM209-233, Δ27-82]

We then decided to move to the shorter version of the protein, despite the fact that it was more difficult to obtain at high yields required for the crystallization trials, since deposited structure in complex with ligands on the PDB, have been solved using this short construct. Firstly, we established a solid protocol for purification, then we moved to crystallization trials. We decided to obtain crystals with Bcl-Xl [ΔTM209-233, Δ27-82] in complex with ABT-737, a known inhibitor of the protein, to have a control of the crystallization properties of the short constructs, which does not crystallize in the apo form.

After several trials, crystals of Bcl-Xl in complex with ABT-737 (figure 3.1-18) were obtained.

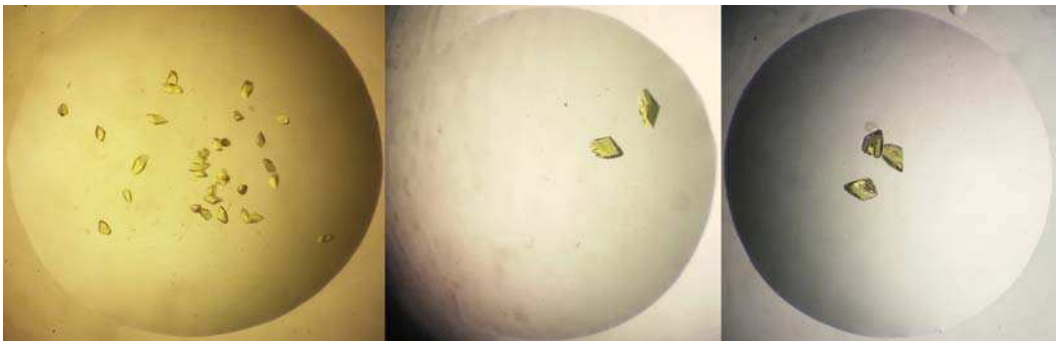


Figure 3.1 - 18 Bcl-Xl [ΔTM209-233, Δ27-82] crystals

Diffraction data were collected on the crystals; data collection and refinement statistics are summarized in table 3.1-5.

	ABT737
Data collection	
Wavelenght (Å)	1
Space group	C2
Cell dimensions	
a b c (Å)	151,855 68,57 76,58
α β γ (°)	90 100 90
Resolution (Å)	46.43-1.89 (1.93-1.89)
R_{sym}	0,065(0,331)
R_{meas}	0,078 (0,397)
$CC_{(1/2)}$	0,998 (0,856)
< I/σI >	11,3 (3.6)
Completeness (%)	97,7 (83,3)
Multiplicity	3,4 (3,1)
Refinement	
Resolution (Å)	46,43-1,89
Number of reflections	59575
R_{work}/R_{free}	24,42/28,68
B-factors (Å²)	
protein	41,6
ligands	37,17
r.m.s. deviations	
bond lenght (Å)	0,011
bond angles (°)	1,256
Ramachandran plot	
favoured (%)	97,46
outliers (%)	0,36

Table 3.1 - 5 data collection and refinement statistics

With this dataset it was possible to solve the structure of Bcl-Xl in complex with ABT-737. In comparison with the deposited PDB structure (2YXJ), mine crystallized in a different space group, with the protein being a domain-swapped dimer (figure 3.1-19).

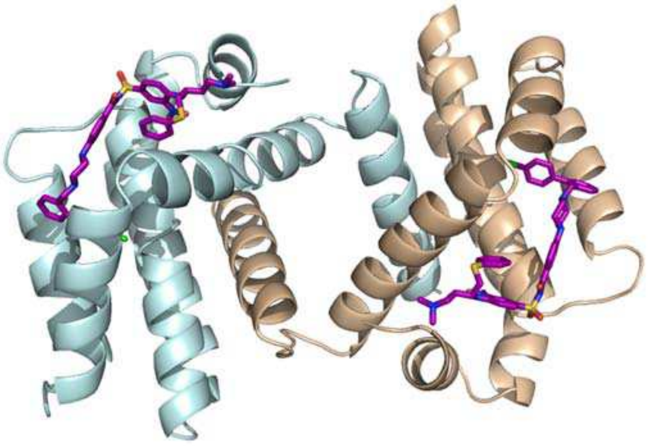


Figure 3.1 - 19 Domain-swapped dimer of Bcl-Xl. The two proteins exchange the α1 helix forming the dimer. In purple, the two molecules of ABT-737.

ABT-737 binds in the same hydrophobic pocket as in the 2YXJ structure, as shown in figure 3.1-20.

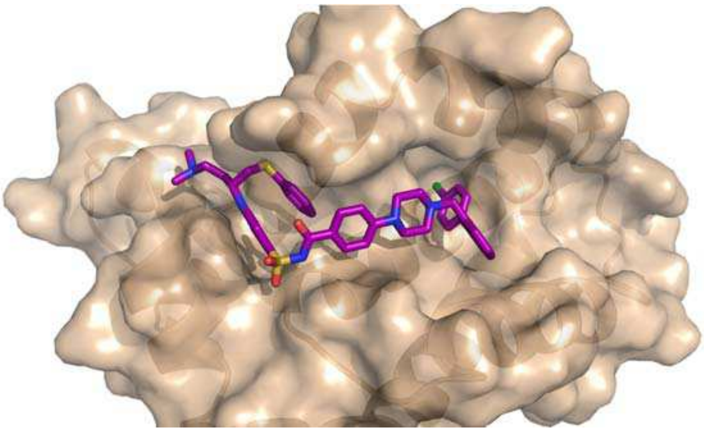


Figure 3.1 - 20 ABT-737 bound to Bcl-Xl

Figure 3.1-21 shows the refined $2F_O$-F_C map of the molecules of ABT-737 contoured at 1 σ.

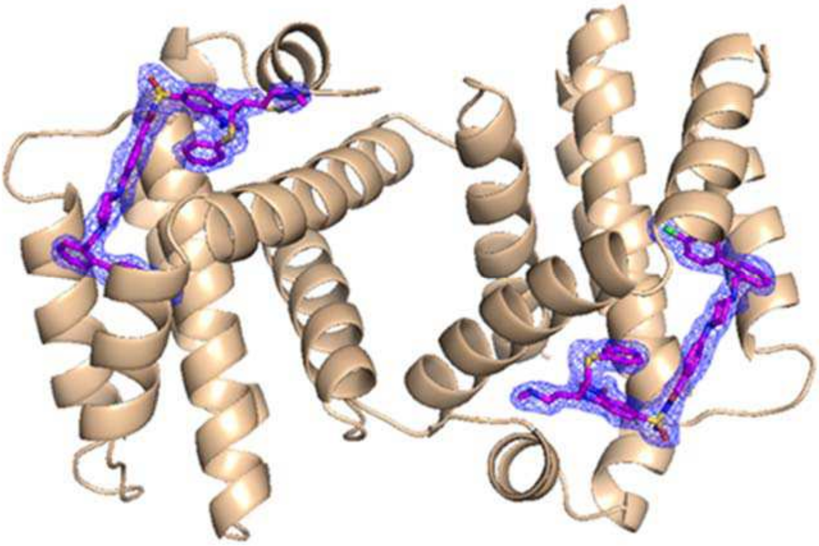


Figure 3.1 - 21 $2F_O$-F_C map of the molecules of ABT-737 contoured at 1 σ

The four ABT molecules establish similar contacts with the Bcl-Xl molecules, that include hydrophobic interactions, hydrogen bonds, π-π and π-cation interactions as shown in figure 3.1-22.

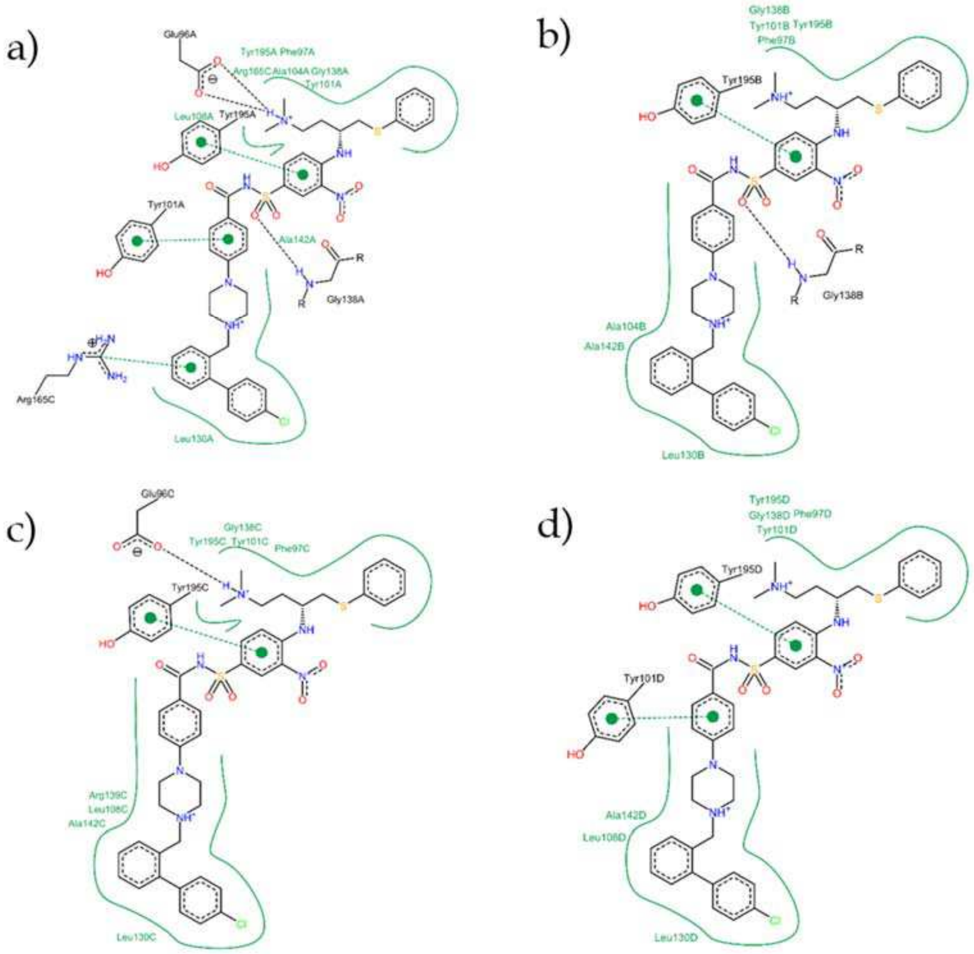


Figure 3.1 - 22 Black dashed lines indicate hydrogen bonds. Green solid line show hydrophobic interactions and green dashed lines show π-π and π-cation interactions.

After managing to set up a solid crystallization protocol for the short construct of Bcl-Xl, which does not crystallize in the apo form, I have set up crystallization trials with the molecules provided by our collaborators. So far, I have not obtained crystals probably due to the low affinity and/or solubility of the molecules.

Conclusions

Two constructs of the anti-apoptotic protein Bcl-xL were expressed and purified. The longer version, Bcl-xL [ΔTM209-233] showed high production yields and good crystallization reproducibility. Several structures were solved both in the apo form and in presence of the ligands that we are studying. No electronic density for the ligands was however found.

On the other hand, the shorter construct Bcl-xL [ΔTM209-233, Δ27-82] was more difficult to produce and crystallize. Crystals grew in presence of the known inhibitor ABT-737 and the structure of the complex was solved.

After managing to set up a solid crystallization protocol for the short construct of Bcl-Xl, which does not crystallize in the apo form, crystallization trials with the molecules provided by our collaborators were performed. So far crystals have not been obtained. This is probably due to the low affinity and/or solubility of the molecules. It will be necessary to improve them by in silico guided analysis, adding functional groups that could possibly enhance their affinity and solubility, before proceeding with further crystallization trials.

3.2 CASQ

Introduction

Calsequestrin (CASQ)

Calsequestrin (CASQ) is a 45 kDa protein able to bind Ca^{2+} with low affinity and high capacity. It is responsible for Ca^{2+} storage in the lumen of the sarcoplasmic reticulum (SR) of striated muscles in relationship with activation of muscle contraction. Two genes encode for two isoforms of CASQ: CASQ1 is expressed in fast skeletal fibers and CASQ2 in cardiac muscle and slow skeletal fibers (Schiaffino and Reggiani, 2011). In patients suffering with vacuolar aggregate myopathy, a muscular disorder characterized by the presence of vacuoles containing aggregated proteins of the SR, a mutation in *CASQ1* was found (Tomelleri et al., 2006; Rossi et al., 2014). CASQ1 can be present as a monomer, or it can undergo polymerization in Ca^{2+}- dependent manner. Low Ca^{2+} concentrations cause the formation of dimes by favoring front-to-front interactions between two adjacent monomers. Higher Ca^{2+} concentrations cause the formation of polymers, mediated by back-to-back interactions (Park et al., 2003).

A large macromolecular machinery controls Ca2+ release. CASQ1 interacts with the Stromal Interaction Molecule 1 (STIM1), localized on the SR (Wang et al., 2015; Zhang et al., 2016). The function of STIM1, together with the Ca^{2+} release-activated Ca^{2+} modulator 1 (ORAI1) on the plasma membrane, is to allow the replenishing of depleted intracellular Ca^{2+} store. This mechanism is known as Store-Operated Ca^{2+} Entry (SOCE) and CASQ1 may have an inhibitory effect on the pathways that regulate it (Shin et al., 2003; Zhao et al., 2010).

In patients affected by tubular aggregate myopathy (TAM), mutations in *STIM1* and *ORAI1* have recently been identified (Böhm et al., 2013, Böhm et al., 2017; Endo et al., 2015; Okuma et al., 2016). TAM is a progressive condition characterized by cramps, musclular weakness, myalgia and the presence of tubular aggregates in muscle fibers. These aggregates originate from the SR membrane (Engel et al., 1970) as a result of processes not fully understood. They contain various SR proteins including, among others, sarco-endoplasmic reticulum Ca^{2+} ATPase (SERCA), triadin and CASQ1

(Salviati et al., 1985; Chevessier et al., 2005; Boncompagni et al., 2011). It has been proposed that these tubular aggregates result from the aggregation of SR-derived membranes possibly after misfolding and aggregation of SR proteins caused by hereditary or acquired conditions (Kuncl et al., 1989; Chevessier et al., 2005; Agbulut et al., 2000; Schiaffino 2012; Giacomello et al., 2015).

We have recently published the identification and characterization of three mutations in the *CASQ1* gene in patients suffering with TAM in the in collaboration with the group of prof. Sorrentino (University of Siena). The characterization of these mutated variants of CASQ1 indicated that they have altered Ca^{2+}-dependent polymerization and reduced ability to store Ca^{2+} (Barone et al., 2017).

Experimental procedures

Dynamic Light Scattering

Dynamic light scattering (DLS) was used to analyze the Ca^{2+}-dependent aggregation of the proteins, using a Zetasizer Nano™ (Malvern Instruments) and a low volume quartz cuvette.

Prior to data collections, the samples were centrifuged at 12000 rpm for 10 minutes and filtered using 0.45 µm filters. Measurements were performed at 25 °C using a protein concentration of 3 mg/ml in 20 mM Tris-HCl, 300 mM KCl, pH 7.4, with the following Ca^{2+} concentrations: 0, 3, 5 and 10 mM. 10 mM EGTA was added at the sample with 10 mM Ca^{2+} in order to evaluate the reversibility of the aggregation process.

Structural analysis

Structural analysis of CASQ1 mutations has been performed based on the crystal structure of hCASQ1 (PDB code: 3UOM) using the program PYMOL.

Results and Discussion

These results are part of the article titled "Identification and characterization of three novel variants in the CASQ1 gene in four patients with tubular aggregate myopathy" that we have recently published (Barone et al, 2017).

DLS measurements on were performed on the WT protein and mutants found in patients suffering from tubular aggregate myopathy in order to see the different calcium dependent aggregation properties of these mutated forms in respect to the wild-type protein found in healthy subjects. The interaction between calsequestrin and calcium is fundamental for the function of the protein and so is the understanding of the different calcium-associated behaviors of the mutated forms of CASQ found in the patients.

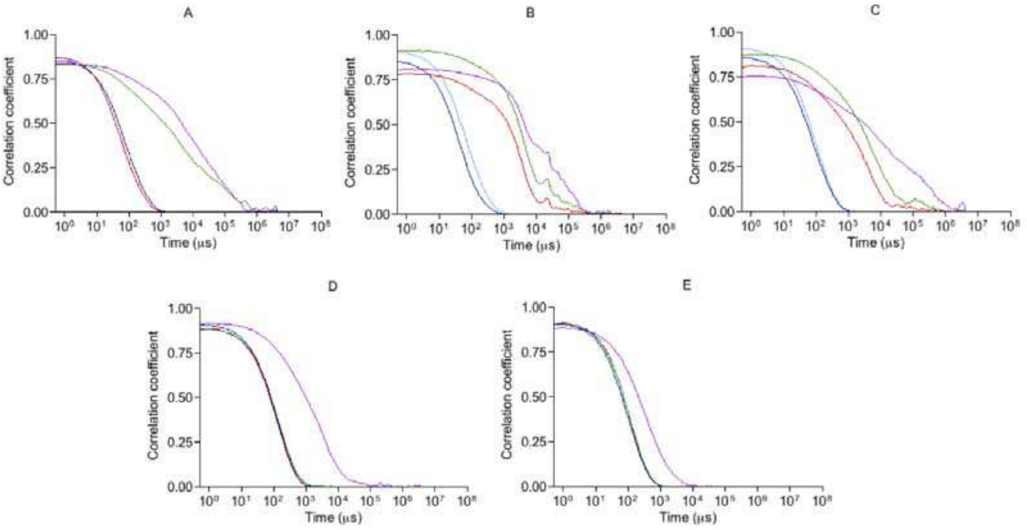


Figure 3.2 - 1 DLS measured correlation curves at increasing CaCl2 concentrations (cyan 0, red 3, green 5 and magenta 10 mM calcium) for wt-CASQ1 (A), mutant D244G (B), mutant I385T (C), mutant G103D (D) and mutant D44N (E). In absence of calcium the correlograms are very similar for the five samples, indicating a similar initial state. In all cases the aggregation process induced by calcium is reversible as indicated by the effect of the addition of 10 mM EGTA (blue).

As shown in figure 3.2-1, the wild-type protein did not appreciably change its aggregation state between 0 and 3 mM Ca^{2+}. At 5 mM Ca^{2+}, the signal revealed a more aggregated sample and at 10 mM Ca^{2+} the sample appeared even more aggregated. The D244G variant had the highest tendency to aggregate, showing aggregation already at 3 mM Ca^{2+}, with the correlogram shifted to higher correlation times. Variant I385T showed a similar behavior, starting aggregating from 3 mM Ca^{2+} on. In contrast, variants G103D and D44N did not change their aggregation state up to 5 mM Ca^{2+} and were found aggregated only at 10 mM Ca^{2+}. With all proteins, aggregation was fully reversed by the addition of EGTA.

These results, along with a structural analysis based on the positions of the mutations in the protein, allowed to rationalize at structural level the effects of the mutations on the aggregation properties and on the severity of the onset of the disease. Using the available crystal structures of WT and D244G CASQ1 (Wang et al., 1998; Lewis et al., 2015), we noticed the three CASQ1 mutations identified in TAM patients are located at sites that can affect the interactions between CASQ1 subunits.

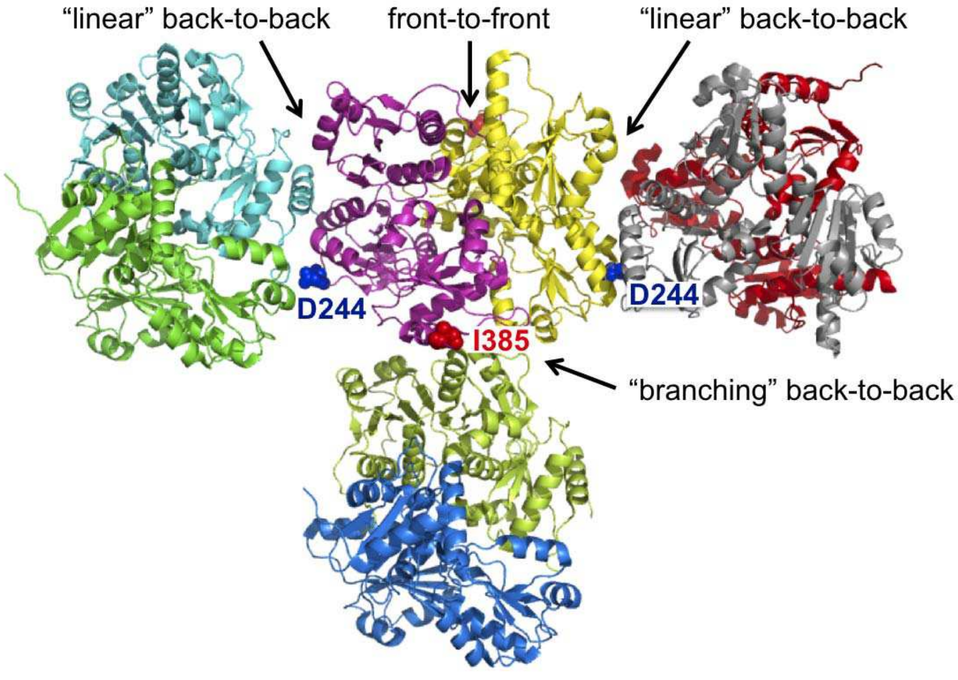


Figure 3.2 - 2 Proposed model of polymerization based on the crystal structure of CASQ1 in presence of high calcium (PDB-ID 3UOM). A dimer of CASQ1 held together by front-to-front interactions (yellow and magenta monomers) can interact with other dimers both via "linear" back-to-back contacts (dimers green/cyan and red/grey) and via "branching" back-to-back contacts (dimer blue/olivegreen), originating a branched polymer. Mutation in position 244 (blue spheres for D244) interferes with linear back-to-back contacts while in position 385 (red spheres for I385) with branching back-to-back contacts.

Initial studies with the D244G mutation (Lewis et al., 2015) indicated that it may decrease the association of dimers through linear back-to-back interactions. As shown in figure 3.2-2, the I385T mutation is located at the interface of "lateral" back-to-back contacts that are responsible for the formation of the branched polymers formed by WT CASQ1 (Perni et al., 2013). Moreover, the negatively charged C-terminal tail, immediately downstream of I385, has been shown to be important to hamper the formation of not-specific polymerization products (Park et al., 2004; Beard and Dulhunty, 2015). Hence, it can be hypothesized that the I385T mutation favors the assembly of atypical, not-fully functional, polymeric forms. On the other hand, residues 44 and 103 are positioned at the interface responsible for the formation of CASQ1 dimers (figures 3.2-2 and 3.2-3).

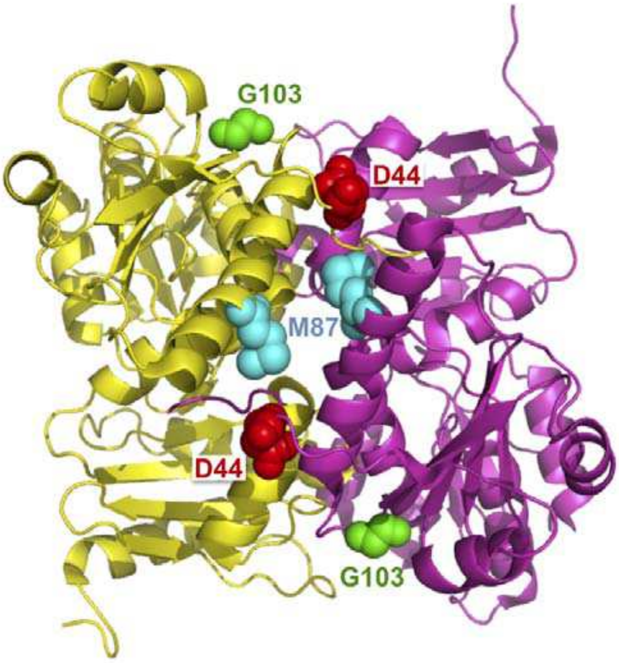


Figure 3.2 - 3 Quaternary assembly of the CASQ1 dimer. One molecule (yellow) interacts with the partner (magenta) through an extended region (around 2430 Å2 interface area) forming the so-called front-to-front contact. Residues studied in this paper are shown as sphere, D44 in red, M87 in cyan and G103 in green. All belong to the front-to-front interaction surface.

Accordingly, the D44N and G103D mutations, can affect the interactions that strengthen the front-to-front dimerization of the protein, resulting in dimers less efficient to support the formation of back-to-back tetramers, similarly to what proposed for the M87T mutation (Lewis et al., 2015, figure 3.2-3). It can be noted that the CASQ1 mutations that have increased polymerization (D244G and I385T) are located at sites of back-to-back interactions, while CASQ1 mutations that have a reduced polymerization (D44N and G103D) are positioned at sites of front-to-front interactions.

Final considerations

Single crystal x-ray crystallography is a powerful technique to study protein-protein and protein-ligand interactions.

The anomalous signal, in addition its utility in solving the phase problem, can be exploited to obtain fine details about the atomic structure , as in the case of the anions bound to the STAS domain of prestin, and the placing of the bromine atoms of the CDK2 inhibitor K137.

Single crystal x-ray crystallography is also used to screen potential ligands, coupled with in silico docking procedures. However, as in the case of the allosteric inhibitors of CDK2, for which binding constants were also obtained in the ANS displacement test, the affinity proved too low to obtain crystals of the complexes. The same happened with the in silico-screened ligands for BCL-XL. Indeed the study of protein-protein and protein-ligands interactions is a complex field that requires a multi-disciplinar approach.

BIBLIOGRAPHY

PART 1

Adams P. D., Afonine P. V., Bunkóczi G., Chen V. B., I. Davis W., Echols N., Headd J. J., Hung L.-W., Kapral G. J., Grosse-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C. and Zwart P. H., "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Cryst.* **D66**: 213-221 (2010).

Aravind L., Koonin E.V., "The STAS domain—a link between anion transporters and antisigma- factor antagonists." *Curr Biol.* **10**: R53–5 (2000).

Ashmore J.F., "A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier." *J. Physiol.* **388**: 323-347 (1987).

Babu, M. *et al.* "Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism." *Structure* **18**: 1450–1462 (2010).

Birke AS. & Javelle A. "Prestin and the good vibrations" *Biochem J.* **473**: 2425–2427 (2016).

Brennich, M. E., Kieffer, J., Bonamis, G., De Maria Antolinos, A., Hutin, S., Pernot P., and Round, A. "Online data analysis at the ESRF BioSAXS beamline." *J. Appl. Cryst.* **49**: 203-212 (2016).

Dallos P. "The active coclea." *J. Neurosci.* **2**: 4575-4585. (1992).

Dallos P., Zheng J., Cheatham M.A. "Prestin and the cochlear amplifier." *J. Physiol.* **76**: 37-42 (2006).

Dallos, P. & Fakler, B. "Prestin, a new type of motor protein." *Nat. Rev. Mol. Cell Biol.* **3**: 104–111 (2002).

Dawson, P. A. & Markovich, D. "Pathogenetics of the human SLC26 transporters." *Curr. Med. Chem.* **12**: 385–396 (2005).

De Maria Antolinos, A., Pernot, P., Brennich, M. E., Kieffer, J., Bowler, M. W., Delageniere, S., Ohlsson, S., Malbet Monaco, S., Ashton, A., Franke, D., Svergun, D., McSweeney, S., Gordon, E., and Round, A. "ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments." *Acta Crystallogr. D Biol. Crystallogr.* **71**: 76-85 (2015).

Dorwart, M. R., Shcheynikov, N., Yang, D. & Muallem, S. "The solute carrier 26 family of proteins in epithelial ion transport." *Physiology (Bethesda)* **23**: 104–114 (2008).

Emsley P. and Cowtan K. "Coot: model-building tools for molecular graphics." *Acta Crystallographica* **D60**: 2126-2132 (2004).

Emsley P., Lohkamp B., Scott W.G. and Cowtan K. "Features and Development of Coot." *Acta Crystallographica* **D66**: 486-501 (2010).

Evans P.R., "An introduction to data reduction: space-group determination, scaling and intensity statistics" *Acta Crystallographica* **D67**: 282-292 (2011).

Evans P., "Scaling and assessment of data quality." *Acta Crystallographica* **D62**: 72–82 (2005).

Fischer H., de Oliveira Neto M., Napolitano H.B., Craievich A.F., Polikarpov I. "The molecular weight of proteins in solution can be determined from a single SAXS measurement on a relative scale." **43**: 101-109 (2010).

Franke, D. and Svergun, D.I. "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." *J. Appl. Cryst.*, **42**: 342-346 (2009).

Géléoc G.S.G., Holt J.R., "Auditory amplification: outer hair cells pres the issue." *Trends Neurosci.* **26**: 115-117 (2003).

Greetsma E.R., Chang Y.N., Shaik F.R., Neldner Y., Pardon E., Steyaert J., Dutzler R., "Structure of a prokaryotic fumarate transporter reveals the architecture of the SLC26 family." *Nat Struct Mol Biol.* **10**: 803-8 (2015).

Gorbunov D., Sturlese M., Nies F., Kluge M., Bellanda M., Battistutta R., Oliver D. "Molecular architecture and the structural basis for anion interaction in prestin and SLC26 transporters." *NatCommun.* **5**: 3622-3634 (2014).

Hediger M.A., Romero M.F., Peng J.B., Rolfs A., Takanaga H., Bruford E.A. "The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins—Introduction." *Pflugers Arch* **447**: 465–468 (2004).

Kabsch W., "XDS", *Acta Cryst.* **D66**: 125-132 (2010).

Keller J.P., Homma K., Duan C., Zheng J., Cheatham M.A., Dallos P., "Functional Regulation of the SLC26-Family Protein Prestin by Calcium/Calmodulin" *The Journal of Neuroscience* **34(4)**: 1325–1332 (2014).

Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J., Svergun D.I. "PRIMUS - a Windows-PC based system for small-angle scattering data analysis." *J. Appl. Cryst.* **36**: 1277-1282 (2003).

Konarev P.V., and Svergun D.I. "A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems." *IUCr J.* **2**: 352-360 (2015).

Liu X.Z., Ouyang X.M., Xia X.J., Zheng J., Pandya A., Li F., Du L.L., Welch K.O., Petit C., Smith R.J.H., Webb B.T., Yan D., Arnos K.S., Corey D., Dallos P., Nance W.E., Chen Z.Y., "Prestin, a cochlear motor protein, is detective in non-syndromic hearing loss." *Hum. Mol. Genet.* **12**: 1155-1162 (2003).

Lolli G., Pasqualetto E., Costanzi E., Bonetto G., Battistutta R. "The STAS domain of mammalian SLC26A5 prestin harbours an anion-binding site." *Biochem J*, **473(4)**: 365-70 (2016).

Ludwig J., Oliver D., Frank G., Klöcker N., Gummer A.W., Fakler B., "Reciprocal electromechanical properties of rat prestin: The motor molecule from rat outer hair cells." *Proc. Natl. Acad. Sci. USA* **98**: 4178–4183 (2001).

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. "Phaser crystallographic software." *J. Appl. Cryst* **40**: 658-674 (2007).

Mount D.B., Romero M.F. "The SLC26 gene family of multifunctional anion exchangers." *Pflugers Arch.* **447**: 710–721 (2004).

Oliver D., He D.Z., Klöcker N., Ludwig J., Schulte U., Waldegger S., Ruppersberg J.P., Dallos P., Fakler B. "Intracellular anions as the voltage-sensor of prestin, the outer hair cell motor protein." *Science* **292**: 2340–2343 (2001).

Pasqualetto E., Aiello R., Gesiot L., Bonetto G., Bellanda M., Battistutta R. "Structure of the cytosolic portion of the motor protein prestin and functional role of the STAS domain in SLC26/SulP anion transporters." *J. Mol. Biol.* **400**: 448–462 (2010).

Pernot P., Round A., Barrett R., De Maria Antolinos A., Gobbo A., Gordon E., Huet J., Kieffer J., Lentini M., Mattenet M., Morawe C., Mueller-Dieckmann C., Ohlsson S., Schmid W., Surr J., Theveneau P., Zerrad L., McSweeney S. "Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution." *J. Synchrotron Radiat.* **20**: 660-664 (2013).

Petoukhov M.V., Svergun D.I., "Global rigid body modelling of macromolecular complexes against small-angle scattering data." *Biophys. J.,* **89**: 1237-1250 (2009).

Petoukhov M.V., Franke D., Shkumatov A.V., Tria G., Kikhney A.G., Gajda M., Gorba C., Mertens H.D., Konarev P.V., Svergun D.I. "New developments in the ATSAS program package for small-angle scattering data analysis." *J. Appl. Cryst.* **45**: 342-350 (2012).

Povey S., Lovering R., Bruford E., Wright M., Lush M., Wain H. “The HUGO Gene Nomenclature Committee (HGNC).” *Hum Genet* **109**: 678–680 (2001).

Price G.D., Howitt S.M., “Topology mapping to characterize cyanobacterial bicarbonate transporters: BicA (SulP/SLC26 family) and SbtA”, *Mol Membr Biol*, **31(6)**: 177–182 (2014).

Rambo R.P., Tainer J.A. “Accurate assessment of mass, models and resolution by small-angle scattering.” *Nature* **496**: 477-481 (2013).

Saier M.H. Jr, Eng B.H., Fard S., Garg J., Haggerty D.A., Hutchinson W.J., Jack D.L., Lai E.C., Liu H.J., Nusinew D.P., Omar A.M., Pao S.S., Paulsen I.T., Quan J.A., Sliwinski M., Tseng T.T., Wachi S., Young G.B. “Phylogenetic characterization of novel transport protein families revealed by genome analyses.” *Biochim Biophys Acta* **1422**: 1–56 (1999).

Sambrook J., Russel D.W. “Molecular Cloning. A Laboratory Manual.” *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. U.S.A.* (2001).

Santos-Sacchi J. “Reversible inhibition of voltage-dependent outer hair cell motility and capacitance.” *J. Neurosci.* **11**: 3096–3110 (1991).

Schaechinger T.J., Oliver D. “Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers.” *Proc Natl Acad Sci USA* **104**: 7693–7698 (2007).

Schlessinger A. Matsson P., Shima J.E., Pieper U., Yee S.W., Kelly L., Apeltsin L., Stroud R.M., Ferrin T.E., Giacomini K.M., Sali A. "Comparison of human solute carriers." *Protein Science* **19**: 412–428 (2010).

Sharma A.K., Ye L., Baer C.E., Shanmugasundaram K., Alber T., Alper S.L., Rigby A.C. "Solution Structure of the Guanine Nucleotide-binding STAS Domain of SLC26-related SulP Protein Rv1739c from Mycobacterium tuberculosis." *J. Biol. Chem.* **286**: 8534–8544 (2011).

Sharma A.K., Rigby A.C., Alper S.L. "STAS domain structure and function." *Cell. Physiol. Biochem.* **28**: 407–422 (2011).

Shehata W.E., Brownell W.E., Dieler R. "Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pig cochlea." *Acta Otolaryngol.* **111**: 707–718 (1991).

Svergun D.I. "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing." *Biophys J.* **76(6)**: 2879-2886 (1999).

Tidow H., Nissen P. "Structural diversity of calmodulin binding to its target sites" *FEBS J.*, **280**: 5551-5565 (2013).

Toth T., Deak L., Fazakas F., Zheng J., Muszbek L., Sziklai I. "A new mutation in the human *pres* gene and its effect on prestin function." *Int. J. Mol. Med.* **20**: 545-550 (2007).

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., Wilson K.S. "Overview of the CCP4 suite and current developments." *Acta Crytsallographica* **D67**: 235-242 (2011).

Zheng J., Long K.B., Shen W., Madison L.D., Dallos P. "Prestin topology: localization of protein epitopes in relation to the plasma membrane." *NeuroReport* **12**: 1929-1935 (2001).

Zheng J., Madison L.D., Oliver D., Fakler B., Dallos P. "Prestin is the motor protein of cochlear outer hair cells." *Nature* **405**: 149–155 (2000).

Zheng, J. Du G.G., Matsuda K., Orem A., Aguiñaga S., Deák L., Navarrete E., Madison L.D., Dallos P. "The C-terminus of prestin influences nonlinear capacitance and plasma membrane targeting." *J. Cell. Sci.* **118**: 2987–2996 (2005).

PART 2

Ahn N. G., Resing K.A. “Toward the phospho-proteome.” *Nat. Biotechnol.* **19**: 317-318 (2001).

Cohen P. “The regulation of protein function by multisite phosphorylation –a 25 year update.” *Trends Bio-chem. Sci.* **25**: 596-601 (2000).

Hanks S.K., Hunter T. “The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.” *FASEB J.* **9**: 576-596 (1995).

Krebs E.G. “An accidental biochemist.” *Annu. Rev. Biochem.* **67**: xii-xxxii (1998).

Kornev A.P., Haste N.M., Taylor S.S., Ten Eyck L.F. “Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.” *Proc. Natl Acad. Sci. USA* **103(47)**: 17783–17788 (2006).

Kornev A.P., Taylor S.S., Ten Eyck L.F. “A helix scaffold for the assembly of active protein kinases.” *Proc Natl Acad Sci USA.* **105**: 14377-14382 (2008).

Manning G., Whyte D.B., Martinez R., Hunter T., Sudarsanam S. “The Protein Kinase Complement of the Human Genome.” *Science* **298(5600)**: 1912-1934 (2002).

Manning G., Plowman G. D., Hunter T., Sudarsanam S. “Evolution of protein kinase signaling from yeast to man.” *Trends Biochem. Sci.* **27**: 514-520 (2002 b).

Taylor S.S., Kornev A.P. “Protein kinases: evolution of dynamic regulatory proteins.” *Trends Biochem. Sci.* **36**: 65-77 (2011).

Taylor S.S., Keshwani M.M., Steichen J.M., Kornev A.P. “Evolution of the eukaryotic protein kinases as dynamic molecular switches.” *Phil. Trans. R. Soc. B* **367**: 2517-2528 (2012).

Walsh D.A., Perkins J.P., Krebs E.G. “An adenosine 3′,5′-monophosphate-dependant protein kinase from rabbit skeletal muscle.” *J. Biol. Chem.* **243**: 3763-3765 (1968).

PART 2.1

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis W.I., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., Zwart P.H., "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Cryst.* **D66**: 213-221 (2010).

Ali S., Heathcote D.A., Kroll S.H., Jogalekar A.S., Scheiper B., Patel H., Brackow J., Siwicka A., Fuchter M.J, Periyasamy M., "The development of a selective cyclin-dependent kinase inhibitor that shows antitumor activity." *Cancer Res.* **69**: 6208–6215 (2009).

Battistutta R., Cozza G., Pierre F., Papinutto E., Lolli G., Sarno S., O'Brien S. E., Siddiqui-Jain A., Haddach M., Anderes K., Ryckman D. M., Meggio F., Pinna L. A., "Unprecedented selectivity and structural determinants of a new class of protein kinase CK2 inhibitors in clinical trials for the treatment of cancer." *Biochemistry* **50(39)**:8478-88 (2011).

Barrett C.P., Noble M.E., "Molecular motions of human cyclin-dependent kinase 2." *J. Biol. Chem.* **280**: 13993–14005 (2005).

Betzi S., Alam R., Martin M., Lubbers D.J., Han H., Jakkaraj S.R., Georg G.I., Schonbrunn E., "Discovery of a potential allosteric ligand binding site in CDK2." *ACS Chem. Biol.* **6**: 492–501 (2011).

Brown N.R., Noble M.E., Endicott J.A., Johnson L.N., "The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases." *Nat. Cell Biol.* **1**: 438–443 (1999).

Brown N.R., Noble M.E., Lawrie A.M., Morris M.C., Tunnah P., Divita G., Johnson, L.N., Endicott J.A., "Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity." *J. Biol. Chem.* **274**: 8746–8756 (1999).

Brown N.R., Lowe E.D., Petri E., Skamnaki V., Antrobus R., Johnson L.N., "Cyclin B and cyclin A confer different substrate recognition properties on CDK2." *Cell Cycle* **6**: 1350–1359 (2007).

Christodoulou M.S., Caporuscio F., Restelli V., Carlino L., Cannazza G., Costanzi E., Citti C., Lo Presti L., Pisani P., Battistutta R., Broggini M., Passarella D., Rastelli G. "Probing an allosteric pocket of CDK2 with small-molecules" *ChemMedChem* **12(1)**: 33-41 (2017).

Cox S., Radzio-Andzelm E., Taylor S.S., "Domain movements in protein kinases." *Curr. Opin. Struct. Biol.* **4**: 893–901 (1994).

Cozza G., Zanin S., Sarno S., Costa E., Girardi C., Ribaudo G., Salvi M., Zagotto G., Ruzzene M., Pinna L. A. "Design, validation and efficacy of bisubstrate inhibitors specifically affecting ecto-CK2 kinase activity." *Biochem J.* **471(3)**:415-30 (2015).

Davies T.G., Tunnah P., Meijer L., Marko D., Eisenbrand G., Endicott J.A., Noble M.E., "Inhibitor binding to active and inactive CDK2: The crystal structure of CDK2-cyclin A/indirubin-5-sulphonate." *Structure* **9**: 389–397 (2001).

De Bondt H.L., Rosenblatt J., Jancarik J., Jones H.D., Morgan D.O., Kim S.H., "Crystal structure of cyclin-dependent kinase 2." *Nature*, **363**: 595–602 (1993).

De Vivo M., Bottegoni G., Berteotti A., Recanatini M., Gervasio F.L., Cavalli A. "Cyclin-dependent kinases: Bridging their structure and function through computations." *Future Med. Chem.* **3**: 1551–1559 (2011).

Draetta G., Beach D. "Activation of cdc2 protein kinase during mitosis in human cells: cell cycle-dependent phosphorylation and subunit rearrangement." Cell **54**: 17–26 (1988).

Emsley P., Cowtan K. "Coot: model-building tools for molecular graphics." *Acta Crystallographica* **D60**: 2126-2132 (2004).

Emsley P., Lohkamp B., Scott W.G., Cowtan K. "Features and Development of Coot." *Acta Crystallographica* **D66**: 486-501 (2010).

Endicott J. A., Noble M. E., Tucker J. A. "Cyclin-dependent kinases: inhibition and substrate recognition." *Curr. Opin. Struct. Biol.* **9**: 738–744 (1999).

Esposito L., Indovina P., Magnotti F., Conti D., Giordano A., "Anticancer therapeutic strategies based on CDK inhibitors." *Curr. Pharm. Des.* **19**: 5327–5332 (2013).

Evans P. "Scaling and assessment of data quality." *Acta Crystallographica* **D62**: 72–82 (2005).

Evans P.R. “An introduction to data reduction: space-group determination, scaling and intensity statistics” *Acta Crystallographica* **D67**: 282-292 (2011).

Gray N., Detivaud L., Doerig C., Meijer L., “ATP-site directed inhibitors of cyclin-dependent kinases.” *Curr. Med. Chem.* **6**: 859–875 (1999).

Holmes J.K., Solomon M.J., “The role of Thr160 phosphorylation of Cdk2 in substrate recognition.” *Eur. J. Biochem.* **268**: 4647–4653 (2001).

Jeffrey P.D., Russo A.A., Polyak K., Gibbs E., Hurwitz J., Massague J., Pavletich N.P. “Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex.” *Nature* **376**: 313–320 (1995).

Kabsch W., “XDS”, *Acta Cryst.* **D66**: 125-132 (2010).

Kawana H., Tamaru J.-I., Tanaka T. Hirai A., Saito Y., Kitagawa M., Mikata A., Harigaya K., Kuriyama T. “Role of p27Kip1 and cyclin-dependent kinase 2 in the proliferation of non-small cell lung cancer.” *Am. J. Pathol.* **153**: 505–513 (1998).

Knighton D.R., Zheng J.H., Ten Eyck L.F., Ashford V.A., Xuong N.H., Taylor S.S., Sowadski J.M. “Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.” *Science* **253**: 407–414 (1991).

Lees J.A., Weinberg R.A., “Tossing monkey wrenches into the clock: New ways of treating cancer.” *Proc. Natl. Acad. Sci. USA* **96**: 4221–4223 (1999).

Li Y., Zhang J., Gao W., Zhang L., Pan Y., Zhang S., Wang Y. "Insights on Structural Characteristics and Ligand Binding Mechanisms of CDK2." *Int. J. Mol. Sci.* **16(5)**: 9314-40 (2015).

Lolli G., Johnson L.N., "CAK-Cyclin-dependent Activating Kinase: A key kinase in cell cycle control and a target for drugs?" *Cell Cycle* **4**: 572–577 (2005).

Malumbres M., Pevarello P., Barbacid M., Bischoff J.R., "CDK inhibitors in cancer therapy: What is next?" *Trends Pharmacol. Sci.*, **29**: 16–21 (2008).

Martin M.P., Alam R., Betzi S., Ingles D.J., Zhu J.Y., Schonbrunn E., "A novel approach to the discovery of small-molecule ligands of CDK2." *Chembiochem* **13**: 2128–2136 (2012).

Matsushime H., Ewen M.E., Strom D.K., Kato J.Y., Hanks S.K., Roussel M.F., Sherr C.J. "Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins." *Cell* **71**: 323–334 (1992).

Meyerson M., Harlow E. "Identification of G1 kinase activity for cdk6, a novel cyclin D partner." *Mol. Cell. Biol.* **14**: 2077–2086 (1994).

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. "Phaser crystallographic software." *J. Appl. Cryst* **40**: 658-674 (2007).

Morgan D.O. "Principles of CDK regulation." *Nature* **374**: 131–134 (1995).

Nelson P.J., Shankland S.J. “Therapeutics in renal disease: the road ahead for antiproliferative targets” *Nephron Exp. Nephrol.* **103**: 6-15 (2005).

Nolen B., Taylor S., Ghosh G. “Regulation of protein kinases: Controlling activity through activation segment conformation.” *Mol. Cell* **15**: 661–675 (2004).

Norbury C., Nurse P. “Animal cell cycles and their control” *Annu. Rev. Biochem.* **61**: 441–470 (1992).

Pagano M., Pepperkok R., Verde F., Ansorge W., Draetta G. “Cyclin A is required at two points in the human cell cycle.” *EMBO J.* **11**: 961–971 (1992).

Palmieri L., Rastelli G., “αC displacement as a general approach for allosteric modulation of protein kinases” *Drug Discov. Today* **18(7-8)**: 407-14 (2013).

Pavletich N.P. “Mechanisms of cyclin-dependent kinase regulation: Structures of cdks, their cyclin activators, and cip and INK4 inhibitors.” *J. Mol. Biol.* **287**: 821–828 (1999).

Pines J. “Cyclins: Wheels within wheels.” *Cell Growth Differ.* **2**: 305–310 (1991).

Pines J. “Four-dimensional control of the cell cycle.” *Nat. Cell Biol.* **1**: 73–79 (1999).

Rastelli G., Anighoro A., Chripkova M., Carrassa L., Broggini M. "Structure-based discovery of the first allosteric inhibitors of cyclin-dependent kinase 2." *Cell Cycle* **13**: 2296–2305 (2014).

Russo A.A., Jeffrey P.D., Pavletich N.P. "Structural basis of cyclin-dependent kinase activation by phosphorylation." *Nat. Struct. Biol.* **3**: 696–700 (1996).

Sambrook J., Russel D.W. "Molecular Cloning. A Laboratory Manual." *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. U.S.A.* (2001).

Shapiro G.I. "Cyclin-dependent kinase pathways as targets for cancer treatment." *J. Clin. Oncol.* **24**: 1770–1783 (2006).

Sherr C.J. "Cancer cell cycles." *Science* **274**: 1672–1677 (1996).

Smith P.D., Crocker S.J., Jackson-Lewis V., Jordan-Sciutto K.L., Hayley S., Mount M.P., O'Hare M.J., Callaghan S., Slack R.S., Przedborski S. "Cyclin dependent kinase 5 is a mediator of dopaminergic neuron loss in a mouse model of Parkinson's disease" *Proc. Natl. Acad. Sci.* **100**: 13650-13655 (2003).

Suryadinata R., Sadowski M., Sarcevic B. "Control of cell cycle progression by phosphorylation of cyclin-dependent kinase (CDK) substrates." *Biosci. Rep.* **30(4)**: 243-55 (2010).

Tsai L.H., Lee M.S., Cruz J. "Cdk5, a therapeutic target for Alzheimer's disease?" *Biochim. Biophys. Acta-Proteins Proteom.* **1697**: 137-142 (2004).

Wang J., Liu S., Fu Y., Wang J.H., Lu Y. "Cdk5 activation induces hippocampal CA1 cell death by directly phosphorylating NMDA receptors" *Nat. Neurosci.* **6**: 1039-1047 (2003).

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., Wilson K.S. "Overview of the CCP4 suite and current developments." *Acta Crytsallographica* **D67**: 235-242 (2011).

PART 2.2

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis W.I., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C. and Zwart P.H., "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Cryst.* **D66**: 213-221 (2010).

Battistutta R., Lolli G. "Structural and functional determinants of protein kinase CK2α: facts and open questions." *Mol Cell Biochem*. **356**: 67-73 (2011).

Burnett G., Kennedy E.P. "The enzymatic phosphorylation of proteins." *J. Biol. Chem.* **211**: 969-980 (1954).

Canton D.A., Zhang C., Litchfield D.W. "Assembly of protein kinase CK2: investigation of complex formation between catalytic and regulatory subunits using a zinc-finger deficient mutant of CK2β." *Biochem J.* **358**: 87-94 (2001).

Chantalat L., Leroy D., Filhol O., Nueda A., Benitez M. J., Chambaz E.M., Cochet C., Dideberg O. "Crystal structure of the human protein kinase CK2 regulatory subunit reveals its zinc finger-mediated dimerization." *EMBO J.* **18**: 2930-2940 (1999).

Emsley P., Cowtan K. "Coot: model-building tools for molecular graphics." *Acta Crystallographica* **D60**: 2126-2132 (2004).

Emsley P., Lohkamp B., Scott W.G., Cowtan K. “Features and Development of Coot.” *Acta Crystallographica* **D66**: 486-501 (2010).

Evans P. “Scaling and assessment of data quality.” *Acta Crystallographica* **D62**: 72–82 (2005).

Evans P.R. “An introduction to data reduction: space-group determination, scaling and intensity statistics” *Acta Crystallographica* **D67**: 282-292 (2011).

Filhol O., Nueda A., Martel V., Gerber-Scokaert D., Benitez M.J., Souchier C., Saoudi Y., Cochet C. “Live-cell fluorescence imaging reveals the dynamics of protein kinase CK2 individual subunits.” *Mol. Cell. Biol.* **23**: 975–987 (2003).

Glover C.V. “A filamentous form of Drosophila casein kinase II.” *J Biol Chem.* **261**: 14349-14354 (1986).

Guerra B., Issinger O.G. “Protein kinase CK2 in human diseases.” *Curr. Med. Chem.* **15**: 1870–1886 (2008).

Hanahan D., Weinberg R. A. “The hallmarks of cancer.” *Cell* **100**: 57-705 (2000).

Hübner G.M., Larsen J.N., Guerra B., Niefind K., Vrecl M., Issinger O.G. “Evidence for aggregation of protein kinase CK2 in the cell: a novel strategy for studying CK2 holoenzyme interaction by BRET2.” *Mol. Cell. Biochem.* **397**: 285–293 (2014).

Kabsch W., “XDS”, *Acta Cryst.* **D66**: 125-132 (2010).

Leroy D., Schmid N., Behr J. P., Filhol O., Pares S., Garin J., Bourgarit J. J., Chambaz E. M., Cochet C. “Direct identification of a polyamine binding domain on the regulatory subunit of the protein kinase casein kinase 2 by photoaffinity labeling.” *J Biol Chem.* **270**: 17400-17406 (1995).

Litchfield D.W., Lozeman F.J., Cicirelli M.F., Harrylock M., Ericsson L.H., Piening C.J., Krebs E.G. “Phosphorylation of the beta subunit of casein kinase II in human A431 cells. Identification of the autophosphorylation site and a site phosphorylated by p34cdc2.” *J. Biol. Chem.* **266(30)**: 20380-20389 (1991).

Lolli G., Pinna L.A., Battistutta R. “Structural determinants of protein kinase CK2 regulation by autoinhibitory polymerization.” *ACS Chem Biol* **7(7)**: 1158-1163 (2012).

Lolli G., Ranchio A., Battistutta R. “Active form of the protein kinase CK2 α2β2 holoenzyme is a strong complex with symmetric architecture.” *ACS Chem. Biol.* **9**: 366–371 (2014).

Lolli G., Naressi D., Sarno S., Battistutta R. “Characterization of the oligomeric states of the CK2 α2β2 holoenzyme in solution” *Biochemical Journal* **474**: 2405–2416 (2017).

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. “Phaser crystallographic software.” *J. Appl. Cryst* **40**: 658-674 (2007).

Niefind K., Guerra B., Pinna L.A., Issinger O.G., Schomburg D. “Crystal structure of the catalytic subunit of protein kinase CK2 from *Zea mays* at 2.1 Å resolution.” *EMBO J.* **17**: 2451-2462 (1998).

Niefind K., Guerra B., Ermakowa I., Issinger O.G. “Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme.” *EMBO J.* **20**: 5320-5331 (2001).

Niefind K., Raaf J., Issinger O.G. “Protein kinase CK2 in health and disease: Protein kinase CK2: from structures to insights.” *Cell Mol Life Sci.* **66**: 1800-1816 (2009).

Niefind K., Battistutta R. “Structural bases of protein kinase CK2 function and inhibition. In Protein Kinase CK2” (Pinna, L.A., ed.) Wiley-Blackwell, Oxford, U.K 3–75, (2013).

Pagano M.A., Sarno S., Poletto G., Cozza G., Pinna L.A., Meggio, F. “Autophosphorylation at the regulatory beta subunit reflects the supramolecular organization of protein kinase CK2.” *Mol. Cell. Biochem.* **274**: 23−29 (2005).

Papinutto E., Ranchio A., Lolli G., Pinna L.A., Battistutta R. “Structural and functional analysis of the flexible regions of the catalyticα-subunit of protein kinase CK2.” *J. Struct. Biol.* **177**: 382–391 (2012).

Pinna L.A. “Protein kinase CK2: a challenge to canons.” *J Cell Sci.* **115**: 3873-3878 (2002).

Pinna L.A. (ed.) “Protein Kinase CK2.” Wiley-Blackwell, Oxford, U.K (2013).

Ruzzene M., Pinna L. A. “Addiction to protein kinase CK2: a common denominator of diverse cancer cells?” *Biochimica et Biophysica Acta* **1804(3)**: 499-504 (2010).

Sambrook J., Russel D.W. “Molecular Cloning. A Laboratory Manual.” *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. U.S.A.*, (2001).

Schnitzler A., Olsen B.B., Issinger O.G., Niefind K. “The protein kinase CK2 Andante holoenzyme structure supports proposed models of autoregulation and trans-autophosphorylation.” *J. Mol. Biol.* **426**: 1871–1882 (2014).

St-Denis N.A., Litchfield D.W. “From birth to death: The role of protein kinase CK2 in the regulation of cell proliferation and survival.” *Cell. Mol. Life Sci.* **66**: 1817-1829 (2009).

Theis-Febvre N., Martel V., Laudet B., Souchier C., Grunwald D., Cochet C., Filhol O. “Highlighting protein kinase CK2 movement in living cells.” *Mol Cell Biochem* **274**: 15-22 (2005).

Valero E., De Bonis S., Filhol O., Wade R.H., Langowski J., Chambaz E.M., Cochet C. “Quaternary structure of casein kinase 2 characterization of multiple oligomeric states and relation with its catalytic activity.” *J. Biol. Chem.* **270**: 8345–8352 (1995).

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., Wilson K.S. “Overview of the CCP4 suite and current developments.” *Acta Crytsallographica* **D67**: 235-242 (2011).

PART 3.1

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis W.I., Echols N., Headd J.J., Hung L.W., Kapral G.J., Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C. and Zwart P.H., "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Cryst.* **D66**: 213-221 (2010).

Besbes S., Mirshahi M., Pocard M., Billard C. "New dimension in therapeutic targeting of Bcl-2 family proteins." *Oncotarget* **6**: 12862-71 (2015).

Czabotar P.E., Lee E.F., Thompson G.V., Wardak A.Z., Fairlie D.W., Colman P.M. "Mutation to Bax beyond the BH3 Domain disrupts interactions with pro-apoptotic survival proteins and promotes apoptosis." *The journal of biological chemistry* **286**: 7123-7131 (2011).

Emsley P., Cowtan K. "Coot: model-building tools for molecular graphics." *Acta Crystallographica* **D60**: 2126-2132 (2004).

Emsley P., Lohkamp B., Scott W.G., Cowtan K. "Features and Development of Coot." *Acta Crystallographica* **D66**: 486-501 (2010).

Evans P. "Scaling and assessment of data quality." *Acta Crystallographica* **D62**: 72–82 (2005).

Evans P.R. "An introduction to data reduction: space-group determination, scaling and intensity statistics" *Acta Crystallographica* **D67**: 282-292 (2011).

Follis A.V., Chipul J.E., Fisher J.C., Yun M.K., Grace C.R., Nourse A., Baran K., Ou L., Min L., White S.W., Green D.R., Kriwacki R.W. "PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis." *Nature Chemical Biology* **9**: 163-168 (2013).

Gaulard P. "Expression of the bcl-2 gene product in follicular lymphoma." *Am. J. Pathol* **140(5)**: 1089-1095 (1992).

Gewies A. "Introduction to Apoptosis." *ApoRewiev* 1-26 (2003).

Harada N. "Expression of Bcl-2 family of proteins in fresh myeloma cells." *Leukemia* **12(11)**: 1817-1820 (1998).

Higashiyama M., Doi O., Kodoama K., Yokouchi H., Tateishi R. "High prevalence of bcl-2 oncoprotein expression in small cell lung cancer." *Anticancer Res.* **15(2)**: 503-505 (1995).

Kabsch W. "XDS." *Acta Cryst.* **D66**: 125-132 (2010).

Kerr J.F., Wyllie A.H., Currie A.R. "Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics." *British Journal of Cancer* **26**: 239-257 (1972).

Keshgegian A. A., Johnston E., Cnaan A. "Bcl-2 oncoprotein positivity and hogh MIB-1 (Ki-67) proliferative rate are independent predictive markers for recurrence in prostate carcinoma." *Am. J. Clin. Pathol.* **110**: 443-439 (1998).

Lee E.F., Czabotar P.E., Smith B.J., Deshayes K., Zobel K., Colman P. M., Fairlie W.D. "Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family." *Nature* **14**: 1711-13 (2007).

Leiter U., Schmid R.M., Kaskel P., Peter R.U., Krahn G. "Antiapoptotic bcl-2 and bcl-xL in advanced malignant melanoma." *Arch. Dermatol. Res.* **292(5)**: 225-232. (2000).

Lessene G., Czabotar P., Colman P.M. "Bcl-2 family antagonists for cancer therapy." *Nature Reviews* **7**: 989-1000 (2008).

Lipponen P. "Apoptosis suppressing protein bcl-2 is expressed in well-differentiated breast carcinomas with favourable prognosis." *J. Pathol* **177(1)**: 49-55 (1995).

Liu X., Beugelsdijk A., Chen J. "Dynamics of the BH3-only protein binding interface of Bcl-xL." *Biophysical Journal* **109**: 1049-57 (2015).

Lodish H.F., Berk A., Kaiser C.A., Krieger M., Bretscher, Ploegh, "Molecular Cell Biology." New York: Katherine Ahr Parker (2013).

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., Read R.J. "Phaser crystallographic software." *J. Appl. Cryst* **40**: 658-674 (2007).

Min K. H., Reynods, C.P. “Bcl-2 Inhibitors: Targeting Mitochondrial Apoptotic Pathways in Cancer Therapy.” *Clinical Cancer Research* **15(4)**: 1126-1132 (2009).

Muchmore S.W., Sattler M., Liang H., Meadows R.P., Harlan J.E., Yoon H.S., Nettesheim D., Chang B.S., Thompson C.B., Wong S.L., Ng S.L., Fesik S.W. “X-ray and NMR structure of human Bcl-xL, an inhbitor of programmed cell death.” *Nature* **381**: 335-341 (1996).

Nguyen M., Marcellus R.C., Roulston A., Watson M., Serfass L., Murthy Madiraju S. R., Goulet D., Viallet J., Bélec L., Billot X., Acoca S., Purisima E., Wiegmans A., Cluse L., Johnstone R.W., Beauparlant P., Shore G.C. “Small molecule obatoclax (GX15-070) antagonizes Mcl-1 and overcomes Mcl-1-mediated resistace to apoptosis.” *Proc. Natl. Acad. Sci.* **104**: 19512-17 (2007).

Obersatain A., Jeffrey P., Shi Y. “Crystal structure of the Bcl-XL-Beclin 1 peptide complex: Beclin 1 is a novel BH3-only protein.” *Journal of Biological Chemistry* **282**: 13123-32 (2007).

Oltersdorf T., Elmore S.W., Shoemaker A.R., Armstrong R.C., Augeri D.J., Belli B.A., Bruncko M., Deckwerth T.L., Dinges J., Hajduk P.J., Joseph M.K., Kitada S., Korsmeyer S.J., Kunzer A.R., Letai A., Li C., Mitten M.J., Nettesheim D.G., Ng S., Nimmer P.M., O'Connor J.M., Oleksijew A., Petros A.M., Reed J.C., Shen W., Tahir S.K., Thompson C.B., Tomaselli K.J., Wang B., Wendt M.D., Zhang H., Fesik S.W., Rosenberg S.H. “An inhibitor of Bcl-2 family proteins induces regression of solid tumors.” *Nature* **435**: 677-681 (2005).

Petros A.M., Nettesheim D.G., Wang Y., Olejniczak E.T., Meadows R.P., Mack J., Swift K., Matayoshi E.D., Zhang H., Thompson C.B., Fesik S.W. "Rationale for Bcl-XL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies." *Protein Science* **9**: 2528-34 (2000).

Petros A.M., Olejniczak E.T., Fesik S.W. "Structural biology of the Bcl-2 family of proteins." *Biochimica et Biophysica Acta* **1664(2-3)**: 83-94 (2004).

Rajan S., Choi M., Baek K., Yoon H.S. "Bh3 induced conformational changes in Bcl-Xl revealed by crystal structure and comparative analysis." *Proteins: structure, function and bioinformatics* **83**: 1262-72 (2015).

Sambrook J., Russel D.W. "Molecular Cloning. A Laboratory Manual." *Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. U.S.A.*, (2001).

Sattler M., Liang H., Nettesheim D., Meadows R.P., Harlan J.E., Eberstadt M., Yoon H.S., Shuker S.B., Chang B.S., Minn A.J., Thompson C.B., Fesik S.W. "Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis." *Science* **275**: 983-986 (1997).

Shamas-Din A., Kale J., Leber B., Andrews D.W. "Mechanism of Action of Bcl-2 Family Proteins." *Cold Spring Harb Perspect Biol* **5(4)**: a008714. (2013).

Touzeau C., Dousset C., Le Gouille S., Sampath D., Leverson J.D., Souers A.J., Maïga S., Béné M.C., Moreau P., Pellat-Deceunynck C., Amiot M. "The Bcl-2 specific BH3 mimetic ABT-199: a promising targeted therapy for t(11;14) multiple myeloma." *Leukemia* **28**: 210-12 (2014).

Tse C., Shoemaker A.R., Adickes J., Anderson M.G., Chen J., Jin S., Johnson E.F., Marsh K.C., Mitten M.J., Nimmer P., Roberts L., Tahir S.K., Xiao Y., Yang X., Zhang H., Fesik S., Rosenberg S.H., Elmore S.W. "ABT-263: A potent and orally bioavailable Bcl-2 family inhibitor." *Cancer Res* **68(9)**: 3421-28 (2008).

Tsujimoto Y., Croce C.M. "Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphomas." *Proc. Natl. Acad. Sci. USA* **83**: 5214-18 (1986).

Vela L., Isabel M. "Bcl-2 family of proteins as drug targets for cancer chemotherapy: the long way of BH3 mimetics from bench to beside." *Current Opinion in Pharmacology* **23**: 74-81 (2015).

Wang G., Nikolovska-Coleska Z., Yang C.Y., Wang R., Tang G., Guo J., Shangary S., Qiu S., Gao W., Yang D., Meagher J., Stuckey J., Krajewski K., Jiang S., Roller P.P., Abaan H.O., Tomita Y., Wang S. "Structure-based design of potent small-molecule inhibitors of anti-apoptotic Bcl-2 proteins." *J Med Chem* **49**: 6139-42 (2006).

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., Wilson K.S. "Overview of the CCP4 suite and current developments." *Acta Crytsallographica* **D67**: 235-242 (2011).

Youle R. J., Strasser A. "The BCL-2 protein family: opposing activities that mediate cell death." *Nat Rev Mol Cell Biol* **9(1)**: 47-59. (2008).

PART 3.2

Agbulut O., Destombes J., Thiesson D., Butler-Browne G. "Age-related appearance of tubular aggregates in the skeletal muscle of almost all male inbred mice." *Histochem Cell Biol* **114**: 477–481 (2000).

Barone V., Del Re V., Gamberucci A., Polverino V., Galli L., Costanzi E., Toniolo L., Berti G., Malandrini A., Ricci G., Siciliano G., Vattemi G., Tomelleri G., Pierantozzi E., Spinozzi S., Volpi N., Fulceri R., Battistutta R., Reggiani C., Sorrentino V. "Identification and characterization of three novel variants in the CASQ1 gene in four patients with tubular aggregate myopathy." *Hum mutat* (2017).

Beard N.A., Dulhunty A.F. "C-terminal residues of skeletal muscle calsequestrin are essential for calcium binding and for skeletal ryanodine receptor inhibition." *Skelet Muscle* **5**: 6 (2015).

Böhm J., Chevessier F., De Paula A.M., Koch C., Attarian S., Feger C., Hantaï D., Laforêt P., Ghorab K., Vallat J., Fardeau M., Figarella-Branger D., Pouget J., Romero N.B., Koch M., Ebel C., Levy N., Krahn M., Eymard B., Bartoli M., Laporte J. "Constitutive activation of the calcium sensor STIM1 causes tubular-aggregate myopathy." *Am J Hum Genet* **92**: 271-278 (2013).

Böhm J., Bulla M., Urquhart J.E., Malfatti E., Williams S.G., O'Sullivan J., Szlauer A., Koch C., Baranello G., Mora M., Ripolone M., Violano R., Moggio M., Kingston H., Dawson T., DeGoede C.G., Nixon J., Boland A., Deleuze J.F., Romero N., Newman W.G., Demaurex N., Laporte J. "ORAI1 mutations with distinct channel gating defects in tubular aggregate myopathy." *Hum Mutat* **38**: 426-438 (2017).

Boncompagni S., Protasi F., Franzini-Armstrong C. "Sequential stages in the age-dependent gradual formation and accumulation of tubular aggregates in fast twitch muscle fibers: SERCA and calsequestrin involvement." *Age (Dordr)* **34(1)**: 27-41 (2012).

Chevessier F., Bauche-Godard S., Leroy J.P., Koenig J., Paturneau-Jouas M., Eymard B., Hantai D., Verdiere-Sahuque M. "The origin of tubular aggregates in human myopathies." *J Pathol* **207**: 313-323 (2005).

Endo Y., Noguchi S., Hara Y., Hayashi Y.K., Motomura K., Miyatake S., Murakami N., Tanaka S., Yamashita S., Kizu R., Bamba M., Goto Y., Matsumoto N., Nonaka I., Nishino I. "Dominant mutations in ORAI1 cause tubular aggregate myopathy with hypocalcemia via constitutive activation of store-operated Ca2+ channels." *Hum Mol Genet* **24**: 637-648 (2015).

Engel W.K., Bishop D.W., Cunningham G.G. "Tubular aggregates in type II muscle fibers: ultrastructural and histochemical correlation." *J Ultrastruct Res* **31**: 507–525 (1970).

Giacomello E., Quarta M., Paolini C., Squecco R., Fusco P., Toniolo L., Blaauw B., Formoso L., Rossi D., Birkenmeier C., Peters L.L., Francini F., Protasi F., Reggiani C., Sorrentino V. "Deletion of small ankyrin 1 (sAnk1) isoforms results in structural and functional alterations in aging skeletal muscle fibers." *Am J Physiol Cell Physiol* **308**: C123-38 (2015).

Kuncl R.W., Pestronk A., Lane J., Alexander E. "The MRL +/+ mouse: a new model of tubular aggregates which are gender- and age-related." *Acta Neuropathol* **78**: 615–620 (1989).

Lewis K.M., Ronish L.A., Ríos E., Kang C. “Characterization of Two Human Skeletal Calsequestrin Mutants Implicated in Malignant Hyperthermia and Vacuolar Aggregate Myopathy.” *J Biol Chem* **290**: 28665-28674 (2015).

Okuma H., Saito F., Mitsui J., Hara Y., Hatanaka Y., Ikeda M., Shimizu T., Matsumura K., Shimizu J., Tsuji S., Sonoo M. “Tubular aggregate myopathy caused by a novel mutation in the cytoplasmic domain of STIM1.” *Neurol Genet* **2(1)**:e50 (2016).

Park H., Wu S., Dunker A.K., Kang C. “Polymerization of calsequestrin. Implications for Ca2+ regulation.” *J Biol Chem* **278**: 16176-16182 (2003).

Park H., Park I.Y., Kim E., Youn B., Fields K., A. Keith Dunker K.A., Chul Hee Kang C. “Comparing Skeletal and Cardiac Calsequestrin Structures and Their Calcium Binding.” *J Biol Chem* **279**: 18026-18033 (2004).

Perni S., Close M., Franzini-Armstrong C. “Novel details of calsequestrin gel conformation in situ.” *J Biol Chem* **288**: 31358-31362 (2013).

Rossi D., Vezzani B., Galli L., Paolini C., Toniolo L., Pierantozzi E., Spinozzi S., Barone V., Pegoraro E., Bello L., Cenacchi G., Vattemi G., Tomelleri G., Ricci G., Siciliano G., Protasi F., Reggiani C., Sorrentino V. “A mutation in the CASQ1 gene causes a vacuolar myopathy with accumulation of sarcoplasmic reticulum protein aggregates.” *Hum Mutat* **35(10)**: 1163-1170 (2014).

Salviati G., Pierobon-Bormioli S., Betto R., Damiani E., Angelini C., Ringel S.P., Salvatori S., Margreth A. “Tubular aggregates: sarcoplasmic reticulum origin, calcium storage ability, and functional implications." *Muscle Nerve* **8(4)**: 299–306 (1985).

Schiaffino S., Reggiani C. "Fiber types in mammalian skeletal muscles." *Physiol Rev* **91**: 1447–1531 (2011).

Schiaffino S. "Tubular aggregates in skeletal muscle: just a special type of protein aggregates?" *Neuromuscul Disord* **22**: 199–207 (2012).

Shin D.W., Pan Z., Kim E.K., Lee J.M., Bhat M.B., Parness J., Kim D.H., Ma J. "A retrograde signal from calsequestrin for the regulation of store-operated Ca2+ entry in skeletal muscle." *J Biol Chem.* **278**: 3286-92 (2003).

Tomelleri G., Palmucci L., Tonin P., Mongini T., Marini M., L'Erario R., Rizzuto N., Vattemi G. "SERCA1 and calsequestrin storage myopathy: a new surplus protein myopathy." *Brain* **129**: 2085-2092 (2006).

Wang S., Trumble W.R., Liao H., Wesson C.R., Dunker A.K., Kang C.H. "Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum." *Nat Struct Biol* **5**: 476-483 (1998).

Wang L., Zhang L., Li S., Zheng Y., Yan X., Chen M., Wang H., Putney J.W., Luo D. "Retrograde regulation of STIM1-Orai1 interaction and store-operated Ca2+ entry by calsequestrin." *Sci Rep* **5**: 11349 (2015).

Zhang L., Wang L., Li S., Xue J., Luo D. "Calsequestrin-1 Regulates Store-Operated Ca2+ Entry by Inhibiting STIM1 Aggregation." *Cell Physiol Biochem* **38(6)**: 2183-2193 (2016).

Zhao X., Min C.K., Ko J.K., Parness J., Kim D.H., Weisleder N., Ma J. "Increased store-operated Ca2+ entry in skeletal muscle with reduced calsequestrin-1 expression" *Biophys J.* **99**: 1556-64 (2010).